Study of ZZ+jets and Electroweak ZZ+2jets Production in Proton-Proton Collisions at $\sqrt{s} = 13\,\text{TeV}$ in Four-Lepton Events Using the CMS Detector at the LHC

by

He He

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Physics)

at the

University of Wisconsin – Madison

2025

Date of final oral examination: 12/05/2025

The dissertation is approved by the following members of the Final Oral Committee:

Matthew Herndon, Professor of Physics

Kevin Black, Professor of Physics

Lisa Everett, Professor of Physics

Oliver Schmitz, Professor of Nuclear Engineering and Engineering Physics

Abstract

Two studies are presented on diboson production in association with jets in the fully leptonic final states, pp $\rightarrow$ (Z/γ^*)(Z/γ^*) + jets $\rightarrow$ $2\ell 2\ell'$ + jets, ($\ell, \ell' =$ e or μ) in proton-proton collisions at a center-of-mass energy of 13 TeV. The data sample corresponds to an integrated luminosity of 138 fb^{-1} collected with the CMS detector at the LHC. In the first study, search is performed for the electroweak (EW) production of two Z bosons in association with two jets (EW ZZ+2jets) using a deep neural network (DNN) approach, and the production is measured with an observed (expected) significance of 4.4 (3.4) standard deviations. In the second study, differential distributions and normalized differential cross sections of diboson ZZ production associated with different numbers of jets (ZZ+jets) are measured as a function of jet multiplicity, transverse momentum p_{T}, pseudorapidity η, invariant mass and $\Delta\eta$ of the highest-p_{T} and second-highest-p_{T} jets, and as a function of invariant mass of the four-lepton system for events with various jet multiplicities. These differential cross sections are compared with theoretical predictions that mostly agree with the experimental data. However, in a few regions we observe discrepancies between the predicted and measured values. Further improvement of the predictions is required to describe the ZZ+jets production in the whole phase space.

Acknowledgements

I could not have reached this stage in my life without the help and encouragement of many many wonderful people. Unfortunately I am not able to list every one of these people within this limited space of text, but they are all in my mind along with my precious memories.

I would like to first thank the people in the UW-Madison CMS group. My supervisor, Prof. Matthew Herndon, has given me countless crucial and invaluable helps, detailed advice and guidance, in a large number of aspects throughout my entire Ph.D. study, which I am all really grateful for. Alexander Savin has taught me a lot of useful knowledge and ways of thinking during my physics analyses, and I really appreciate the large amount of helpful suggestions, encouragement, and patient explanations from him. Prof. Sridhara Dasu always gives me a lot of useful advice and important reminders. I also enjoy meeting with Prof. Wesley Smith, Prof. Kevin Black and Prof. Tulika Bose, who have also made essential contributions in creating such an active and energetic atmosphere inside the group as seen today. I was lucky to meet several excellent senior students (now alumni) since early in my Ph.D. study: Kenneth Long has given me many useful helps and suggestions in crucial stages of my Ph.D. studies, and Usama Hussain has provided me with important guidance when I struggled to dive into my physics analyses at the beginning. Although I haven't formally got to know Nathaniel Woods, I have learned and benefited a lot from his notes and thesis, and his well-organized and well-structured coding styles. I would like to also thank all the people who have contributed to the development of the ZZ analysis framework of codes used by the UW-Madison CMS group (including the three senior students mentioned). I have always enjoyed talking and gathering with my fellow groupmates and also my friends, to name a few of them: Abhishikth Mal-

lampalli, Victor Shang, Ho Fung Tsoi, Mitanshu Thakore, Susmita Mondal, Justin Marquez, Dylan Teague and Wren Vetens, and I always remember their help, support and encouragement. During my work with the CMS muon CSC, Isabelle De Bruyn and Armando Lanaro have also given me friendly help and guidance, and I couldn't have had a smooth start with my detector work without their help. I have also received strong support from people in the computing team, in particular Carl Vuosalo and Dan Bradley, who have been providing very prompt and useful information, diagnostics and instructions. Besides people in the UW CMS group, I would also like to thank many people in the UW-Madison physics department for helping to create such a friendly and welcoming environment.

As one would naturally expect, I have also received a lot of help from other people in the CMS Collaboration, besides those already mentioned within the UW CMS group. Riccardo Bellan and Roberto Covarelli in the CMS group from Università degli Studi di Torino and INFN, and Claude Charlot from Laboratoire Leprince-Ringuet, have provided important inputs, feedbacks and helps for the ZZ+jets analysis and for the EW ZZ+2jets analysis in this thesis. The results for the ZZ+jets analysis in this thesis have also been published by the CMS Collaboration in Ref. [1] (in this thesis we inherit some of the text from this paper), and this was achieved with the effort of the entire CMS Collaboration, including effort from those who have reviewed, commented on the paper and facilitated with the publication. In particular, the ARC members for the ZZ+jets analysis, Nicholas Hadley, Francesca Cavallari, Barbara Gonzalez, and Darien Wood have given many helpful comments and suggestions. As explained in Section 1.7, the ZZ+jets analysis in this thesis is an extension to the CMS inclusive ZZ analysis [2], and we have used the previous CMS ZZ+jets analysis [3] as an important reference for analysis methods in the current ZZ+jets analysis. On the other hand, the EW ZZ+2jets analysis in this thesis serves as a cross check for the published CMS

analysis [4]. I would also like to thank all the enthusiastic and responsible people in the CMS muon CSC team, in particular Mikhail Ignatenko and Jian Wang for their trust, patience and prompt guidance. Some of the text in this thesis discussing the CMS detector, event reconstruction and object identification is from the internal standard text provided within the CMS Collaboration.

Regarding the ZZ+jets analysis, I would like to also thank the authors of [5] for providing the samples and detailed information necessary for making the comparison to the nNNLO+PS predictions, which is a very important part of the results.

In my daily life, I have been receiving warm greetings and encouragements from Yulang Cen and Yuqian Xu, two of my best friends since childhood, and Yuqing Liang, a wonderful friend of mine since high school, as well as from several of my good old friends. It has been my very pleasure to have these long-lasting friendships and to be shared with important moments and feelings in their life.

Finally and very importantly, I would like to thank my family members, without whom I would have gone nowhere. My father, Jianfeng He, and my mother, Yanyu Yang, have been giving me plainly visible love and care throughout my life up until now. My dear wife, Xueting Xie, have always been very thoughtful and supportive, and I cherish my every day with her. I cannot thank them enough for all their love and for all the hope and happiness they have given me in my life.

In addition, many Feynman diagrams in this thesis were created with the graphical tool FeynGame [6–8], and I would like to thank the people behind it for creating such an excellent tool.

Contents

List of Figures

List of Tables

Chapter 1

The Standard Model

1.1 Introduction

Particle physics studies the elementary particles and their interactions. The Standard Model (SM) is a powerful and unified theory for particle physics that provides successful description of all current experimental data within uncertainty. It describes interactions among elementary particles by the exchange of particles. On the other hand, there have been experimental observations that cannot be accounted for by the SM, indicating the existence of physics beyond the SM (BSM) and the need for a more complete theory. For introduction and discussion on the SM we can refer to Refs. [9–17] and the Particle Data Group (PDG) review [18]. For specific discussion on quantum chromodynamics theory and topics related to simulation, one can refer to [12, 19–31], and "Monte Carlo event generators (rev.)" in the PDG review [18]. For topics on SM measurements one can refer also to [32]. Note that different authors can use different sign conventions when discussing about the SM, and for a comprehensive overview of different sign conventions one can refer to Ref. [33]. We have used these mentioned sources as our references in this chapter, along with other references cited

in the main discussion.

1.2 Matter and Forces

Currently all the known matter in this universe is formed by combinations of twelve fundamental spin-half fermions (and their antiparticles) as listed in Table 1.1, where spin is an intrinsic property of a particle associated with its internal angular momentum. These fermions are categorized into quarks and leptons by the type of interactions they participate in. The electron, muon, tau and neutrinos are collectively referred to as the leptons. These fermions are also grouped into three generations. The electron, electron neutrino, up quark and down quark belong to the first generation. Particles in the second and third generations behave like copies of the corresponding particles in the first generation, in the sense that they have the same kind of physical interactions, other than difference in mass which has various physical consequences such as those in decaying. Each of these particles has an antiparticle with the same mass but opposite charge.

There are currently four known fundamental forces: gravity, electromagnetic force, strong force and weak force. Our discussion about SM focus on the latter three forces, where each force is described by a quantum field theory (QFT) corresponding to the exchange of spin-1 boson(s) known as gauge boson. Electromagnetic interaction concerns electrically charged particles, and is described by the theory of quantum electrodynamics (QED) and the exchange of virtual photon. Quark and charged leptons carry electric charge whereas neutrinos do not carry electric charge. The strong force is responsible for the nuclear force binding protons and neutrons together in the atomic nucleus. It is described by quantum chromodynamics (QCD) and the exchange of a massless gauge boson called gluon. Among the fundamental fermions,

only the quarks carry the charge called "color charge" and correspondingly feel the strong force. The weak interactions is responsible for some of the important particle decays. It include the charged-current interactions mediated by the W^+ and W^- bosons with a mass of 80.4 GeV, and the neutral-current interactions mediated by the neutral Z boson with a mass of 91.2 GeV. All fundamental fermions are involved in weak interactions and we say they carry weak charge. The large masses of the W and Z bosons results in the suppression of the interaction probability at low energies (see for example eqn. (1.146)), causing a longer lifetime for particles that only decay through weak interactions.

The final element of SM is the Higgs boson. As will be discussed in Section 1.3, the Brout-Englert-Higgs (BEH) spontaneous electroweak symmetry-breaking mechanism [34–39] predicts the existence of a spin-0 scalar particle, the SM Higgs boson, which gives masses to massive elementary particles, and explains the large masses of the W and Z bosons that distinguish them from the massless gauge bosons, photon and gluon. Experimentally, the Higgs boson was discovered in the LHC by the CMS and ATLAS experiments in 2012 [40–42], and was measured to have a mass of around 125 GeV.

1.3 Electroweak Unification and Symmetry Breaking

1.3.1 The GSW Model

The GSW model developed by Glashow, Salam and Weinberg in the 1960s is a model describing a unified picture of electromagnetic and weak interactions. In the GSW model, we have a set of massless fermions and gauge bosons, and they obtain masses

Table 1.1: Table of fundamental fermions listed with their generation, electric charge Q, and masses. To be precise, $\nu_{\rm e}$, ν_μ, ν_τ do not have definite masses, but represent combination of quantum states ν_1, ν_2, ν_3 (mass eigenstates) that have definite masses ($< 0.8\,\text{eV}$ with 90% CL from Ref. [18]). The nominal mass values in this table are taken from Ref. [18].

	Lepton	Q	mass (GeV)
First generation	electron (e)	-1	0.000511
	neutrino ($\nu_{\rm e}$)	0	-
Second generation	muon (μ)	-1	0.106
	neutrino (ν_μ)	0	-
Third generation	tau (τ)	-1	1.78
	neutrino (ν_τ)	0	-

	Quark	Q	mass (GeV)
First generation	down (d)	$-\frac{1}{3}$	0.00216
	up (u)	$+\frac{2}{3}$	0.00470
Second generation	strange (s)	$-\frac{1}{3}$	0.0935
	charm (c)	$+\frac{2}{3}$	1.27
Third generation	bottom (b)	$-\frac{1}{3}$	4.18
	top (t)	$+\frac{2}{3}$	173

through the BEH mechanism. We can introduce the appropriate gauge fields by requiring invariance of the Lagrangian under local gauge transformation, and in GSW model we require invariance under $\mathrm{U}(1)_Y \times \mathrm{SU}(2)_L$. We start with the fermionic Dirac Lagrangian without mass terms:

$$\mathscr{L}_{\rm ferm} = \sum_{f=\text{lepton,quarks}} \bar{f} i\gamma^\mu \partial_\mu f. \tag{1.1}$$

We can group the left-handed electron and neutrino into a doublet:

$$L = \begin{pmatrix} \nu_{\rm e} \\ {\rm e}^- \end{pmatrix}_L \tag{1.2}$$

and we say that this doublet is with total weak isospin $I_{\rm w} = \frac{1}{2}$ and the upper and lower components have third component of isospin $I_{\rm w}^{(3)}\,(\nu_{\rm e}) = +\frac{1}{2}$ and $I_{\rm w}^{(3)}\,({\rm e}^-) = -\frac{1}{2}$ respectively. Similarly we have other doublets:

$$\begin{pmatrix}\nu_\mu\\ \mu^-\end{pmatrix}_L, \quad \begin{pmatrix}\nu_\tau\\ \tau^-\end{pmatrix}_L, \quad \begin{pmatrix}\mathrm{u}\\ \mathrm{d}\end{pmatrix}_L, \quad \begin{pmatrix}\mathrm{c}\\ \mathrm{s}\end{pmatrix}_L, \quad \begin{pmatrix}\mathrm{t}\\ \mathrm{b}\end{pmatrix}_L \tag{1.3}$$

These doublets are affected by the gauge transformation

$$\varphi(x) \to \varphi'(x) = \exp\left[ig_w \boldsymbol{\alpha}(x)\cdot\mathbf{T}\right]\varphi(x) \tag{1.4}$$

where $\mathbf{T} = \frac{1}{2}\sigma$ are the three generators of the SU(2) group written in terms of the Pauli matrices. On the other hand, the right-handed particle chiral states are treated as weak isospin singlets with $I_\mathrm{w} = I_\mathrm{w}^{(3)} = 0$ and therefore unaffected by the mentioned SU(2) gauge transformation:

$$\mathrm{e}_R^-, \quad \mu_R^-, \quad \tau_R^-, \quad \mathrm{u}_R, \quad \mathrm{c}_R, \quad \mathrm{t}_R, \quad \mathrm{d}_R, \quad \mathrm{s}_R, \quad \mathrm{b}_R \tag{1.5}$$

and the three types of ν_R. This explains the "L" subscript in $\mathrm{SU}(2)_L$. Note that we have been talking about particle states in the flavor basis, which is different from the mass basis for neutrinos and quarks. We can then also consider a local U(1) transformation affecting both the left- and right-handed particle chiral states:

$$\psi(x) \to \psi'(x) = \exp\left[ig'\frac{Y}{2}\zeta(x)\right]\psi(x), \tag{1.6}$$

where Y is the weak hypercharge, a charge newly introduced that is specific to the particle type and chiral states. We must require $Y_{\mathrm{e}_L} = Y_{\nu_L}$, otherwise the $\mathrm{U}(1)_Y$ transformation can introduce phase difference between the components of the $\mathrm{SU}(2)_L$ doublet and break the $\mathrm{SU}(2)_L$ symmetry. Therefore we can use the label Y_L for both components. In order to preserve invariance under the local gauge transformations, we replace the derivative in the Lagrangian with the covariant derivative:

$$\partial_\mu \to D_\mu = \partial_\mu + ig_w\mathbf{T}\cdot\mathbf{W}_\mu + ig'\frac{Y}{2}B_\mu \tag{1.7}$$

and the gauge fields transform according to:

$$W_\mu^k \to W_\mu'^k = W_\mu^k - \partial_\mu \alpha^k - g_{\mathrm{w}} \epsilon^{ijk} \alpha^i W_\mu^j \tag{1.8}$$

$$B_\mu \to B'_\mu = B_\mu - \partial_\mu \zeta(x) \tag{1.9}$$

Here the levi-Civita symbol

$$\varepsilon^{ijk} = \begin{cases} 1, \text{ if } ijk = 123, 231, \text{ or } 312 \\ -1, \text{ if } ijk = 132, 213, \text{ or } 321 \\ 0, \text{ otherwise} \end{cases} \tag{1.10}$$

is the structure constant for SU(2).

Now we can multiply out the covariant derivative to get the interaction terms. In the following calculations, we focus on e and ν_{e} for simplicity, but the same calculation can be repeated exactly for other leptons and for quarks. The terms corresponding to the U(1) transformation are:

$$\begin{aligned} \mathscr{L}_{\mathrm{U}(1)} &= \bar{L} i\gamma^\mu \left(ig' \frac{Y_L}{2} B_\mu \right) L + \bar{e}_R i\gamma^\mu \left(ig' \frac{Y_{\mathrm{e}_R}}{2} B_\mu \right) e_R + \bar{\nu}_R i\gamma^\mu \left(ig' \frac{Y_{\nu_R}}{2} B_\mu \right) \nu_R \\ &= -\frac{g'}{2} B_\mu \left[Y_L \left(\bar{\nu}_L \gamma^\mu \nu_L + \bar{e}_L \gamma^\mu e_L \right) + Y_{\mathrm{e}_R} \bar{e}_R \gamma^\mu e_R + Y_{\nu_R} \bar{\nu}_R \gamma^\mu \nu_R \right] \end{aligned} \tag{1.11}$$

The terms corresponding to the SU(2) transformation are:

$$\begin{aligned} \mathscr{L}_{\mathrm{SU}(2)} &= \bar{L} i\gamma^\mu \left[i g_w \frac{\sigma^i}{2} W_\mu^i \right] L \\ &= -\frac{g_w}{2} \begin{pmatrix} \bar{\nu}_L & \bar{e}_L \end{pmatrix} \gamma^\mu \begin{pmatrix} W_\mu^3 & W_\mu^1 - iW_\mu^2 \\ W_\mu^1 + iW_\mu^2 & -W_\mu^3 \end{pmatrix} \begin{pmatrix} \nu_L \\ e_L \end{pmatrix} \\ &= -\frac{g_w}{2} \begin{pmatrix} \bar{\nu}_L & \bar{e}_L \end{pmatrix} \gamma^\mu \begin{pmatrix} W_\mu^3 & \sqrt{2} W_\mu^+ \\ \sqrt{2} W_\mu^- & -W_\mu^3 \end{pmatrix} \begin{pmatrix} \nu_L \\ e_L \end{pmatrix} \\ &= -\frac{g_w}{2} \Big[\bar{\nu}_L \gamma^\mu \nu_L W_\mu^3 - \bar{e}_L \gamma^\mu e_L W_\mu^3 + \sqrt{2} \bar{\nu}_L \gamma^\mu e_L W_\mu^+ + \sqrt{2} \bar{e}_L \gamma^\mu \nu_L W_\mu^- \Big] \end{aligned} \tag{1.12}$$

where we have redefined the fields with

$$\mathrm{W}_{\mu}^{\pm} = \frac{1}{\sqrt{2}} \left(\mathrm{W}_{\mu}^{1} \mp i \, \mathrm{W}_{\mu}^{2} \right) \tag{1.13}$$

Collecting the $\bar{\nu}_L \gamma^\mu \nu_L$ terms from the SU(2) and U(1) terms, we have

$$(-\frac{g'}{2} Y_L B_\mu - \frac{g_w}{2} W_\mu^3) \bar{\nu}_L \gamma^\mu \nu_L \tag{1.14}$$

Therefore we can define a new field $Z_\mu \propto g' Y_L B_\mu + g_w W_\mu^3$ which is coupled to the neutrino and actually corresponds to the Z boson. Since the neutrino doesn't couple to the electromagnetic field, we can also define $A_\mu \propto g_w B_\mu - g' Y_L W_\mu^3$ which is orthogonal to Z_μ provided that B_μ and W_μ^i are orthogonal to each other, and let it be the electromagnetic field. Assuming B_μ and W_μ^i are in addition normalized, we can normalize the two new fields with

$$\begin{aligned} A_\mu &= \frac{g_w B_\mu - g' Y_L W_\mu^3}{\sqrt{g_w^2 + g'^2 Y_L^2}} = +B_\mu \cos\theta_w + W_\mu^3 \sin\theta_w, \\ Z_\mu &= \frac{g' Y_L B_\mu + g_w W_\mu^3}{\sqrt{g_w^2 + g'^2 Y_L^2}} = -B_\mu \sin\theta_w + W_\mu^3 \cos\theta_w \end{aligned} \tag{1.15}$$

where we have introduced the electroweak mixing angle θ_w by defining:

$$\begin{aligned} \sin\theta_w &= \frac{g'}{\sqrt{g'^2 + g_w^2}}, \\ \cos\theta_w &= \frac{g_w}{\sqrt{g'^2 + g_w^2}} \end{aligned} \tag{1.16}$$

and choosing a convenient convention

$$Y_L = -1 \tag{1.17}$$

Since Y_L always appears along with g', we are free to choose this convention and g' will have the corresponding definition. Inverting eqn. (1.15) gives:

$$\begin{aligned} B_\mu &= +A_\mu \cos\theta_w - Z_\mu \sin\theta_w, \\ W_\mu^3 &= +A_\mu \sin\theta_w + Z_\mu \cos\theta_w \end{aligned} \tag{1.18}$$

Substituting these into the previous equations allow us to express in terms of the fields A_μ and Z_μ. Now collecting the electron terms from the U(1) and SU(2) Lagrangian terms, we have:

$$-A_\mu \left\{ \bar{e}_L \gamma^\mu e_L \left[\frac{1}{2} g' Y_{\mathrm{e}_L} \cos\theta_w - \frac{1}{2} g_w \sin\theta_w \right] + \bar{e}_R \gamma^\mu e_R \left[\frac{1}{2} g' Y_{\mathrm{e}_R} \cos\theta_w \right] \right\} \tag{1.19}$$

In QED, the Feynman rule associated with the interaction vertex is

$$-iQe\gamma^\mu \tag{1.20}$$

and the interaction between electron and electromagnetic field corresponds to the Lagrangian terms:

$$-Q_{\mathrm{e}} e A_\mu (\bar{e}_L \gamma^\mu e_L + \bar{e}_R \gamma^\mu e_R) \tag{1.21}$$

Here e is the unit charge and Q is the charge number, with $Q_{\mathrm{e}} = -1$ for electron. Therefore to match with QED we have from the coefficient of the left-handed terms:

$$Q_{\mathrm{e}} e = \frac{1}{2} g' Y_{\mathrm{e}_L} \cos\theta_w - \frac{1}{2} g_w \sin\theta_w = -\frac{g' g_w}{\sqrt{g'^2 + g_w^2}} \tag{1.22}$$

which gives

$$e = \frac{g' g_w}{\sqrt{g'^2 + g_w^2}} = g_w \sin\theta_w = g' \cos\theta_w \tag{1.23}$$

and from the coefficient of the right-handed term:

$$Q_{\mathrm{e}} e = \frac{1}{2} g' Y_{\mathrm{e}_R} \cos\theta_w = -\frac{1}{2} \frac{g' g_w}{\sqrt{g'^2 + g_w^2}} \tag{1.24}$$

which gives

$$Y_{\mathrm{e}_R} = 2Y_L = -2 \tag{1.25}$$

Collecting the neutrino terms in the Lagrangian similar to eqn. (1.19) and using $Q_\nu = 0$, we have

$$Y_{\nu_R} = 0 \tag{1.26}$$

We can repeat the above calculation for other leptons and quarks, and there is a generalized relation:

$$Q = \frac{Y}{2} + T_3 \tag{1.27}$$

Where Q, Y, $T_3 = \frac{1}{2}\sigma_3$ here are operators with eigenvalues being the charge number, weak hypercharge and third component of the isospin $I_{\mathrm{w}}^{(3)}$, but they also can refer to the values for the corresponding fermion depending on the context. We can for example see that this relation is true for both the electron and neutrino.

Now we turn to the terms coupling to the Z_μ field, which corresponds to what we call the neutral current interaction. Collecting the electron terms from the Lagrangian and substituting eqn.(1.23) and Y_L, Y_{e_R} gives:

$$\begin{aligned} & Z_\mu \left\{ \left[\frac{g_w}{2} \cos\theta_w + \frac{g'}{2} Y_L \sin\theta_w \right] \bar{e}_L \gamma^\mu e_L + \frac{g'}{2} Y_{\mathrm{e}_R} \sin\theta_w \bar{e}_R \gamma^\mu e_R \right\} \\ = & Z_\mu \left\{ \left[\frac{e}{\cos\theta_w \sin\theta_w} \left(\frac{1}{2} - \sin^2\theta_w \right) \right] \bar{e}_L \gamma^\mu e_L + \frac{e}{\cos\theta_w \sin\theta_w} \left(-\sin^2\theta_w \right) \bar{e}_R \gamma^\mu e_R \right\} \end{aligned} \tag{1.28}$$

and both the left- and right-handed electron terms can be collectively written with:

$$-Z_\mu \frac{e}{\sin\theta_w \cos\theta_w} \left(T_3 - Q \sin^2\theta_w \right) \left(\bar{e}_L \gamma^\mu e_L + \bar{e}_R \gamma^\mu e_R \right) \tag{1.29}$$

These interaction terms also apply to a general fundamental fermion:

$$\begin{aligned} & - Z_\mu \frac{e}{\sin\theta_w \cos\theta_w} \left(T_3 - Q \sin^2\theta_w \right) \left(\bar{f}_L \gamma^\mu f_L + \bar{f}_R \gamma^\mu f_R \right) \\ & = -Z_\mu \frac{e}{\sin\theta_w \cos\theta_w} \left(c_L \bar{f}_L \gamma^\mu f_L + c_R \bar{f}_R \gamma^\mu f_R \right) \\ & = -Z_\mu \frac{1}{2} \frac{e}{\sin\theta_w \cos\theta_w} \bar{f} \left(c_V \gamma^\mu - c_A \gamma^\mu \gamma^5 \right) f \end{aligned} \tag{1.30}$$

with

$$\begin{aligned} c_L &= T_3 - Q\sin^2\theta_w, \\ c_R &= -Q\sin^2\theta_w, \\ c_V &= T_3 - 2Q\sin^2\theta_w, \\ c_A &= T_3 \end{aligned} \tag{1.31}$$

and we have used the fact that

$$\begin{aligned} \bar{f}_L\gamma^\mu f_L &= \bar{f}\gamma^\mu\frac{1}{2}\left(1-\gamma^5\right)f \equiv \bar{f}\gamma^\mu P_L f \\ \bar{f}_R\gamma^\mu f_R &= \bar{f}\gamma^\mu\frac{1}{2}\left(1+\gamma^5\right)f \equiv \bar{f}\gamma^\mu P_R f \end{aligned} \tag{1.32}$$

P_L, P_R are the left- and right-handed projection operators satisfying:

$$\begin{aligned} P_L^2 &= P_L, \\ P_R^2 &= P_R, \\ P_L + P_R &= 1, \\ P_L P_R &= 0 \end{aligned} \tag{1.33}$$

Clearly parity is violated in neutral current interaction with the Z boson, and we have organized the couplings into the vector and axial-vector parts corresponding to c_V and C_A. The experimentally measured value of $\sin^2\theta_w$ is 0.23149 (collider average) [18]. A summary of the Q, $I_{\rm w}^{(3)}$ and Y assignments of the fundamental fermions and their couplings to the Z boson is summarized in Table 1.2. The Feynman rule associated with the Z and fermion interaction vertex is

$$-i\frac{1}{2}\frac{e}{\sin\theta_w\cos\theta_w}\gamma^\mu\left[c_V - c_A\gamma^5\right] \tag{1.34}$$

The remaining terms in the Lagrangian corresponds to charged current interaction with W^+ and W^- in the SM:

$$-\frac{g_{\rm w}}{\sqrt{2}}\left(\bar{\nu}_L\gamma^\mu e_L W_\mu^+ + \bar{e}_L\gamma^\mu\nu_L W_\mu^-\right) \tag{1.35}$$

Table 1.2: Summary of the Q, $I_{\mathrm{w}}^{(3)}$ and Y assignments of the fundamental fermions and their couplings to the Z boson from Ref. [9], assuming $\sin^2\theta_w = 0.23146$, which causes negligible difference compared with using the value $\sin^2\theta_w = 0.23149$ [18] cited above.

fermion	Q	$I_{\mathrm{W}}^{(3)}$	Y_L	Y_R	c_L	c_R	c_V	c_A
$\nu_{\mathrm{e}}, \nu_\mu, \nu_\tau$	0	$+\frac{1}{2}$	-1	0	$+\frac{1}{2}$	0	$+\frac{1}{2}$	$+\frac{1}{2}$
$\mathrm{e}^-, \mu^-, \tau^-$	-1	$-\frac{1}{2}$	-1	-2	-0.27	+0.23	-0.04	$-\frac{1}{2}$
$\mathrm{u}, \mathrm{c}, \mathrm{t}$	$+\frac{2}{3}$	$+\frac{1}{2}$	$+\frac{1}{3}$	$+\frac{4}{3}$	+0.35	-0.15	+0.19	$+\frac{1}{2}$
$\mathrm{d}, \mathrm{s}, \mathrm{b}$	$-\frac{1}{3}$	$-\frac{1}{2}$	$+\frac{1}{3}$	$-\frac{2}{3}$	-0.42	+0.08	-0.35	$-\frac{1}{2}$

where we see the W^+ and W^- only couples to left-handed chiral states, which is maximally violating parity. Note that the right-handed neutrino is not coupled to the photon and W, Z bosons, and has not yet been observed. The Feynman rule associated with this vertex is

$$-i\frac{g_w}{\sqrt{2}}\frac{1}{2}\gamma^\mu\left(1-\gamma^5\right) \tag{1.36}$$

From the interaction terms in eqn. (1.36), we see that one entry in the SU(2) fermion doublet can be transformed into the other by interaction with a charged force carrier (the W^+ and W^- bosons). Also from the interaction terms in eqn. (1.21) and (1.30), particle and the corresponding antiparticle can annihilate into a neutral gauge boson (photon or Z boson). The corresponding Feynman diagrams are shown in Fig. 1.1. The interaction strength is proportional to e (or g' and g_w), the constants C_A and C_V, and also affected by the propagator as in eqn. (1.146), which is associated with the gauge boson and has mass terms in the denominator for massive gauge bosons. Note that particle-antiparticle annihilation into photon cannot be achieved via a single vertex in Fig. 1.1 (right) since this violates four-momentum conservation, but can be achieved for example with a diagram involving two such vertices. Fig. 1.1

(right) also implies another diagram with the same structure showing that a particle can emit or absorb a neutral force carrier without being transformed.

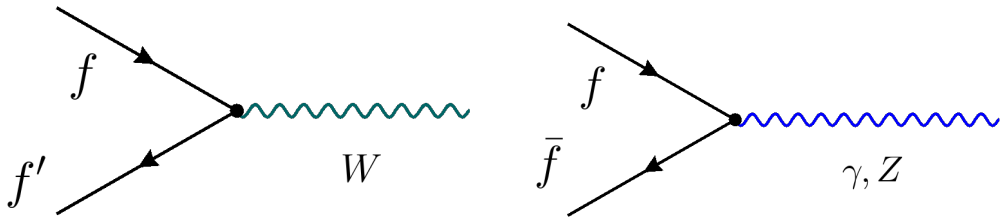

Figure 1.1: Feynman diagrams showing interactions via the charged and neutral force carriers.

We also need to include the kinematic terms of the gauge fields in the Lagrangian, which is given by:

$$\mathscr{L}_{\mathrm{W,B}} = -\frac{1}{4} W^i_{\mu\nu} W^{i\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} \tag{1.37}$$

where we also sum over the index i, and

$$\begin{aligned} W^i_{\mu\nu} &= \partial_\mu W^i_\nu - \partial_\nu W^i_\mu - g_w \varepsilon^{ijk} W^j_\mu W^k_\nu \\ B_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu \end{aligned} \tag{1.38}$$

The GSW model reproduces the couplings in QED and electroweak interactions. However, mass terms for fundamental fermions and gauge bosons are not allowed in the Lagrangian under the requirement of gauge invariance. Consider a fermion mass term

$$\begin{aligned} m\bar{f}f &= m\bar{f}\left(P_L + P_R\right) f \\ &= m\bar{f}P_L P_L f + m\bar{f}P_R P_R f \\ &= m\left(\bar{f}_R f_L + \bar{f}_L f_R\right) \end{aligned} \tag{1.39}$$

Clearly this term is not invariant under SU(2). Meanwhile, a mass term for a gauge field such as

$$\frac{1}{2} m_B^2 B^\mu B_\mu \tag{1.40}$$

is clearly not gauge invariant considering how the gauge fields transform. Therefore these mass terms are not allowed in the Lagrangian. We know from experiments that the fundamental fermions and the W and Z bosons have nonzero masses. To resolve this contradiction, the masses are introduced through the BEH mechanism.

1.3.2 The BEH Mechanism in the Standard Model

In the GSW model, we can introduce an isospin doublet of complex scale fields:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \tag{1.41}$$

which we call the Higgs doublet. The lower component is neutral in electric charge, and the upper component has $Q = +1$. The Lagrangian for this doublet is given by

$$\mathcal{L}_\phi = (\partial_\mu \phi)^\dagger (\partial^\mu \phi) - V(\phi) \tag{1.42}$$

with the Higgs potential

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda \left(\phi^\dagger \phi\right)^2 \tag{1.43}$$

This is the minimal Higgs model that can generate the masses properly in the SM. If we set $\mu^2 < 0$, the set of minima for this potential satisfies:

$$\phi^\dagger \phi = \frac{1}{2}\left(\phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2\right) = -\frac{\mu^2}{2\lambda} \equiv \frac{v^2}{2} \tag{1.44}$$

To make the Lagrangian invariant under local $\mathrm{U}(1)_Y \times \mathrm{SU}(2)_L$, we can again replace the derivative with the covariant derivative:

$$\partial_\mu \to D_\mu = \partial_\mu + i g_w \mathbf{T} \cdot \mathbf{W}_\mu + i g' \frac{Y}{2} B_\mu \tag{1.45}$$

Using eqn. (1.27) and substituting the $I_{\mathrm{w}}^{(3)}$ and Q for the lower component, we have

$$Y_{\mathrm{Higgs}} = 1 \tag{1.46}$$

for the doublet. We are free to choose our minimum of the potential energy (vacuum) to be located at

$$\phi_0 = \langle 0|\phi|0\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\v\end{pmatrix} \tag{1.47}$$

and expand the field around this minimum

$$\phi(x) = \frac{1}{\sqrt{2}}\begin{pmatrix}\phi_1(x) + i\phi_2(x)\\v + \eta(x) + i\phi_4(x)\end{pmatrix} \tag{1.48}$$

ϕ_0 is called the vacuum expectation value (vev). The choice of vacuum is not invariant under $\mathrm{U}(1)_Y \times \mathrm{SU}(2)_L$, and we call this "spontaneous symmetry breaking". Note that making this expansion hides the underlying $\mathrm{U}(1)_Y \times \mathrm{SU}(2)_L$ symmetry of the Lagrangian, but doesn't remove it. This choice of vacuum is consistent with the real world observation, because we can see directly that the ϕ_0 is still invariant under a subgroup of $\mathrm{U}(1)_Y \times \mathrm{SU}(2)_L$:

$$\phi_0 \to \phi' = \exp\left[i\alpha(x)Q\right]\phi_0 = \exp\left[i\alpha(x)\left(\frac{Y}{2} + T_3\right)\right]\phi_0 = \phi_0 \tag{1.49}$$

where Q can be thought of as a linear combination of the generators of $\mathrm{U}(1)_Y \times \mathrm{SU}(2)_L$ (T_3 and Y). We can recognize this subgroup to be the electromagnetic $\mathrm{U}(1)_{EM}$. Therefore the apparent form of the Lagrangian after the expansion still respects $\mathrm{U}(1)_{EM}$ symmetry and the photon remains massless as we would expect.

Since the Lagrangian itself is invariant under $\mathrm{U}(1)_Y \times \mathrm{SU}(2)_L$, we are free to choose a gauge such that

$$\phi(x) = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\v + h(x)\end{pmatrix} \tag{1.50}$$

this gauge is called the unitary gauge. We now first look at the mass terms of the

gauge fields by substituting the vev (disregarding $V(\phi_0)$ which is just a constant):

$$\begin{aligned}
&\left[\left(D_\mu\phi\right)^\dagger\left(D^\mu\phi\right)\right](\phi_0) \\
&= g_w^2\frac{v^2}{8}\begin{pmatrix} 0 & 1 \end{pmatrix}\begin{pmatrix} \frac{g'}{g_w}B_\mu + W_\mu^3 & W_\mu^1 - iW_\mu^2 \\ W_\mu^1 + iW_\mu^2 & \frac{g'}{g_w}B_\mu - W_\mu^3 \end{pmatrix} g^{\mu\nu}\begin{pmatrix} \frac{g'}{g_w}B_\nu + W_\nu^3 & W_\nu^1 - iW_\nu^2 \\ W_\nu^1 + iW_\nu^2 & \frac{g'}{g_w}B_\nu - W_\nu^3 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix} \\
&= g_w^2\frac{v^2}{8}\left[\left(W_\mu^1\right)^2 + \left(W_\mu^2\right)^2 + \left(\frac{g'}{g_w}B_\mu - W_\mu^3\right)^2\right].
\end{aligned} \tag{1.51}$$

We can already recognize from eqn. (1.15) and (1.16) that

$$\frac{g'}{g_w}B_\mu - W_\mu^3 = -Z_\mu/\cos\theta_w \tag{1.52}$$

therefore we can read off the Z mass from the mass term $\frac{1}{2}m_Z^2 Z_\mu^2$

$$m_Z = \frac{1}{2\cos\theta_w}g_w v \tag{1.53}$$

Similarly by using eqn. (1.13) we can recognize

$$g_w^2\frac{v^2}{8}\left[\left(W_\mu^1\right)^2 + \left(W_\mu^2\right)^2\right] = g_w^2\frac{v^2}{8}\left[2W_\mu^+W^{-\mu}\right] = m_W^2 W_\mu^+ W^{-\mu} \tag{1.54}$$

giving

$$m_W = \frac{v}{2}g_w \tag{1.55}$$

One can therefore determine the value of v from m_W and g_W. We have the current value $v = 246.22\,\text{GeV}$ according to the review titled "Electroweak model and constraints on new physics" in Ref. [18].

1.3.3 Full Gauge Sector of the SM

Combining with the Lagrangian terms in eqn. (1.37) and expanding in terms of A_μ, Z_μ, $W_\mu^\pm$ and the field strength tensors

$$\begin{aligned}
Z_{\mu\nu} &= \partial_\mu Z_\nu - \partial_\nu Z_\mu, \\
F_{\mu\nu} &= \partial_\mu A_\nu - \partial_\nu A_\mu
\end{aligned} \tag{1.56}$$

we have the part of Lagrangian containing the gauge fields given by (see [11]):

$$\begin{aligned}
\mathcal{L}_{\text{gauge}} =& -\frac{1}{4}F_{\mu\nu}^2 - \frac{1}{4}Z_{\mu\nu}^2 - \frac{1}{2}\left(\partial_\mu W_\nu^+ - \partial_\nu W_\mu^+\right)\left(\partial^\mu W^{-\nu} - \partial^\nu W^{-\mu}\right) \\
&+ \frac{1}{2}m_Z^2 Z^\mu Z_\mu + m_W^2 W_\mu^+ W^{-\mu} \\
&+ ie\cot\theta_w\left[Z^{\mu\nu}W_\mu^+ W_\nu^- - \left(\partial_\mu W_\nu^+ - \partial_\nu W_\mu^+\right)Z^\mu W^{-\nu} + \left(\partial_\mu W_\nu^- - \partial_\nu W_\mu^-\right)Z^\mu W^{+\nu}\right] \\
&+ ie\left[F^{\mu\nu}W_\mu^+ W_\nu^- - \left(\partial_\mu W_\nu^+ - \partial_\nu W_\mu^+\right)A^\mu W^{-\nu} + \left(\partial_\mu W_\nu^- - \partial_\nu W_\mu^-\right)A^\mu W^{+\nu}\right] \\
&+ \frac{1}{2}\frac{e^2}{\sin^2\theta_w}\left(W^{+\mu}W_\mu^+ W^{-\nu}W_\nu^- - W^{+\mu}W_\mu^- W^{+\nu}W_\nu^-\right) \\
&+ e^2\left(A^\mu W_\mu^+ A^\nu W_\nu^- - A_\mu^2 W^{+\nu}W_\nu^-\right) + e^2\cot^2\theta_w\left(Z^\mu W_\mu^+ Z^\nu W_\nu^- - Z_\mu^2 W^{+\nu}W_\nu^-\right) \\
&+ e^2\cot\theta_w\left(W_\mu^+ W_\nu^- A^\mu Z^\nu + W_\mu^- W_\nu^+ A^\mu Z^\nu - 2W^{+\mu}W_\mu^- A^\nu Z_\nu\right)
\end{aligned} \tag{1.57}$$

The triple and quartic interaction vertex factors can be read off from this Lagrangian. Fig. 1.2 shows two example vertices (drawing W^+ as particle and W^- as antiparticle), with the corresponding vertex factors in the Feynman rule given by:

$$\begin{aligned}
W^+W^-Z\text{:}\ & -ie\cot\theta_w\left[g^{\mu\nu}\left(p_1-p_2\right)^\lambda + g^{\nu\lambda}\left(p_2-p_3\right)^\mu + g^{\lambda\mu}\left(p_3-p_1\right)^\nu\right] \\
W^+W^-ZZ\text{:}\ & ie^2\cot^2\theta_w\left[g^{\alpha\mu}g^{\beta\nu} + g^{\alpha\nu}g^{\beta\mu} - 2g^{\alpha\beta}g^{\mu\nu}\right]
\end{aligned} \tag{1.58}$$

Other vertices include the ones replacing one or more of the Z boson in the shown diagrams with photon, and also the $W^+W^-W^+W^-$ vertex.

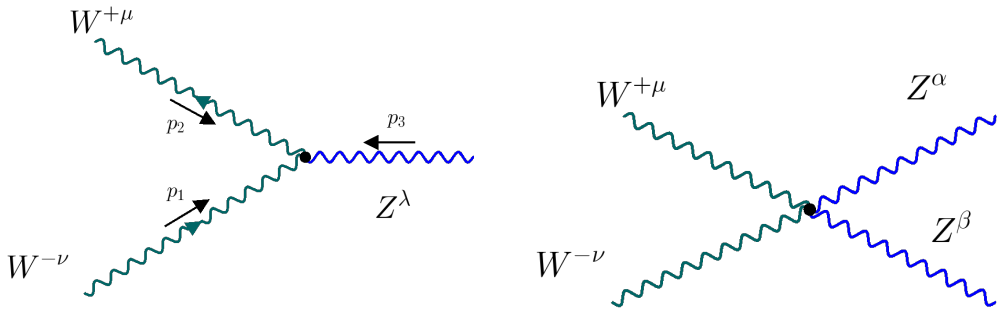

Figure 1.2: Example Feynman diagrams showing triple and quartic gauge boson interactions.

Now we add back the h field, which is known as the Higgs boson, and obtain the Lagrangian terms involving h (see [11]):

$$\begin{aligned}\mathcal{L}_{\text{Higgs}} =& -\frac{1}{2}h\left(\partial_\mu\partial^\mu + m_h^2\right)h - g_w\frac{m_h^2}{4m_W}h^3 - \frac{g^2}{32}\frac{m_h^2}{m_W^2}h^4 \\ &+ 2\frac{h}{v}\left(m_W^2 W_\mu^+ W^{-\mu} + \frac{1}{2}m_Z^2 Z_\mu Z^\mu\right) + \left(\frac{h}{v}\right)^2\left(m_W^2 W_\mu^+ W^{-\mu} + \frac{1}{2}m_Z^2 Z_\mu Z^\mu\right)\end{aligned} \tag{1.59}$$

where

$$m_h = \sqrt{-2\mu^2} = \sqrt{2\lambda v^2} \tag{1.60}$$

We can read off the coupling between the Higgs boson and the W or Z boson from the Lagrangian. The trilinear terms correspond to the Feynman diagrams in Fig. 1.3 with vertex given by:

$$\text{HWW: } i\frac{e}{\sin\theta_w}m_W g^{\mu\nu}, \quad \text{HZZ: } i\frac{e}{\sin\theta_w\cos\theta_w}m_Z g^{\mu\nu} \tag{1.61}$$

and we see the coupling strength is proportional to the gauge boson mass.

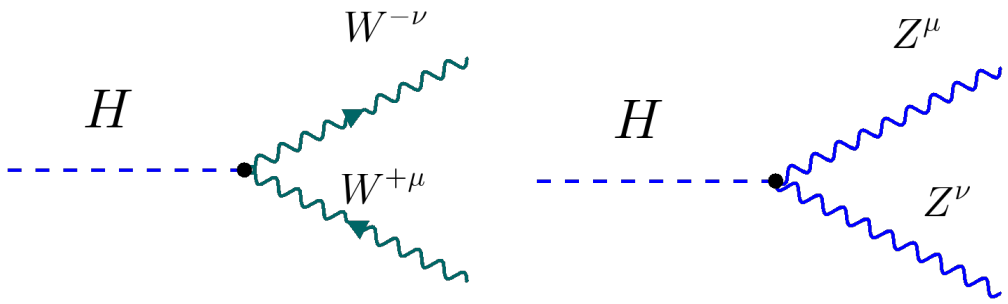

Figure 1.3: Feynman diagrams showing trilinear interaction vertices between the Higgs boson and the Z and W bosons.

1.3.4 Fermion Masses

The fermion masses can be generated by introducing a Lagrangian term like

$$\mathcal{L}_{\text{YukDn}} = -y\bar{L}\phi e_R + h.c. \tag{1.62}$$

where h.c. stands for Hermitian conjugate, and the constant y is known as the Yukawa coupling. After symmetry breaking, substituting eqn. (1.50), we have

$$\mathcal{L}_{\text{YukDn}_\text{e}} = -\frac{y_\text{e}}{\sqrt{2}} v \left(\bar{\text{e}}_L \text{e}_R + \bar{\text{e}}_R \text{e}_L\right) - \frac{y_\text{e}}{\sqrt{2}} h \left(\bar{\text{e}}_L \text{e}_R + \bar{\text{e}}_R \text{e}_L\right) \tag{1.63}$$

which generates a mass term of $-m_e \left(\bar{e}_L e_R + \bar{e}_R e_L\right)$ with $m_e = \frac{y_\text{e}}{\sqrt{2}} v$, and we have interaction vertex between Higgs boson and fermion as shown in Fig. 1.4, with coupling strength proportional to m_f. The Yukawa coupling value is not predicted by the Higgs mechanism, but can be chosen to be consistent with the electron mass m_e given the known value of $v \approx 246\,\text{GeV}$. Since the Yukawa coupling also appears in the coupling between Higgs boson and fermions, this theory of the Higgs mechanism and fermion masses can be put into tests. Interestingly, for the top quark, the Yukawa coupling is almost exactly unity.

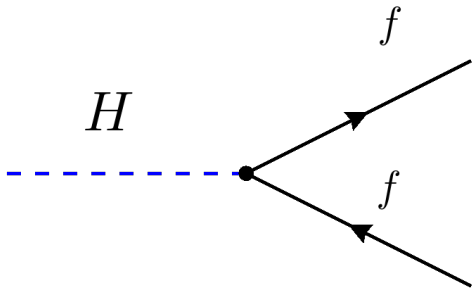

Figure 1.4: Feynman diagrams showing trilinear interaction vertices between the Higgs boson and the fermion. The fermion is in mass basis.

$\mathcal{L}_{\text{YukDn}}$ can be used to generate mass for electron and other charged leptons, as well as the down-type quarks (d, s, b). To generate mass terms for the fundamental fermions at the upper component of the SU(2) doublet, we can replace ϕ with

$$\phi_c = i\sigma_2 \phi^* = \begin{pmatrix} \phi^{0*} \\ -\phi^- \end{pmatrix} \tag{1.64}$$

which has hypercharge -1, and we can verify that $\bar{L}\phi_c$ is SU(2) invariant. Then a term like $-y\bar{L}\phi_c \nu_R$ can give masses to neutrino and up-type quarks. The neutrino can

obtain mass through other possible mechanisms, and for a more detailed discussion for neutrino masses one can for example refer to Ref. [11].

Now we focus on the quark sector. For quarks we can write the corresponding terms as

$$\mathcal{L}_{\text{quarks}} = -y_d^{ij}\bar{L}^i\phi d_R^j - y_u^{ij}\bar{L}^i\phi_c u_R^j + \text{ h.c.} \tag{1.65}$$

where the i, j indices label the three quark generations, and for example $u_R^j = (u_R, c_R, t_R)$. The two y^{ij} matrices are 3×3 matrices of the Yukawa coupling coefficients. We have been writing the quarks states in the flavor basis. To diagonalize the two matrices y_d and y_u and switch to the mass basis, we can perform a singular value decomposition:

$$y_d = U_d M_d K_d^\dagger, \quad y_u = U_u M_u K_u^\dagger \tag{1.66}$$

where the two U matrices and two K matrices are unitary matrices, and the two M matrices are diagonal matrices with nonnegative real entries. This decomposition is always possible (see for example Ref. [43]). If we redo the calculation in Section 1.3.1 for the quarks (using values in Table 1.2), we have the Lagrangian for the quarks in the flavor basis is given by

$$\begin{aligned}
\mathcal{L}_{\text{flavor-basis}} = &\begin{pmatrix} \bar{u}_L & \bar{d}_L \end{pmatrix}^i \left[i\partial_\mu\gamma^\mu - \gamma^\mu \begin{pmatrix} \frac{g'}{6}B_\mu + \frac{g_w}{2}W_\mu^3 & \frac{g_w}{\sqrt{2}}W_\mu^+ \\ \frac{g_w}{\sqrt{2}}W_\mu^- & \frac{g'}{6}B_\mu - \frac{g_w}{2}W_\mu^3 \end{pmatrix} \right] \begin{pmatrix} u_L \\ d_L \end{pmatrix}^i \\
&+ \bar{u}_R^i \left(i\partial_\mu\gamma^\mu - g'\frac{2}{3}\gamma^\mu B_\mu \right) u_R^i + \bar{d}_R^i \left(i\partial_\mu\gamma^\mu + g'\frac{1}{3}\gamma^\mu B_\mu \right) d_R^i \\
&- \frac{v}{\sqrt{2}} \left[\bar{d}_L^i \left(U_d M_d K_d^\dagger \right)^{ij} d_R^j + \bar{u}_L^i \left(U_u M_u K_u^\dagger \right)^{ij} u_R^j + \text{ h.c. } \right] \\
&+ \text{Higgs interaction terms}
\end{aligned} \tag{1.67}$$

If we switch to the mass basis (u', d') by

$$d_R^i = K_d^{ij} d_R'^j, \quad u_R^i = K_u^{ij} u_R'^j, \quad d_L^i = U_d^{ij} d_L'^j, \quad u_L^i = U_u^{ij} u_L'^j \tag{1.68}$$

the mass terms become

$$-m_d^j \left(\bar{d}_L'^j d_R'^j + \bar{d}_R'^j d_L'^j\right) - m_u^j \left(\bar{u}_L'^j u_R'^j + \bar{u}_R'^j u_L'^j\right) \tag{1.69}$$

With the masses being the corresponding diagonal entries of the matrices $\frac{v}{\sqrt{2}}M$. In the SU(2) and U(1) part of the Lagrangian, the terms containing the combination $\bar{u}^i ... u^i$ and $\bar{d}^i ... d^i$ still have the same form since U and K are unitary, and therefore we just replace u, d with u' and d' in these terms. On the other hand, the terms involving W^+ and W^- mix the up- and down-type quarks, and they become

$$-\frac{g_w}{\sqrt{2}} \left[W_\mu^+ \bar{u}_L'^i \gamma^\mu (V)^{ij} d_L'^j + W_\mu^- \bar{d}_L'^i \gamma^\mu \left(V^\dagger\right)^{ij} u_L'^j\right] \tag{1.70}$$

where

$$V \equiv U_u^\dagger U_d = \begin{pmatrix} V_{11} & V_{12} & V_{13} \\ V_{21} & V_{22} & V_{23} \\ V_{31} & V_{32} & V_{33} \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \tag{1.71}$$

is known as the Cabibbo-Kobayashi-Maskawa (CKM) matrix, which is non-diagonal, and the mixing in eqn. (1.70) means that quarks in higher generation can decay to lower generation via the W boson.

1.4 Quantum Chromodynamics

1.4.1 Gluons, Color Confinement and Hadronization

In QCD, only quarks have a type of conserved charge called "color". The three color states in QCD can be represented by

$$r = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad g = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad b = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \tag{1.72}$$

There are also correspondingly 3 anticolor states $\bar{r}$, $\bar{g}$, $\bar{b}$. To introduce the QCD gauge field, we require the Lagrangian to be invariant under SU(3) local gauge transformation acting on the color space:

$$\psi(x) \to \psi'(x) = \exp\left[ig_S \boldsymbol{\alpha}(x) \cdot \hat{\mathbf{T}}\right] \psi(x) \tag{1.73}$$

where $\hat{\mathbf{T}} = \{T^a\}$ are the eight generators of SU(3) given by Gell-Mann matrices λ^a:

$$T^a = \frac{1}{2}\lambda^a \quad (a = 1, \ldots, 8) \tag{1.74}$$

which acts on the color states, and results in zero for fields that do not carry color charge. To preserve gauge invariance, we extend the covariant derivative in the previous section by:

$$D_\mu = \partial_\mu + ig'\frac{Y}{2}B_\mu + ig_w\frac{\sigma^i}{2}W_\mu^i + ig_S\frac{\lambda^a}{2}G_\mu^a \tag{1.75}$$

and the new gauge field G_μ^a transforms according to

$$G_\mu^k \to G_\mu^{k'} = G_\mu^k - \partial_\mu \alpha^k - g_S f_{ijk}\alpha^i G_\mu^j \tag{1.76}$$

Here f_{ijk} is the structure constant of SU(3) defined by

$$[\lambda_i, \lambda_j] = 2if_{ijk}\lambda_k \tag{1.77}$$

This results in an additional Lagrangian term for quarks:

$$-g_S\bar{q}\frac{\lambda^a}{2}G^a_\mu\gamma^\mu q \tag{1.78}$$

where q stands for the quark field and the quark carries a color factor in all the Lagrangian terms introduced so far. The Feynman rule associated with this QCD vertex is

$$c_j^\dagger\left\{-\frac{1}{2}ig_S\lambda^a\gamma^\mu\right\}c_i = -\frac{1}{2}ig_S\lambda^a_{ji}\gamma^\mu \tag{1.79}$$

where c_j and c_i are the color states associated with the quarks, with r, g, b states labeled by $i, j = 1, 2, 3$. The G^a_μ field corresponds to gluons which mediates the QCD interactions. Gluons have to carry both color charge and anticolor charge to conserve colors. There are eight gluon states forming a color octet:

$$\begin{aligned}
|1\rangle &= (r\bar{g} + g\bar{r})/\sqrt{2} \qquad & |5\rangle &= -i(r\bar{b} - b\bar{r})/\sqrt{2} \\
|2\rangle &= -i(r\bar{g} - g\bar{r})/\sqrt{2} & |6\rangle &= (g\bar{b} + b\bar{g})/\sqrt{2} \\
|3\rangle &= (r\bar{r} - g\bar{g})/\sqrt{2} & |7\rangle &= -i(g\bar{b} - b\bar{g})/\sqrt{2} \\
|4\rangle &= (r\bar{b} + b\bar{r})/\sqrt{2} & |8\rangle &= (r\bar{r} + g\bar{g} - 2b\bar{b})/\sqrt{6}
\end{aligned} \tag{1.80}$$

Corresponding to the eight Gell-Mann matrices:

$$\lambda_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \lambda_4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \lambda_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \lambda_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$
$$\lambda_2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \lambda_5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}, \lambda_7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \lambda_8 = \frac{1}{\sqrt{3}}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix} \tag{1.81}$$

One more possibility is a color singlet state

$$|9\rangle = (r\bar{r} + b\bar{b} + g\bar{g})/\sqrt{3} \tag{1.82}$$

but it is believed that there does not exist simple, non-composite color singlet states according to experimental observation. The kinetic term for G^a_μ is given by

$$-\frac{1}{4}G^a_{\mu\nu}G^{a\mu\nu} \tag{1.83}$$

with the field strength tensor given by

$$G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu - g_S f^{abc} G^b_\mu G^c_\nu \tag{1.84}$$

Therefore the Lagrangian contain terms corresponding to triple and quartic self-interaction of the gluon, as shown in Fig 1.5, and one can refer to Ref. [33] for the Feynman rules.

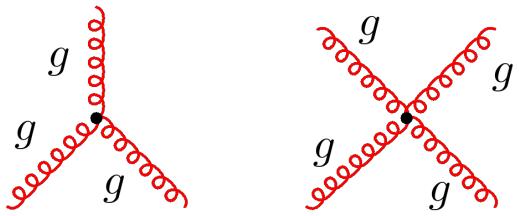

Figure 1.5: Feynman diagrams showing the triple and quartic self-interaction of the gluon.

Currently no free quarks and gluons have been observed in experiments. This is explained by a hypothesis called color confinement, which states that objects with nonzero color charges have to be confined in a color singlet state and cannot be propagating as free particles. This also means that gluons are confined within color singlet objects and cannot propagate for long distance. Currently all confirmed hadronic states correspond to SU(3) color singlet states that can be formed by quarks and antiquarks, $q\bar{q}$ (meson), qqq (baryon), $\bar{q}\bar{q}\bar{q}$ (antibaryon), or combinations of these states such as pentaquark states $\bar{q}qqqq$. There is currently no analytic proof for color confinement, but a qualitative understanding of its likely origin can be obtained, as discussed for example in Ref. [9]. We can think about the interaction between quarks

in terms of the exchange of virtual gluons. Since gluon carry color charges, there are attractive interactions among these exchanged gluons, which confine the color field lines into a tube between the two quarks. The self-interaction of gluons also causes α_S to increase with decreasing energy as discussed in the next section, resulting in large values of α_S at low energy scale. In the non-relativistic limit, the interactions between two quarks or between a quark and an antiquark can be modeled with the potential [9, 44, 45]:

$$V(r) = C\frac{\alpha_S}{r} + \sigma r \tag{1.85}$$

where C is a constant that can be either positive or negative, and $\sigma \sim 1$ GeV/fm experimentally, which corresponds to a very strong force. At relatively large distance, the second term dominates and the potential energy stored between the quarks increases linearly with the separation. Therefore, there will be a very huge energy stored between two separated free quarks, causing quarks to always confine to colorless hadrons.

As a result of color confinement, high energy quarks produced in collisions do not propagate as free particle, but form jets of color singlet particles, and this process is known as hadronization. A qualitative description of the hadronization process is shown in Fig. 1.6, which is for the case of a $q\bar{q}$ pair but the description applies similarly to other cases. As energetic quarks from a hard collision separate, the energy stored between them increases with distance and becomes sufficient for producing new $q\bar{q}$ pairs from vacuum, and breaking the color field into smaller "strings" is energetically favorable. This process can continue until all quarks have sufficiently low energy to combine and form colorless hadrons, which are observed as jets. Currently the precise process of hadronization is not well understood from first principle, but there exists a number of phenomenological models that provide reasonable description of experimental data.

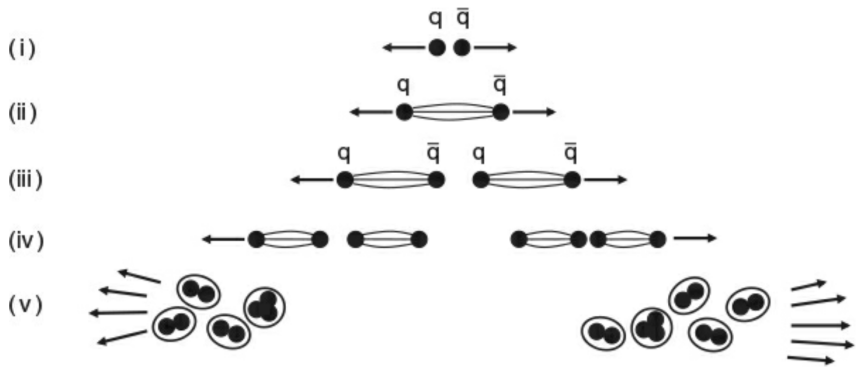

Figure 1.6: Qualitative picture of the steps in the hadronization process. Figure from Ref. [9].

1.4.2 Renormalization and Asymptotic Freedom

There can be ultra-violet (UV) divergence if we do a calculation based on parameters that are not physical, such as using bare masses and bare couplings in the Lagrangian. In a renormalizable theory such as QED and QCD, the UV divergence can be canceled by introducing a finite amount of counterterms, which have an effect of shifting or rescaling the parameters into physical ones. After renormalization, the coupling constants now depend on the energy scale, and we say that the coupling is "running". In QED and QCD the couplings vary with the scale μ_R known as the renormalization scale:

$$\mu_R^2 \frac{\partial \alpha\left(\mu_R^2\right)}{\partial \mu_R^2} = \beta(\alpha) \tag{1.86}$$

where the β-function $\beta(\alpha)$ can be expanded perturbatively as:

$$-\beta(\alpha) = \sum_{n=0}^{\infty} b_n \alpha^{2+n} = \frac{\beta_0}{4\pi}\alpha^2 + \frac{\beta_1}{(4\pi)^2}\alpha^3 + \ldots \tag{1.87}$$

To first order accuracy, this equation has a solution given by

$$\alpha\left(\mu_R^2\right) \equiv \frac{g^2\left(\mu_R^2\right)}{4\pi} = \frac{\alpha\left(Q^2\right)}{1 + \alpha\left(Q^2\right)\frac{\beta_0}{4\pi}\log\frac{\mu_R^2}{Q^2}} \tag{1.88}$$

which means if we measure α at a reference energy scale Q, we can predict it at energy scale μ_R. In QCD, this solution can be written as:

$$\alpha_S\left(\mu_R^2\right) \equiv \frac{g_S^2\left(\mu_R^2\right)}{4\pi} = \frac{1}{\frac{\beta_0}{4\pi}\log\frac{\mu_R^2}{\Lambda_{\mathrm{QCD}}^2}} \tag{1.89}$$

where $\Lambda_{\mathrm{QCD}} \approx 250\,\mathrm{MeV}$ is referred to as the QCD scale, and

$$\beta_0 = \frac{11N_c - 2N_f}{3} \tag{1.90}$$

with N_f the number of quark flavors and N_C the number of colors. Experimentally $\alpha_S\left(m_Z^2\right) = 0.1180 \pm 0.0009$ [18].

β_0 being positive indicates that the strong coupling decreases with increasing energy scale or decreasing distance, and this behavior is known as asymptotic freedom, which is significantly different from QED. The running of coupling has been confirmed experimentally, as shown in Fig. 1.7. The running coupling diverges for $\mu_R \to \Lambda_{\mathrm{QCD}}$, and the perturbation theory no longer applies.

1.5 Proton-Proton Collisions

The data we use for the ZZ analyses come from proton-proton (pp) collisions in the LHC and our signal processes have a pair of protons as the initial state. Therefore, whether our signal processes are of EW or QCD origin (both are studied in our case), its descriptions require understanding the proton structure, and QCD is the starting point of LHC physics studies.

1.5.1 Factorization, Parton Distribution Function and the DGLAP Equations

In this section we will first introduce a series of concepts and expressions, and then discuss about their physics interpretations. pp collisions at high energies can be

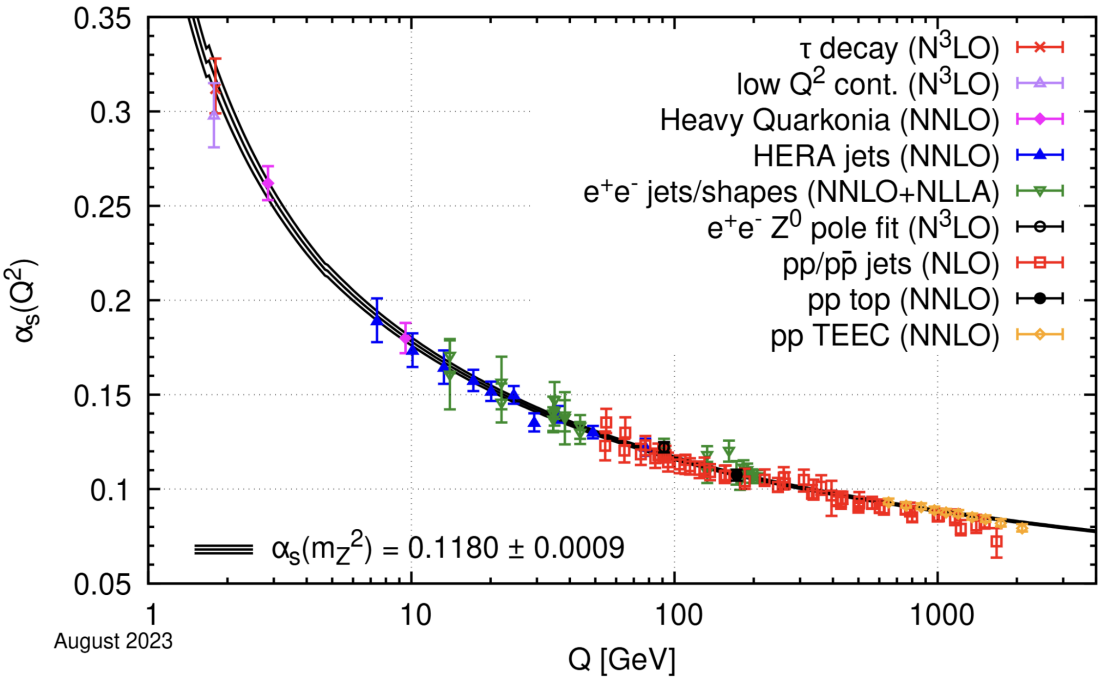

Figure 1.7: Summary of determination of α_S as a function of the energy scale Q compared with the running of the coupling computed at five loops taking as an input the current PDG average, $\alpha_S\left(m_Z^2\right) = 0.1180 \pm 0.0009$. Figure from Ref. [18].

described by the factorization theorem [21]. In general under suitable conditions, the production cross section of n-parton final states from the interaction of two hadrons $h_1 h_2 \to n + X$ can be calculated with the master formula:

$$\sigma = \sum_{a,b} \int_0^1 dx_a dx_b \int f_{a/h_1}\left(x_a, \mu_F^2\right) f_{b/h_2}\left(x_b, \mu_F^2\right) d\hat{\sigma}_{ab\to n}\left(\mu_F, \mu_R\right) \tag{1.91}$$

where $\hat{\sigma}$ is the corresponding partonic cross section, f is the Parton Distribution Function (PDF) that we will discuss in details below, μ_F is the factorization scale at which the PDF is determined, which can be thought of as the scale that separates short- and long-distance physics, and μ_R is the renormalization scale at which the coupling constants are evaluated. In the following discussion, we use variables with the "hat" notation to indicate a partonic quantity (except the regularized splitting function $\hat{P_{ba}}$).

The scales μ_F and μ_R are process independent, and the physics doesn't depend on the choice of these scales, but the choice affects the calculated results and uncertainties due to finite truncation of the perturbative expansion. The scales are typically chosen to be close to the characteristic scales of the relevant processes to avoid unnaturally large logarithms in the perturbation series [12], and the associated theoretical uncertainties are typically evaluated by simultaneously varying the two scales up and down by a factor of two (excluding those combinations where the two scales differ by a factor of four). This is a method to estimate the perturbative uncertainties and account for the residual scale dependence due to the finite truncation. In general we want to choose a range for the scale variations so that the true cross section is expected to be captured within the uncertainties, and there is some dispute on whether two is a sufficiently large factor [12]. We will discuss more on these scale uncertainties in Section 1.5.2, where we talk about the calculation of partonic cross sections at different orders ot accuracy, such as leading order (LO), next-to-leading order (NLO), and next-to-next-to-leading order (NNLO).

In the parton model, there are objects called partons inside the hadron that carries a fraction of the hadron momentum, and can be treated as essentially free during high-energy collisions. The number density of a parton having a fraction x of the hadron momentum is given by the PDF $f_{p/h}(x, Q^2)$, where the label p/h here indicates the type of the parton p and hadron h (typically proton in our discussion), and Q^2 (sometimes represented by the variable t) indicates the energy scale at which the hadron is probed, which was substituted by μ_F^2 in the master formula. We sometimes use simplified notations in the following discussion and ignore the symbol h, p or the arguments in $f_{p/h}(x, Q^2)$. Considering the valence quarks inside a proton, the number of each type of valence quarks is conserved in QED and QCD, which

gives

$$\int dx\,[f_d(x) - f_{\bar{d}}(x)] = 1, \quad \int dx\,[f_u(x) - f_{\bar{u}}(x)] = 2, \quad \int dx\,[f_s(x) - f_{\bar{s}}(x)] = 0 \tag{1.92}$$

Similarly for other remaining quarks. There is no constraint on the number of gluons and therefore on $f_g(x)$. In addition we also require the sum of parton momentum to be equal to the full momentum of the proton, which gives:

$$\sum_j \int dx\,[x f_j(x)] = 1 \tag{1.93}$$

These are known as the sum rules. It turns out that (Ref. [11]) $\int dx\,(f_u(x) + f_d(x))\,x \approx 38\%$ and $\int dx f_g(x) x$ ranges from 35% to 50% depending on Q^2, and the remaining proton momentum is in the sea quarks.

Currently the PDF cannot be calculated from first principle since it is in the non-perturbative regime, but it can be measured at a certain energy scale, and the evolution of the PDF with respect to energy scale is given by the Dokshitser-Gribov-Lipatov-Altarelli-Parisi (DGLAP) equations [46–53]:

$$\begin{aligned} &\frac{\partial}{\partial \log Q^2} \begin{pmatrix} f_{q/h}\,(x, Q^2) \\ f_{g/h}\,(x, Q^2) \end{pmatrix} \\ &\quad = \frac{\alpha_S\,(Q^2)}{2\pi} \int_x^1 \frac{dz}{z} \begin{pmatrix} \hat{P}_{qq}\,(z)\,\hat{P}_{qg}\,(z) \\ \hat{P}_{gq}\,(z)\,\hat{P}_{gg}\,(z) \end{pmatrix} \begin{pmatrix} f_{q/h}\left(\frac{x}{z}, Q^2\right) \\ f_{g/h}\left(\frac{x}{z}, Q^2\right) \end{pmatrix} \end{aligned} \tag{1.94}$$

where the sum over different quark and the corresponding antiquark flavors ($2N_f$ of them) is implicit. Here $\hat{P}_{ba}(z)$ are called the regularized splitting functions, to the

leading order given by:

$$\begin{aligned}
\hat{P}_{qq}^{(0)}(z) =& C_F \left[\frac{1+z^2}{(1-z)_+} + \frac{3}{2}\delta(1-z) \right] \\
\hat{P}_{qg}^{(0)}(z) =& T_R \left[z^2 + (1-z)^2 \right] \\
\hat{P}_{gq}^{(0)}(z) =& C_F \left[\frac{1+(1-z)^2}{z} \right] \\
\hat{P}_{gg}^{(0)}(z) =& 2C_A \left[\frac{z}{(1-z)_+} + \frac{1-z}{z} + z(1-z) \right] \\
& + \frac{11C_A - 4N_f T_R}{6}\delta(1-z)
\end{aligned} \tag{1.95}$$

and the indices of $\hat{P}_{ba}(z)$ corresponds to the parton splitting process $a \to bc$, where the type of c is fixed by a and b. The z in $\hat{P}_{ba}(z)$ represents the energy fraction of b over the parent parton a. The integral of this fraction z is from x to 1 in the DGLAP equation, which is reasonable since the argument x/z in the PDF cannot be larger than 1. We will discuss more on the interpretation of the DGLAP equation in this subsection. C_F and C_A are the Casimir operators of the gauge group in its fundamental and adjoint representations, and in QCD

$$C_F = \frac{N_c^2 - 1}{2N_c} \quad , \quad C_A = N_c \quad \text{and} \quad T_R = \frac{1}{2} \tag{1.96}$$

The plus prescription is defined by imposing the requirement on the integral along with an arbitrary function $f(z)$:

$$\int_0^1 dz \left[g(z) \right]_+ f(z) \equiv \int_0^1 dz g(z) \left[f(z) - f(1) \right] \tag{1.97}$$

so that the divergence of $g(z)$ at $z = 1$ in the integral is canceled provided $f(z)$ is smooth enough, and

$$\frac{1}{[1-z]_+} = \frac{1}{1-z} \text{ for } z \neq 1 \tag{1.98}$$

We will discuss the meaning of the plus prescription below. On the other hand, we also have the unregularized splitting functions $P_{ba}(z)$, which are obtained by simply dropping the plus prescription and the terms proportional to $\delta(1-z)$ in eqn.

(1.95) [28, 31]:

$$\begin{aligned} P_{qq}^{(0)}(z) =& C_F \frac{1+z^2}{1-z} \\ P_{qg}^{(0)}(z) =& T_R \left[z^2 + (1-z)^2\right] \\ P_{gq}^{(0)}(z) =& C_F \left[\frac{1+(1-z)^2}{z}\right] \\ P_{gg}^{(0)}(z) =& 2C_A \left[\frac{z}{1-z} + \frac{1-z}{z} + z(1-z)\right] \end{aligned} \tag{1.99}$$

These unregularized splitting functions represent the probability of emission (splitting) that has momentum almost parallel to the parent parton (collinear region), and are used in place of $\hat{P}_{ba}(z)$ when we discuss about parton shower later in this section and in Chapter 3. For clarity, in the discussion from now on sometimes we also write the splitting functions explicitly as $P_{a,bc}(z)$ for $a \to bc$ splitting.

Now we give a physical interpretation of the DGLAP equation, factorization scale, the splitting functions and the plus prescription, which are closely related to each other. For this purpose, we will need to first talk about soft and collinear singularity cancellation, and why we need to introduce what is known as an infrared cutoff. Let us consider a quark from a proton entering a hard interaction, it can have further emissions of gluons, and the amplitude diverges for soft gluon emission with $E \to 0$ (soft singularity), or for gluons with momentum almost parallel to the emitting quark (relative transverse momentum $k_T \to 0$, collinear singularity). For simplicity, in the following discussion we focus on the $q \to qg$ case but there are also effects from other types of parton splittings. We will discuss about such secondary emissions in Section 1.5.4, where we note that these singularities are not causing concerns in measurements of inclusive final states, due to the the cancellation of divergence between real emission and virtual contributions (Kinoshita-Lee-Nauenberg theorem [54, 55]). On the other hand, the collinear singularity associated with initial states is not canceled between real and virtual contributions (whereas soft singularity does cancel

out), as we are going to see.

For concrete expressions, let us look at the the collinear region, where at leading order the one-gluon emission amplitude squared factorizes into the Born contribution times a factor that is independent of the hard interaction (see Ref. [12, 18, 27]):

$$d\sigma_{\text{hard}+g} \approx d\sigma_{\text{hard}} \frac{dt}{t} dz \frac{d\phi}{2\pi} P_{qq}(\alpha_S, z) \approx d\sigma_{\text{hard}} \frac{dt}{t} dz \frac{d\phi}{2\pi} \frac{\alpha_S}{2\pi} P_{qq}^{(0)}(z) \tag{1.100}$$

Note that we are using the unregularized leading splitting function

$$P_{qq}^{(0)} = C_F \frac{1+z^2}{1-z} \tag{1.101}$$

and $P_{ba}\left(\alpha_S\left(k_T^2\right), z\right)$ incorporates the higher order effects:

$$P_{ba}\left(\alpha_S, z\right) = \frac{\alpha_S}{2\pi} P_{ba}^{(0)}(z) + \left(\frac{\alpha_S}{2\pi}\right)^2 P_{ba}^{(1)}(z) + \left(\frac{\alpha_S}{2\pi}\right)^3 P_{ba}^{(2)}(z) + \ldots \tag{1.102}$$

In the context of leading order effects, we often simply write $P_{ba}^{(0)}$ as P_{ba}. Here the energy scale t in the collinear region is called the hardness, which goes to zero in the collinear limit, and can be defined to be $E^2\theta^2$, or the virtuality of the splitting parton p^2 (which is the invariant mass of the splitting products), or the k_T^2 relative to the parent parton, given by

$$p^2 = 2E^2 z(1-z)(1-\cos\theta) \approx z(1-z)E^2\theta^2, \quad k_T^2 = z^2(1-z)^2 E^2\theta^2 \tag{1.103}$$

These definitions of t are equivalent if the region of small values of z and $1-z$ is not important [18], which means when looking at regions where small t is associated with small θ (collinear) and not soft gluon. Note that $z \to 1$ is associated with soft gluon emission in P_{qq} and P_{gg} whereas $z \to 0$ is associated with soft gluon emission in P_{gq} and P_{gg}. The collinear divergence is explicit in $\frac{dt}{t}$ as $t \to 0$. For simplicity of expressions, if we additionally require that it is also in the soft region ($z \to 1$), using eqn. (1.95), the expressions becomes

$$d\sigma_{\text{hard}+g} \approx d\sigma_{\text{hard}} \frac{dt}{t} dz \frac{d\phi}{2\pi} \frac{\alpha_S}{2\pi} \frac{2C_F}{1-z} \tag{1.104}$$

For final states, as mentioned, we know the divergence in the virtual contribution cancels the real emission, which gives

$$d\sigma_{\text{hard}+virtual} \approx d\sigma_{\text{hard}}(-\frac{dt}{t}dz\frac{d\phi}{2\pi}\frac{\alpha_S}{2\pi}\frac{2C_F}{1-z}) \tag{1.105}$$

However, for initial states, the collinear singularity is not canceled, as seen in the following observation (see Ref. [23]). The parton that goes into the hard interaction could have already gone through soft or collinear emission, but the collinear singularity is not canceled by the virtual contribution, as illustrated in Fig. 1.8, where on the

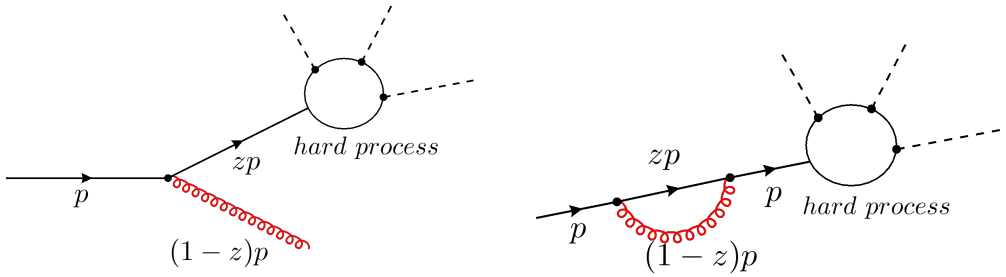

Figure 1.8: Diagrams showing the difference in momentum associated with real and virtual processes when $(1-z)$ is nonzero.

left a quark with original momentum p emits a parton and has remaining momentum zp, and on the right the virtual gluon loop doesn't change the quark momentum. Therefore, the real and virtual contributions are not associated with the same parton momentum that enters the hard interaction. if we write down the momentum dependence of the hard interaction part explicitly for initial states, we have instead

$$d\sigma_{\text{hard}+g} \approx d\sigma_{\text{hard}}(zp)(\frac{dt}{t}dz\frac{d\phi}{2\pi}\frac{\alpha_S}{2\pi}\frac{2C_F}{1-z}) \tag{1.106}$$

and

$$d\sigma_{\text{hard+virtual}} \approx d\sigma_{\text{hard}}(p)(-\frac{dt}{t}dz\frac{d\phi}{2\pi}\frac{\alpha_S}{2\pi}\frac{2C_F}{1-z}) \tag{1.107}$$

and the collinear divergence coming from $\frac{dt}{t}$ in $d\sigma_{\text{hard}+g}$ and $d\sigma_{\text{hard+virtual}}$ does not cancel out when we perform the integration over t:

$$\int_0^{Q^2} \frac{dt}{t} \int dz \frac{d\phi}{2\pi} \frac{\alpha_S}{2\pi} \frac{2C_F}{1-z} (d\sigma_{\text{hard}}(zp) - d\sigma_{\text{hard}}(p)) \tag{1.108}$$

which is a serious concern. On the other hand, the soft singularity does cancel out, since the momentum difference approaches zero for real and virtual contributions as the emission energy approaches zero ($z \to 1$). The case where the gluon emission is soft but not collinear is discussed in Section 1.5.4, where we have a different expression for the factorization factor, and the argument for the cancellation of soft singularity still holds.

To regularize the collinear singularity, we need to impose a lower limit on the parton emission hardness t (e.g. transverse momentum k_T^2), reflecting the fact that we can separate out effect that is below a certain hardness t associated with the probing or hard interaction energy scale Q as the long-distance effect, and in addition, the perturbative description fails at scale close to Λ_{QCD}.

Now we come back to the discussion on the DGLAP equation. Based on the discussion in Ref. [28], we can visualize the relevant processes using the following rough picture, as illustrated in Fig. 1.9: inside a proton before entering a hard interaction, the fluctuation time of the gluon emission and reabsorption for a valence quark scales like k/k_T^2 in the collinear limit [12], using the combination of the uncertainty principle and Lorentz dilation, and is small for large hardness. If the energy scale for a probing hard process is Q_0, with interaction time scale behaving as $1/Q_0$, then for k_T sufficiently larger than Q_0, the emitted gluon is reabsorbed during a short enough time interval without affecting the quark state before the hard interaction, and meanwhile the perturbative calculation is valid. However, for gluon emission with k_T sufficiently smaller than Q_0, the gluon doesn't have enough time to be reabsorbed before the

emitting quark enters the hard interaction, and the gluon permanently carries away part of the quark momentum and therefore affecting the quark state and number density distribution. Furthermore, when k_T is close to or below $\Lambda_{\rm QCD}$, the energy scale becomes too low that the perturbative description is no long valid, due to the running of α_S which should be written as $\alpha_S(k_T^2)$. Therefore, the quark state being probed depends on the probing energy scale Q_0. In other words, the PDF should be written as $f(x, Q_0)$, and the partonic cross section such as $d\sigma_{\rm hard}(p)$ above should be written as $d\sigma_{\rm hard}(p, Q_0)$. We say that the collinear singularity associated with the initial state parton is absorbed into the PDF.

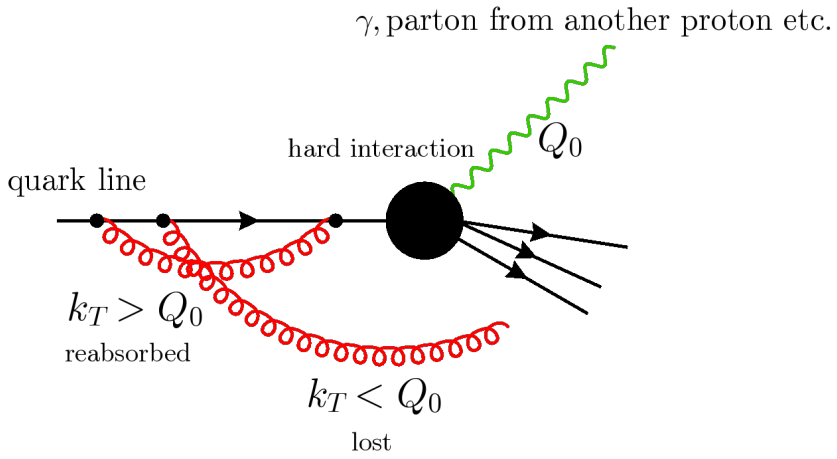

Figure 1.9: Diagram showing how the gluon emission and reabsorption depend on the hard interaction energy scale Q. This is a rough picture and the $k_T > Q_0$ and $k_T < Q_0$ separation is not strict.

Now if we have measured the PDF and α_S at energy scale Q_0, and the hard interaction under consideration is $Q > Q_0$, then we can evolve the PDF and obtain its value at Q by taking into account the additional collinear radiation with hardness between Q_0^2 and Q^2 that doesn't happen fast enough to be reabsorbed before the hard interaction:

$$
\begin{aligned}
f_q\left(x,Q^2\right) =& f_a\left(x,Q_0^2\right)+\int_{Q_0^2}^{Q^2}\frac{dk_T^2}{k_T^2}\int_0^1 dz\int_x^1 dy P_{qq}\left(\alpha_S\left(k_T^2\right),z\right)f_q\left(y,k_T^2\right)\delta\left(x-yz\right)\\
&-\int_{Q_0^2}^{Q^2}\frac{dk_T^2}{k_T^2}\int_0^1 dz\int_0^x dy' P_{qq}\left(\alpha_S\left(k_T^2\right),z\right)f_q\left(x,k_T^2\right)\delta\left(y'-xz\right)\\
=& f_a\left(x,Q_0^2\right)+\int_{Q_0^2}^{Q^2}\frac{dk_T^2}{k_T^2}\int_x^1\frac{dz}{z}P_{qq}\left(\alpha_S\left(k_T^2\right),z\right)f_q\left(x/z,k_T^2\right)\\
&-\int_{Q_0^2}^{Q^2}\frac{dk_T^2}{k_T^2}\int_0^1 dz P_{qq}\left(\alpha_S\left(k_T^2\right),z\right)f_q\left(x,k_T^2\right)
\end{aligned}
\tag{1.109}
$$

This equation can be interpreted as following: the number density of momentum fraction x at scale Q is equal to the number density at scale Q_0 plus two contributions written in the two integrals. The first integral comes from those contributions with momentum fraction $y \geq x$ that has undergone an extra emission with k_T between Q_0 and Q, and with remaining energy fraction z such that $y = x/z$, as constrained by the delta function ($y \geq x$ is indicated in the integration limits for y). Since $x \leq y \leq 1$, the integration limit of z is also between x and 1 after we apply the constraint. The second integral being subtracted comes from those partons that originally have momentum fraction x, and the gluon emissions with k_T between Q_0 and Q are reabsorbed (virtual contributoin), but this is no longer the case when we open the window from Q_0 to Q with higher probing energy scale, and the new momentum fraction reduces to $y' = xz$.

We can now introduce the plus prescription to simplify the notations and combine the two integrals into one. First note that as $z \to 1$, the difference between the first and second integrals disappears and they cancel out, again showing the cancellation of the soft singularity. Let us truncate to leading order $P_{qq}(\alpha_S, z) = \frac{\alpha_S\left(k_T^2\right)}{2\pi}P_{qq}\left(z\right)$. By definition eqn. (1.97), we can write

$$
\left[P_{qq}(z)\right]_+ = \left(\frac{1+z^2}{1-z}\right)_+ \equiv \frac{1+z^2}{1-z} - \delta(1-z)\int_0^1 dy\left(\frac{1+y^2}{1-y}\right) \tag{1.110}
$$

and we can then combine the two integrals:

$$\int_{Q_0^2}^{Q^2} \frac{dk_T^2}{k_T^2} \int_x^1 \frac{dz}{z} \frac{\alpha_S\left(k_T^2\right)}{2\pi} P_{qq}\left(z\right) f_q\left(x/z, k_T^2\right) - \int_{Q_0^2}^{Q^2} \frac{dk_T^2}{k_T^2} \int_0^1 dz \frac{\alpha_S\left(k_T^2\right)}{2\pi} P_{qq}\left(z\right) f_q\left(x, k_T^2\right)$$
$$= \int_{Q_0^2}^{Q^2} \frac{dk_T^2}{k_T^2} \int_x^1 \frac{dz}{z} \left[P_{qq}(z)\right]_+ f_q\left(x/z, k_T^2\right) \tag{1.111}$$

which we can verify by direct substitution of eqn (1.110). We can also obtain the precise expression (see [22])

$$\left[P_{qq}(z)\right]_+ = \left(\frac{1+z^2}{1-z}\right)_+ = \frac{1+z^2}{(1-z)_+} + \frac{3}{2}\delta(1-z) = \hat{P_{qq}}(z) \tag{1.112}$$

which is why we use the regularized equation in the DGLAP equation eqn. (1.94).

The above argument is for gluon emission from quark, but similar calculation can be performed, for example, for gluon splitting into a quark-antiquark pair associated with $P_{qg}(z)$, and we can obtain the general relation taking into account all contributions:

$$f_a\left(x, Q^2\right) = f_a\left(x, Q_0^2\right) + \int_{Q_0^2}^{Q^2} \frac{dk_T^2}{k_T^2} \int_x^1 \frac{dz}{z} \hat{P}_{ba}\left(\alpha_S\left(k_T^2\right), z\right) f_b\left(x/z, k_T^2\right) \tag{1.113}$$

where the sum over index b is implicit. Taking the derivative of eqn. (1.113) with respect to Q^2, we then obtain the DGLAP equation eqn. (1.94). Using the DGLAP equation, we can perform measurements on the PDF and α_S values at a starting scale, and then obtain the PDF at a different scale, for example, by solving the equation using the Mellin transforms (see Refs. [22, 28]).

The factorization scale μ_F is an arbitrary scale that we can substitute into Q_0 to separate which long-distance effects we choose to absorb into the PDF and which effects need to be accounted for through the perturbative calculation. However, it needs to be chosen properly to make the factorization picture and the master formula eqn. (1.91) valid, and will introduce a large logarithm that can spoil the perturbative

description if it is too different from the hard scale Q, as can be seen from the $\frac{dk_T^2}{k_T^2}$ integral.

PDFs have been determined from global fits to a large amount of data (order of 3000 data points for modern PDFs global fits [12]) from a number of different processes from the fixed target experiments, the Tevatron, HERA and the LHC, covering a broad range of x and Q^2. In particular, data from deep inelastic scattering (DIS), Drell-Yan (DY) and jet processes have played a dominant role in the past, and most of the information on the modern PDFs is provided by measurements in DIS experiments data from either HERA-I [56] or the combined HERA-I and HERA-II data [57], which have small uncertainties that are hard to match using the current LHC datasets, and form a very strong constraint for all modern global PDF fits. At the LHC, production of jets, top quarks, and weak bosons provides additional sensitivity to the PDFs and have become increasingly important with the improvement of statistical and systematic uncertainties of the datasets. The accuracy of extrapolating to higher Q^2 using the DGLAP equation depends on the accuracy of the original measurements, accuracy of the evolution codes, and the uncertainty on $\alpha_S(Q^2)$. NLO, NNLO and approximate N^{3}LO (aN^{3}LO) evolution using the DGLAP equation have been achieved [58], and the PDFs have been extracted at LO, NLO, NNLO and even aN^{3}LO from experimental data [32]. As discussed in Ref. [12], the evolution program is expected to extrapolate the PDFs with a reasonable accuracy over the hadron collider kinematic range (a few percent for NLO and NNLO DGLAP, excluding extreme x values). The extrapolation of PDFs to higher Q^2 is required to make predictions for many processes of interest in LHC, and the need for extrapolation will be reduced as more SM processes are included in global PDF fits.

With the above picture we can return to the master formula eqn. (1.91), which is based on the observation that if the hard interaction energy scale Q is high and

the time scale of the interaction characterized by $1/Q$ is much smaller than the life time of a partonic state inside a hadron, we can treat the parton as being stable long before the hard interaction and having a definite momentum fraction x of the hadron h, and therefore the PDF $f(x, \mu_F)$ is universal for different processes independent of the hard interaction. Meanwhile the parton in the hard interaction can be treated as quasi-free. Therefore we can factorize the calculation into a perturbative part associated with quasi-free partons and a non-perturbative part with the PDF. As mentioned in [12], the validity of this factorization has only been proved for the case of DIS and DY production, and the applicability to other processes is only justified by the success of the description and not strict mathematical proof.

1.5.2 Fixed Order Calculation

Perturbative description applies to the hard interaction, and we can calculate the relevant partonic cross sections to different orders of accuracy, such as LO, NLO and NNLO. In general the partonic cross section can be expanded as

$$\hat{\sigma}_{ab} = \alpha_S^k \sum_n \left(\frac{\alpha_S}{\pi}\right)^n \hat{\sigma}_{ab}^{(n)} \tag{1.114}$$

The precise definition of the order of accuracy must be associated with a specific observable. We define LO to be the order at which the calculation gives the first non-trivial contribution to the perturbative expansion of the specific observable, and NLO the first perturbative correction to the same observable and so on. For example, the LO contribution to the $q\bar{q} \to Z$ total cross section has α_S^0 in the coupling, but the LO contribution to $d\sigma/dp_T(Z)$ has α_S^1, since at LO of the total cross section there is no extra emission that causes a deviation of p_T from zero.

With the master formula eqn. (1.91), the LO cross section can be evaluated with

$$\begin{aligned}
&\sigma_n^{(\mathrm{LO})} \equiv \sigma_n^{(\mathrm{Born})} \\
&= \sum_{a,b} \int_0^1 dx_a dx_b f_{a/h_1}\left(x_a, \mu_F\right) f_{b/h_2}\left(x_b, \mu_F\right) \int d\hat{\sigma}_{ab\to n}^{(\mathrm{LO})}\left(\mu_F, \mu_R\right) \\
&= \sum_{a,b} \int_0^1 dx_a dx_b \int d\Phi_n f_{a/h_1}\left(x_a, \mu_F\right) f_{b/h_2}\left(x_b, \mu_F\right) \frac{1}{F} \left|\mathcal{M}_{ab\to n}\right|^2 \left(\Phi_n; \mu_F, \mu_R\right)
\end{aligned} \tag{1.115}$$

Here $\mathcal{M}_{ab\to n}$ is the matrix element which can be evaluated along with Feynman diagrams. F is the Lorentz-invariant flux factor given by

$$F = 4\left[\left(p_a \cdot p_b\right)^2 - p_a^2 p_b^2\right]^{\frac{1}{2}} = 2\hat{s}_{ab} = 2x_a x_b s \tag{1.116}$$

where we have used the approximation $m_a = m_b = 0$, and s is one of the Mandelstam variables

$$s \equiv \left(p_a + p_b\right)^2 = E_{CM}^2 \tag{1.117}$$

The phase space element involving the four-momenta of the final state particles is given by

$$d\Phi_n = \prod_{i=1}^{n} \left[\frac{d^4 p_i}{(2\pi)^4} (2\pi)\delta\left(p_i^2 - m_i^2\right) \Theta\left(p_i^{(0)}\right)\right] (2\pi)^4 \delta^4\left(p_a + p_b - \sum_{i=1}^{n} p_i\right) \tag{1.118}$$

where the terms inside the square bracket is the Lorentz-invariant phase space element.

The cross section at NLO involves corrections to LO from an additional virtual loop and an additional real parton emission:

$$\begin{aligned}
\sigma_n^{(\mathrm{NLO})} &= \sum_{a,b} \int_0^1 dx_a dx_b f_{a/h_1}\left(x_a, \mu_F\right) f_{b/h_2}\left(x_b, \mu_F\right) \int d\hat{\sigma}_{ab\to n}^{(\mathrm{NLO})}\left(\mu_F, \mu_R\right) \\
&= \int d\Phi_{\mathcal{B}} \left[\mathcal{B}_n\left(\Phi_{\mathcal{B}}; \mu_F, \mu_R\right) + \mathcal{V}_n\left(\Phi_{\mathcal{B}}; \mu_F, \mu_R\right)\right] + \int d\Phi_{\mathcal{R}} \mathcal{R}_n\left(\Phi_{\mathcal{R}}; \mu_F, \mu_R\right)
\end{aligned} \tag{1.119}$$

where

$$\begin{aligned}
\mathcal{B}_n\left(\Phi_{\mathcal{B}};\mu_F,\mu_R\right) &= \sum_h^- \left|\mathcal{M}_n^{(b)}\left(\Phi_{\mathcal{B}},h;\mu_F,\mu_R\right)\right|^2 \\
\mathcal{V}_n\left(\Phi_{\mathcal{B}};\mu_F,\mu_R\right) &= \sum_h^- 2\Re\mathfrak{e}\left[\mathcal{M}_n^{(b)}\left(\Phi_{\mathcal{B}},h;\mu_F,\mu_R\right)\mathcal{M}_n^{*(b+2)}\left(\Phi_{\mathcal{B}},h;\mu_F,\mu_R\right)\right] \\
\mathcal{R}_n\left(\Phi_{\mathcal{R}};\mu_F,\mu_R\right) &= \sum_h^- \left|\mathcal{M}_{n+1}^{(b+1)}\left(\Phi_{\mathcal{R}},h;\mu_F,\mu_R\right)\right|^2
\end{aligned} \tag{1.120}$$

and the phase space elements are defined as

$$\begin{aligned}
d\Phi_{\mathcal{B}} &= dx_a dx_b f_{a/h_1}\left(x_a,\mu_F\right) f_{b/h_2}\left(x_b,\mu_F\right)\frac{1}{2\hat{s}_{ab}}d\Phi_n \\
d\Phi_{\mathcal{R}} &= dx_{a'} dx_{b'} f_{a'/h_1}\left(x_{a'},\mu_F\right) f_{b'/h_2}\left(x_{b'},\mu_F\right)\frac{1}{2\hat{s}_{a'b'}}d\Phi_{n+1}
\end{aligned} \tag{1.121}$$

Here the b in $\mathcal{M}_n^{(b)}$ is the perturbative order of the Born contribution, and the summing symbol indicates summing and averaging over the helicity properly. The virtual process has an additional closed loop added to the Born process without changing the external particles, and therefore has the same phase space $\Phi_{\mathcal{B}}$ to integrate over for final states. The NLO virtual contribution to the cross section comes from the interference term between the Born and loop-corrected processes when squaring the matrix element, and this term has the same order in α_S as the real emission matrix element squared. On the other hand, the real emission phase space integration corresponds to $n+1$ final state particles, and this can include processes that replace the incoming particles a, b with a', b', such as replacing an incoming quark a with an incoming gluon that splits into an incoming quark a and an outgoing antiquark $\bar{a}$.

At NLO, additional complication arises compared with LO. The added loop in the virtual contribution involves integration over the four-momentum l in the loop, which is not constrained beyond the conservation of four-momentum at the vertices, and can be infinitely large. Therefore the integral can be potentially infinite if it has the general form of d^4l/l^m for $m \leq 4$. This ultraviolet (UV) divergence needs

to be regularized. Typically we use the method of dimensional regularization, which corresponds to integrating with dimension $D = 4 - 2\varepsilon$ and eventually letting $\varepsilon \to 0$. Then we can use standard practice of renormalization by adding counterterms and redefining the Lagrangian.

On the other hand, as we have mentioned in Section. 1.5.1, in both the virtual and real corrections there are also infrared divergences associated with emission of particle that is either with very low energy (soft) or almost parallel to the emitting particle (collinear). This is expected since we cannot talk about an observable without specifying the detector resolution, unless this observable is insensitive to such unresolvable radiation. According to the Kinoshita-Lee-Nauenberg (KLN) theorem [54, 55], such singularities will cancel out order by order in perturbation theory if we sum over all the initial and final states degenerate in energy, and renormalization is not needed. In other words, if we look at observables that are inclusive over initial and final states soft and collinear configurations, such as an inclusive cross section, then the infrared singularity will be canceled for these observables. Note that as discussed in Section 1.5.1, the KLN cancellation does not occur for the collinear singularity associated with emission from parton coming into the hard interaction, and this singularity needs to be absorbed into the PDF, since the real and virtual contribution in this case correspond to different energy. We call an observable infrared safe if its value is insensitive to soft and collinear emission.

In practice, analytical calculation of the matrix element and integrals only apply to simple case, and also due to the nonanalytic structure of the PDFs, generally in high energy physics we need to use the Monte Carlo (MC) simulation approach to perform the integral. However, the real and virtual integrals are in the different n- and $(n+1)$-particle phase space, and each of the integral has infrared divergence but can be canceled only after the integrals have been evaluated. To address this difficulty

in numerical calculation, a typical method to use is called infrared subtraction. This method construct a real subtraction term $\mathcal{S}_n$ that lives in the (n+1)-particle phase space, and an integrated subtraction term $\mathcal{I}_n^{(\mathcal{S})}$ that lives in the n-particle phase space, such that

$$0 \equiv \int d\Phi_{\mathcal{B}} \mathcal{I}_n^{(\mathcal{S})} \left(\Phi_{\mathcal{B}}; \mu_F, \mu_R\right) - \int d\Phi_{\mathcal{R}} \mathcal{S}_n \left(\Phi_{\mathcal{R}}; \mu_F, \mu_R\right) \tag{1.122}$$

and eqn. (1.119) can be written as

$$\begin{aligned} \sigma_n^{NLO} = & \int d\Phi_{\mathcal{B}} \left[\mathcal{B}_n \left(\Phi_{\mathcal{B}}; \mu_F, \mu_R\right) + \mathcal{V}_n \left(\Phi_{\mathcal{B}}; \mu_F, \mu_R\right) + \mathcal{I}_n^{(\mathcal{S})} \left(\Phi_{\mathcal{B}}; \mu_F, \mu_R\right)\right] \\ & + \int d\Phi_{\mathcal{R}} \left[\mathcal{R}_n \left(\Phi_{\mathcal{R}}; \mu_F, \mu_R\right) - \mathcal{S}_n \left(\Phi_{\mathcal{R}}; \mu_F, \mu_R\right)\right] \end{aligned} \tag{1.123}$$

There have been well-established process-independent ways of constructing such subtraction terms such that both the $d\Phi_{\mathcal{B}}$ and $d\Phi_{\mathcal{R}}$ integrals are individually finite, allowing the use of numerical methods such as MC integration. For more details one can refer to [12]. However, the integrands in eqn. (1.123) are unbounded [26] and therefore we can fill the generated events into distribution plots but cannot unweight the events using the approach discussed in Chapter 3. This means the subtraction method at fixed order is associated with MC integrator and not MC event generator. To generate unweighted events in simulation we need to combine fixed-order calculation with parton shower (PS) discussed in the next subsection and in Chapter 3.

The NNLO calculation is conceptually similar but with substantially more complication. To illustrate the different kinds of contributions, we can refer to Fig. 1.10 (left). This diagram corresponds to a contribution of order α_S^2 to the total amplitude squared that can be achieved through products of different amplitudes, whose diagrams can be obtained by making different cuts (a)-(d) on Fig. 1.10 (left), and the meaning of these cuts are further illustrated in Fig. 1.11. The diagrams shown are

in the case of a DY process but the same kind of ideas apply to more general cases. In case (a) we have contribution from interference between two-loop diagrams and Born-level diagrams:

$$\Re\mathfrak{e}\left[\mathcal{A}^{2-\text{loop}} \times \mathcal{A}^{\text{Born}*}\right] \tag{1.124}$$

In case (b) we have one-loop diagram amplitude squared:

$$\left|\mathcal{A}^{1-\text{loop}}\right|^2 \tag{1.125}$$

In case (c) we have contribution from interference between diagrams with one loop plus one jet and diagrams with one jet:

$$\Re\mathfrak{e}\left[\mathcal{A}^{1-\text{loop}+1-\text{jet}} \times \mathcal{A}^{\text{1-jet}\ *}\right] \tag{1.126}$$

In case (d) we have real emission (2 jets) diagram amplitude squared:

$$\left|\mathcal{A}^{\text{2-jet}}\right|^2 \tag{1.127}$$

In general, modifications at the next order can also occur when new channels open up, such as the $qg \to ZZq$ process that opens at NLO and the $gg \to ZZ$ process at NNLO, which are shown in Fig. 1.12 along with the LO $q\bar{q} \to ZZ$ contribution. The ZZ production processes are discussed later in Section 1.6, where they constitute a concrete example that shows various contributions from different orders of α_S.

For processes with quark-antiquark initial states which is typical at LHC, real emission diagrams containing one or even two gluons can start entering the process at NLO. Due to the abundance of gluons in pp collisions, these new contributions can introduce large corrections not captured by lower-order calculations. Since the channels with new partonic initial states behave like additional LO processes that occur at higher orders, such contributions typically also give rise to a sizable additional scale dependence not captured by uncertainties estimated from scale variations at

lower orders. In general, as discussed in Ref. [12], apart from some pathological cases, higher-order calculations always diminish the scale dependence, and at least an NLO calculation is mandatory in order to have a reliable prediction value [24]. Meanwhile, the scale uncertainty is further reduced at NNLO, which is the first order where one expect to obtain reliable estimate for the scale uncertainty [24], as in the case of ZZ production [59].

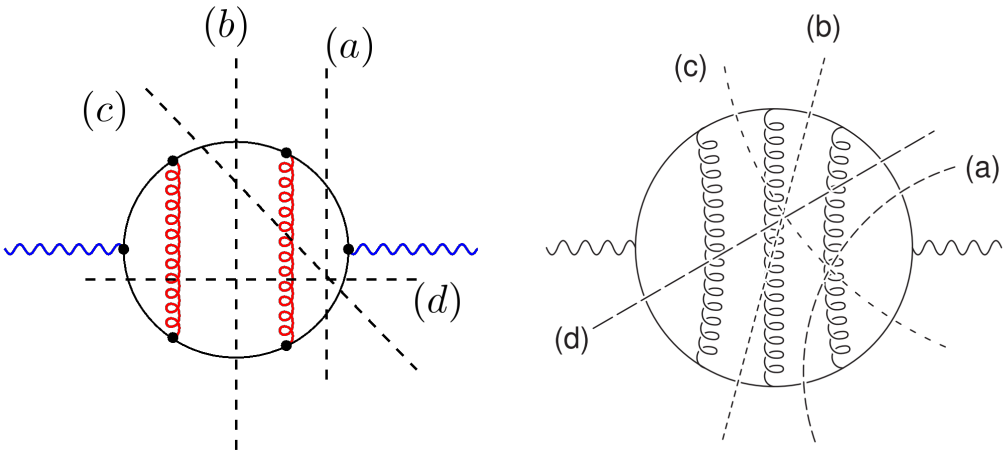

Figure 1.10: A three loop diagram (left) illustrating possible NNLO contributions. It can be cut in all possible ways shown by the dashed lines (a)-(d). A similar diagram (right) is from Ref. [12] but has four loops and corresponds to processes with one more additional jet.

1.5.3 NLO Electroweak Corrections

When discussing physics in hadron collisions it is natural to focus on expansion in order of α_S, since the hard scattering must involve strong interaction and α_S is an order of magnitude larger than the EW coupling in terms of α. However, since $\alpha \sim \alpha_S^2$, one would expect that the NLO EW corrections should be considered along with the NNLO QCD corrections, whereas the latter is already necessary for data description given the current LHC experiment precision. In fact, NLO EW corrections have been

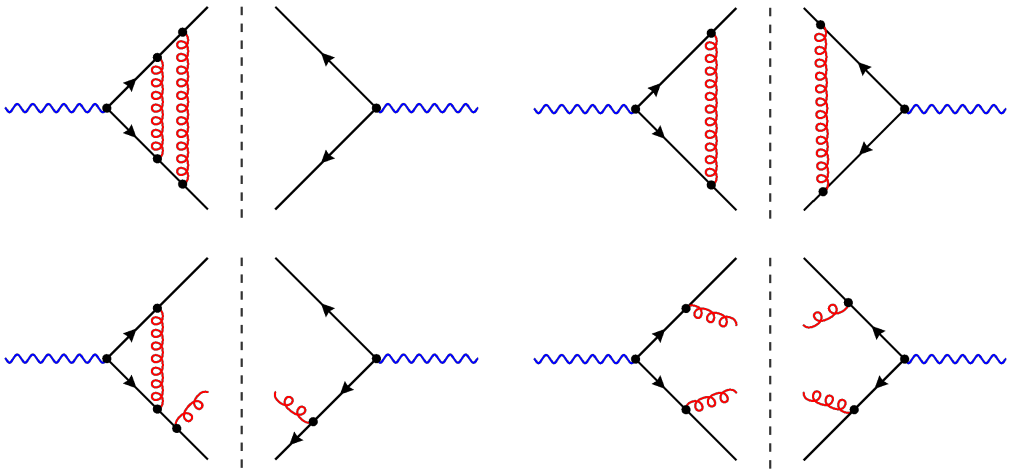

Figure 1.11: Different types of NNLO contributions corresponding to the cuts through (a) (top left), (b) (top right), (c) (bottom left), (d) (bottom right) in Fig. 1.10 (left).

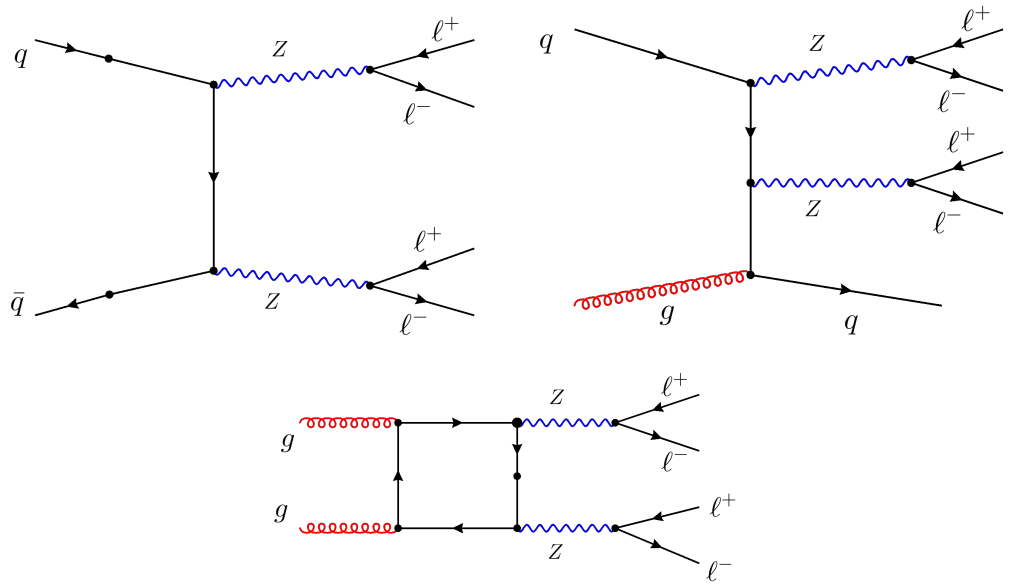

Figure 1.12: Example contributing Feynman diagrams to diboson ZZ production, which include the LO contribution, the qg $\rightarrow$ ZZq process that opens at NLO QCD, and the gg $\rightarrow$ ZZ process at NNLO QCD.

applied as a K-factor in the NNLO sample used in the ZZ+jets analysis as discussed in Section 6.2. A K-factor (See Ref. [12] for more details) is a ratio between higher-order and lower-order cross sections that can be used for corrections. A K-factor can be either global or local (differential).

As discussed in Ref. [12], one can estimate the size of the leading relative EW for the simplest process with

$$\delta^{\mathrm{EW}} = \frac{\hat{\sigma}^{\mathrm{EWCorr}}}{\hat{\sigma}_0} = -(\text{ constant }) \frac{\alpha}{4\pi} \log^2 \left(\frac{s}{m_V^2} \right) \tag{1.128}$$

where s here is the hard scale associated with the process, and m_V is the mass of the massive vector boson involved. This is only a crude approximation but can give some good guidance on the size of the EW corrections to be expected, as illustrated in Fig. 1.13 where δ^{EW} is plotted as a function of $\sqrt{s}$ with the constant in eqn. (1.128) set to 1. This indicates that in regions of large invariant mass or transverse momentum, the NLO EW corrections can become significant and are possible to be as large as their QCD counterparts, which motivates us to take into account this effect.

1.5.4 Secondary Emissions and Resummation

The fixed-order perturbation theory is suitable in describing the hard interactions resulting in production of heavy states (such as vector bosons and Higgs boson) and QCD jets that are hard and well-separated. Pure fix-ordered calculation can be used to evaluate perturbatively well-behaved observables. On the other hand, we need to also describe secondary emission in the soft (associated with gluon) and collinear regime, which includes final state radiation (FSR) emitted by final state particles, and initial state radiation (ISR) emitted by incoming particles before the hard interaction. In this part of the phase space higher order terms are enhanced

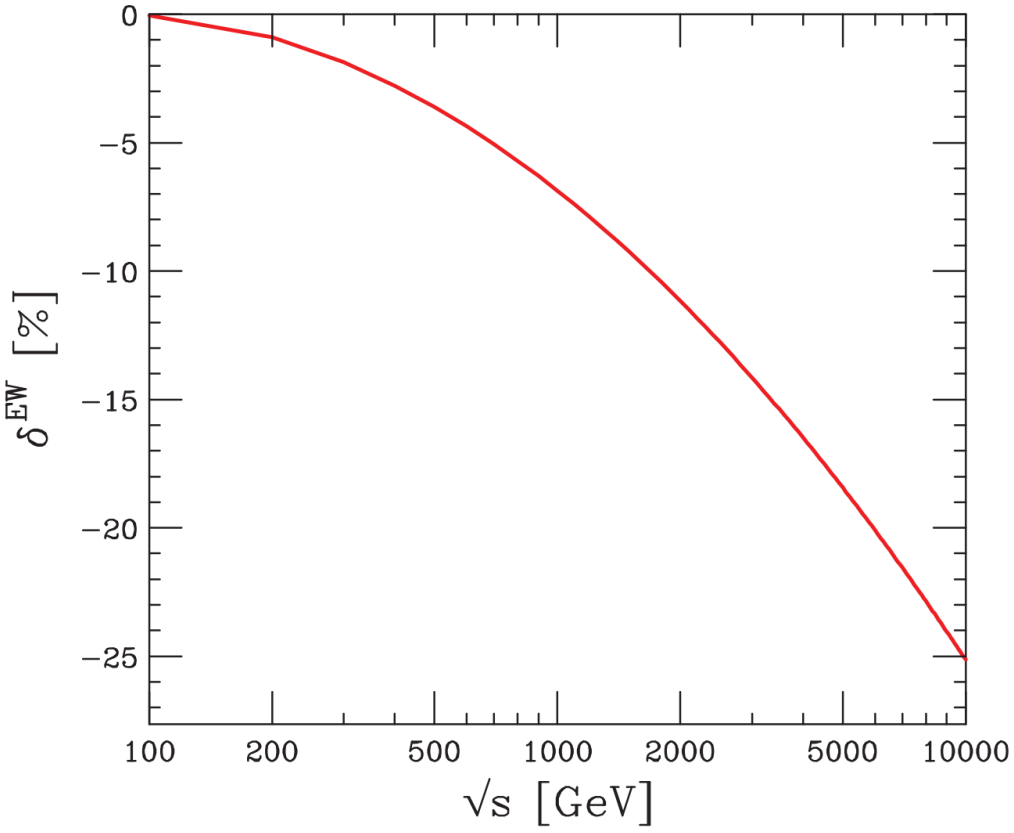

Figure 1.13: A plot from Ref. [12] showing an estimate of the importance of the leading EW correction obtained from eqn. (1.128) with the constant set to 1, plotted as a function of $\sqrt{s}$.

and cannot be neglected. We first focus on the collinear limit. In a process where an outgoing parton a with original energy E splits into b and c, with relative angle θ and the remaining energy fraction $z = E_b/(E_b + E_c) = E_b/E$, in the collinear limit, the squared amplitude can be factorized as:

$$|\mathcal{M}_{n+1}|^2 \, d\Phi_{n+1} \approx |\mathcal{M}_n|^2 \, d\Phi_n \frac{dt}{t} dz \frac{d\phi}{2\pi} \frac{\alpha_S}{2\pi} P_{a,bc}(z) \tag{1.129}$$

which is another way to present eqn. (1.100), and t is the hardness introduced in Section. 1.5.1. Integrating with dt/t from the cutoff scale $t = Q_0^2$ to the hard scale of the primary process $t = Q^2$ and ignoring the soft singularity for now, we have a factor of $\alpha_S \log(Q^2/Q_0^2)$. We can recursively apply eqn. (1.129) to subsequent

splitting, provided that the splitting are strongly ordered in decreasing hardness, which means the subsequent emission has much smaller hardness than all the previous ones, $t \gg t' \gg t'' \ldots$, so that the collinear approximation remains valid. For k emissions we have

$$\sigma_{n+k} \sim \sigma_n \alpha_S^k \int_{Q_0^2}^{Q^2} \frac{dt}{t} \int_{Q_0^2}^{t} \frac{dt'}{t'} \ldots \int_{Q_0^2}^{t^{(k-2)}} \frac{dt^{(k-1)}}{t^{(k-1)}} \sim \sigma_n \left(\frac{\alpha_S}{2\pi}\right)^k \log^k \left(Q^2/Q_0^2\right) \quad (1.130)$$

In this case each α_S is accompanied by a large logarithm and the perturbation theory breaks down. Therefore, we need to account for the large logarithm in all orders, using an approach called resummation. Typically we look at an infrared safe observable that has large logarithm in some parts of phase space and it can be expressed as (see [18]):

$$A\left(\alpha_S\right) \exp\left[L g_0\left(\alpha_S L\right) + g_1\left(\alpha_S L\right) + \alpha_S g_2\left(\alpha_S L\right) + \ldots\right] \quad (1.131)$$

where L represents the large logarithm, α_S is evaluated at a certain energy scale, A can be written as power expansion of α_S starting with constant term, and the functions g_i has power expansion in $\alpha_S L$ starting from the linear term. We say the observable has leading logarithm (LL) accuracy if only g_0 is included, which means there is a double log in the exponential. An observable has next-to-leading-logarithm (NLL) accuracy if g_1 is also included. In both the cases of LL and NLL, only the constant term in A is needed. For NNLL, g_2 is also included and A is up to order α_S. In general, an observable has N^{n}LL accuracy if all g_i functions with $i \leq n$ are fully included, and A is given up to order $n-1$ in α_S, and a dedicated resummation calculation needs to be performed for each new variable [18]. In practice, equivalently the resummation of collinear- and/or soft-enhanced contributions to all orders can be performed with the parton shower method.

1.5.5 Sudakov Form Factor and Parton Showers

The Sudakov form factor $\Delta(t_m, t_n)$ is a crucial concept in a parton shower algorithm. It can be interpreted as the probability of no resolvable emission between scale t_m and t_n, and can be derived as following: the probability of having resolvable emission between t_m and t is

$$P(t_m, t) = 1 - \Delta(t_m, t) \tag{1.132}$$

Taking the derivative with respect to t, we have

$$-\frac{d}{dt}\Delta(t_m, t) = R(t)\Delta(t_m, t) \tag{1.133}$$

where $R(t)$ is the emission probability density at t, and can be obtained from eqn. (1.129):

$$R(t) = \frac{1}{t}\int_{z_{min}(t)}^{z_{max}(t)} dz \int \frac{d\phi}{2\pi}\frac{\alpha_S}{2\pi}P_{a,bc}(z) \tag{1.134}$$

$R(t)\Delta(t_m, t)$ represents the probability density of resolvable emission only happening at t. Now we have a differential equation for $\Delta(t_m, t)$, and solving it gives

$$\Delta_a(t_m, t_n) = \exp\left[-\int_{t_m}^{t_n}\frac{dt}{t}\frac{\alpha_S(t)}{2\pi}\int_{z_{min}(t)}^{z_{max}(t)} dz \int \frac{d\phi}{2\pi}\sum_{bc} P_{a,bc}(z)\right] \tag{1.135}$$

Here we have summed over the contributions from different parton splitting, and explicitly written out the dependence of α_S on the hardness. A more careful treatment suggests that we need to use k_T^2 from eqn. (1.103) as the argument in the running α_S [19, 60]. For the integral in $\Delta(t_m, t_n)$ to be finite, we need to introduce an infrared cut off t_c and only look at hardness range with $t \geq t_c$. t_c classifies what we regard as a resolvable emission. It is typically chosen to be about a few Λ_{QCD} (order of 1 GeV) to avoid the Landau pole of α_S, and below which the perturbation theory eventually breaks down. We can then combine the approximate perturbative treatment above t_c with the nonperturbative hadronization below t_c which we discuss in Section 1.4.1.

The integration limits of the energy fraction z is constrained by the hardness t. The introduction of the infrared cutoff scale t_c guarantees that these z limits avoids the divergence in the splitting function [12], and we can ignore the plus prescription and delta function inside the regularized splitting functions. The splitting with z above z_{max} is treated as too soft to be resolved. Note that without ambiguity, in some cases the order of the arguments in $\Delta(t_m, t_n)$ is switched and it is written instead as $\Delta(t_n, t_m)$, or simply as $\Delta(t_n)$ with the starting scale t_m implicit.

Although the Sudakov form factor only explicitly involves the real resolvable emission in its formula, it has incorporated the enhanced unresolvable real emission below the cutoff t_c and the enhanced virtual contributions through conservation of probability, that is, the sum of splitting and no splitting probability is equal to 1. With this unitarity we also see that the sum of unresolvable virtual and real emission is finite even though they are both divergent, which is a manifestation of the KLN theorem. Performing the integrals with t and z results in a double log (L^2) in the leading contribution, and we see that the Sudakov form factor has a form matching eqn. (1.131), as it is a resummed calculation to all orders.

Using the Sudakov form factor, parton showering in final states can be performed recursively with the following prescription. Start with a parton exiting the interaction vertex with hardness t_n, we solve for the next branching hardness with the equation

$$r = \Delta_a(t_m, t_n) \tag{1.136}$$

where r is a random number uniformly distributed between 0 and 1. If $t_m < t_c$, no splitting occurs and we regard this line as a final parton, and move on to the next parton. If $t_m > t_c$, a new splitting occurs at hardness t_m, with z value and resulting parton species given by probability $P_{a,bc}(z)$. Angular correlations can be accounted for [18], and if they are ignored, then ϕ can be generated uniformly. This

procedure can be applied recursively to all partons and generate an arbitrary number of partons, until no further splitting occurs. Different algorithms may be used to ensure conservation of four-momentum [18].

The parton shower method can also be applied to ISR, but the difference is that the initial hadrons are fixed, and for calculation efficiency we need to start with the partons in the hard interaction and use backward evolution. To derive the Sudakov form factor in this case, we can use the DGLAP equation eqn. (1.94) written in a different notation:

$$\frac{df_{b/h}(x,t)}{d\log t} = \frac{\alpha_S(t)}{2\pi}\sum_a \int_x^1 \frac{dz}{z}\hat{P}_{a,bc}(z)\, f_{a/h}\left(\frac{x}{z},t\right) \tag{1.137}$$

If we start with an incoming parton b at momentum fraction x at hardness t, and change the hardness by δt, this open up a window of contributions from partons at $[t, t+\delta t]$, and the probability of parton b coming from a splitting in this interval is given by the fraction [61]:

$$dP_b = \frac{df_{b/h}(x,t)}{f_{b/h}(x,t)} = \frac{dt}{t}\frac{\alpha_S(t)}{2\pi}\sum_a \int_{z_{min}(t)}^{z_{max}(t)} \frac{dz}{z}P_{a,bc}(z)\,\frac{f_{a/h}\left(\frac{x}{z},t\right)}{f_{b/h}(x,t)} \tag{1.138}$$

and the Sudakov form factor for ISR is given by $1-P_b$, which gives

$$\Delta_b^{ISR}(t,T) = \exp\left\{-\int_t^T \frac{dt'}{t'}\int_{z_{min}(t)}^{z_{max}(t)}\frac{dz}{z}\int\frac{d\phi}{2\pi}\frac{\alpha_S(t')}{2\pi}\sum_{ac}P_{a,bc}(z)\frac{f_{a/h}\left(\frac{x}{z},t'\right)}{f_{b/h}(x,t')}\right\} \tag{1.139}$$

and similarly we can use it to perform parton showering for ISR.

In the collinear picture discussed so far, we have been treating the emission from each parton line as independent, and have ignored the interference effect. Soft gluon emission with four-momentum q, energy ω and polarization ε from an external line with four-momentum p corresponds to a universal, spin-independent factor of

$$F_{\text{soft}} = \frac{p \cdot \varepsilon}{p \cdot q} \tag{1.140}$$

and therefore when we square the matrix elements, there are interference terms that need to be included to account for soft enhancement properly. The cross section has a factor that sums over all pairs of external lines $\{i, j\}$:

$$d\sigma_{n+1} = d\sigma_n \frac{d\omega}{\omega} \frac{d\Omega}{2\pi} \frac{\alpha_S}{2\pi} \sum_{i,j} C_{ij} W_{ij} \tag{1.141}$$

where the radiation function

$$W_{ij} = \frac{\omega^2 p_i \cdot p_j}{p_i \cdot q p_j \cdot q} = \frac{1 - \cos\theta_{ij}}{(1 - \cos\theta_{iq})(1 - \cos\theta_{jq})} \tag{1.142}$$

in the massless approximation for emitting particles, and

$$C_{ij} = -\boldsymbol{Q}_i \cdot \boldsymbol{Q}_j \tag{1.143}$$

is a color factor, with $\boldsymbol{Q}_i$ the color charge vector in a suitable color space representation.

$$d\Omega = d\cos\theta_{iq} d\phi_{iq} \tag{1.144}$$

is the solid angle element of the emitted gluon.

It turns out that we can use a formalism called coherent color branching to compute soft enhancement of all orders and treat both collinear and soft enhancement correctly (see [19] for example). It involves simple modification of the collinear branching procedure described so far by using the evolution variable

$$\zeta = \frac{p_b \cdot p_c}{E_b E_c} \approx 1 - \cos\theta \tag{1.145}$$

and imposing angular ordering $\zeta' < \zeta$ or a factor of $\Theta(\zeta - \zeta')$ for successive branchings. The hardness t for the branching $a \to bc$ becomes $E^2\zeta$ with cutoff $E^2\zeta > t_c$.

Angular ordering is used as the default evolution variable in all versions of HERWIG (see [18, 62]). In PYTHIA8 [63] which is used for the MC samples in this thesis, p_{T}-ordered shower is used by default and it represents a hybrid that combines collinear splitting kernels with dipole (parton pair emission) kinematics [18, 64]. Note that with angular ordering, the first emissions are often not the hardest ones.

We will discuss how the fixed-order calculation and parton shower are combined to obtain description of data with good accuracy in Chapter 3.

1.5.6 Underlying Events

Besides the main hard scattering of interest, there are other processes in pp collisions that need to be taken into account, which corresponds to what is called the underlying events (UE). As mentioned in [65], from the experiment point of view, underlying events contains all activity in a hadronic collision that is not related to the signal particles from the hard process. In particular, a complication arise from the fact that during the hadron collision more than one parton from a single hadron can interact, causing what is known as double-parton scattering or multi-parton interactions (MPI), which can be hard or soft. This effect is beyond the standard factorization discussed so far and needs to be modeled properly. In addition, the remnants of the incident hadron (beam remnants) can also contribute to the overall particle multiplicity. The beam-beam remnants (BBR) refer to what remain after a parton is scattered out of each of the two incident beam hadrons. As we will also discuss in Chapter 3, standard MC generators have adjustable parameters that can control their event modeling behavior, and a set of such parameters that can be adjusted to better fit certain aspects of the experimental data is referred to as a tune. A typical UE tune includes parameters that accounts for effects currently not calculable from first principles alone, such as those related to the MPI energy dependence and

range of color reconnection, and is obtained from fitting data from a large span of center of mass energies from different colliders including the LHC. For more details one can refer to [66–68]. For the analyses in this thesis we used the CMS UE tunes CUETP8M1 [67] (CP5 [68]) tune for the 2016 (2017 and 2018) data-taking period.

1.6 Diboson and Multiboson Physics

In LHC, inclusive multiboson production is dominated by radiation of vector boson off the initial-state quarks from the proton, and this includes EW production where the final-state vector bosons result from pure EW interactions among the radiated vector bosons (and Higgs boson) through scattering or fusion. In particular, it includes pure EW production of two vector bosons associated with 2 jets through such scattering. The multiboson production can also be categorized into processes such as diboson production, triboson production, vector boson fusion (VBF) and vector boson scattering (VBS). Such interactions are important for testing and understanding the gauge structure of SM, as discussed further in the case of diboson production below. As mentioned in Ref. [32], cross sections of several multiboson states have been measured with high precision approaching the 3% total uncertainty level, and currently only the rarest of the multivector boson processes, such as ZZ VBS production (detected with 4σ significance [4]) and some of the triboson production processes involving two or more weak bosons, have not been observed by the CMS experiment.

In this thesis we focus on the production processes of pairs of vector boson (diboson production), which are very important to measure experimentally for the following reasons. First of all they provide a wide range of benchmark cross sections for testing SM, as shown for example in Fig. 1.14, where we see agreement between theory and experiment over all diboson production states. Theoretical predictions at NNLO

QCD are needed for comparison with data, since the current precision of dibson cross section measurement (at a few percent level) requires the predictions to reach the same level of accuracy, and also because NNLO can open new channels that have important contribution. Furthermore, the diboson production provides important test of the non-Abelian gauge structure of the SM theory, and is sensitive to new physics beyond SM. For example, the anomalous triple gauge boson couplings (aTGC) may occur due to BSM modifications to triple gauge boson couplings (TGC) via loop diagrams, or by directly replacing one internal boson line with a virtual heavy BSM particle (see [69]), and can be investigated by making precise measurements of different diboson cross sections, whereas the anomalous quartic gauge boson couplings (aQGC) can be studied with VBS and polarization.

Besides testing SM and probing for new physics, diboson production is closely related to study of the Higgs boson which is one of the most important topics in LHC physics, since it constitutes a significant background for processes with Higgs boson decaying to W or Z pairs and needs to be disentangled from the Higgs signal properly. In this section we discuss various aspects of the diboson ZZ production relevant for analyses present in this thesis.

The high center of mass energy and large statistics in LHC Run 2 has allowed observation of the triboson production for the first time, and the most challenging measurements involve production of three massive vector bosons. Though triboson production is not the signal process studied in this thesis, as discussed in Section 1.6.6, the WWZ, WZZ, ZZZ production constitute irreducible background for the ZZ analyses, for final states that have four genuine leptons (full leptonic decay for WWZ, semileptonic decay for WZZ, ZZZ). This is only a minor background due to the strict requirement of four isolated and well-reconstructed leptons, and also because of the small production cross section of the triboson processes.

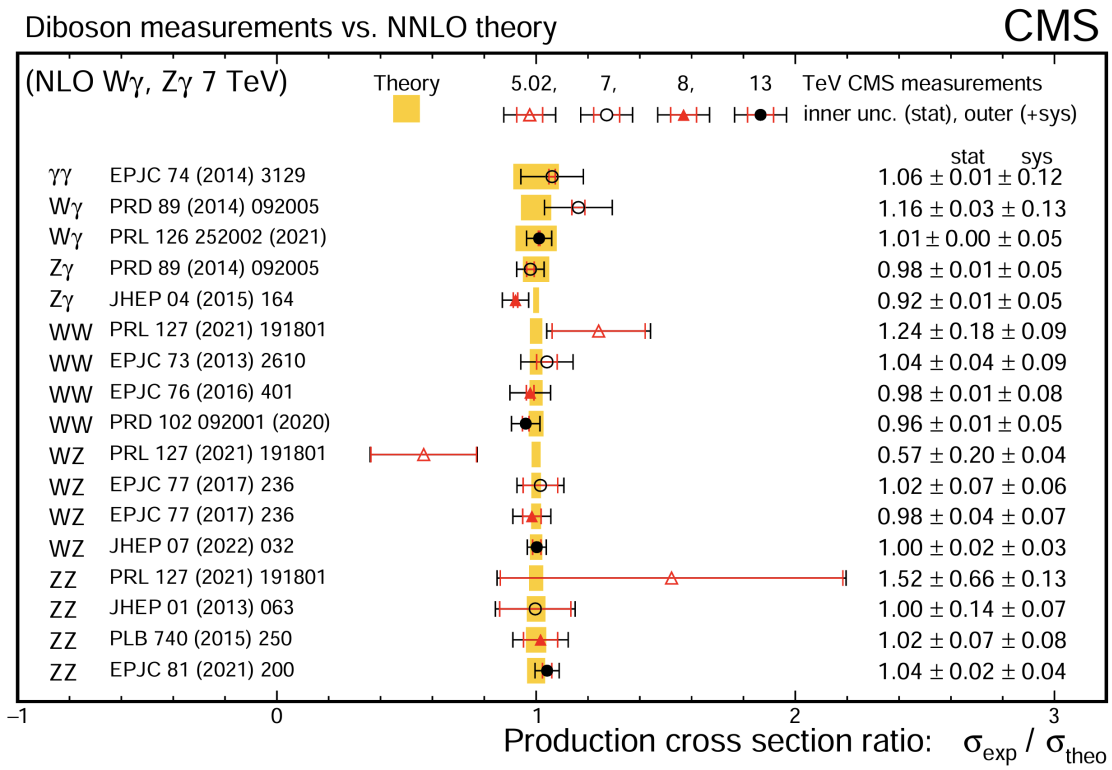

Figure 1.14: Summary of cross section measurements for diboson production shown as a ratio over the NNLO or NLO QCD predictions. This figure is from Ref. [32]. The yellow bands indicate the uncertainties in the theoretical predictions and the error bars on the points are the statistical uncertainties, whereas the outer bars are the combined statistical and systematic uncertainties.

1.6.1 The Z Resonance and Decay

In unitary gauge, the Feynman rule giving the propagator of a massive vector boson is

$$\frac{i}{p^2 - m^2}\left(-g^{\mu\nu} + \frac{p^\mu p^\nu}{m^2}\right) \tag{1.146}$$

Considering the vector boson to be the Z boson, it would appear that the matrix element diverges when $\sqrt{s}$ approaches m_Z. To resolve this, we need to account for the fact that the Z boson is unstable and has a finite lifetime and decay width. We need to include the full decay width Γ_Z and make the following modification:

$$m_Z^2 \to (m_Z - i\Gamma_Z/2)^2 = m_Z^2 - im_Z\Gamma_Z - \frac{1}{4}\Gamma_Z^2 \tag{1.147}$$

The experimentally measured value of m_Z and Γ_Z are ([18])

$$m_Z = 91.1880 \pm 0.0020\text{GeV}, \quad \Gamma_Z = 2.4955 \pm 0.0023\text{GeV} \tag{1.148}$$

Therefore $\Gamma_Z \ll m_Z$ and depending on the context we may ignore the $\frac{1}{4}\Gamma_Z^2$ term to a good approximation. With the modified propagator, if we calculate the cross section of a process like $e^+e^- \to Z \to \mu^+\mu^-$, then to lowest order there will be a peak of the cross section centered at m_Z and with full-width-at-half-maximum (FWHM) equal to Γ_Z. However, higher order corrections such as the effects from initial state radiation (ISR) will distort the shape of the resonance curve plotted as a function of the $\sqrt{s}$ of the system.

The branching ratios $\text{Br}(Z \to f\bar{f}) = \Gamma(Z \to ff)/\Gamma_Z$ measured experimentally are [18]:

$$\begin{aligned} &\text{Br}(Z \to \text{invisible}) = 20.000 \pm 0.055\%, \quad \text{Br}(Z \to \text{ hadrons }) = 69.911 \pm 0.056\%, \\ &\text{Br}\left(Z \to e^+e^-\right) \approx \text{Br}\left(Z \to \mu^+\mu^-\right) \approx \text{Br}\left(Z \to \tau^+\tau^-\right) \approx 3.5\% \end{aligned} \tag{1.149}$$

The charged lepton final states only correspond to a small branching ratio, but have the advantage of being a clean signature with low background.

1.6.2 Nonresonant ZZ Production

At tree level, the nonresonant $pp \to ZZ$ production proceeds mainly through the t- and u-channel quark-antiquark initiated scattering diagrams as shown in Fig. 1.15. This is also called a double-resonant contribution, as contrary to resonant production discussed in Section 1.6.4 where a single intermediate boson is produced. At NLO QCD, as discussed in Section 1.5.2, we have additionally the real emission diagrams that have one extra emission, and the Feynman diagrams can be initiated with gluon. We also have virtual corrections from diagrams with one additional loop, which are

formally NNLO diagrams, but contribute to NLO through interference with the Born-level diagrams. The example NLO diagrams are shown in Fig. 1.16. At NNLO QCD (example diagrams in Fig. 1.17), we have the contributions from squared amplitudes of one-loop diagrams (for example the same loop diagram in Fig. 1.16) and diagrams with two additional emissions. We also have contributions from diagrams with one loop plus one extra emission and two loops through interference. On the other hand, the loop-induced gluon-gluon fusion also opens up at NNLO, which has an important contribution to pp $\rightarrow$ ZZ production due to the high luminosity of gluons, and provides about 60% of the total NNLO effect [70]. This large correction shows the importance of corrections from new intial states at higher orders for accurate determination of the cross sections, as already discussed in Section. 1.5.2. At NNLO, all partonic channels are included for ZZ, and the cross section prediction along with theoretical uncertainties are expected to be reliable [59].

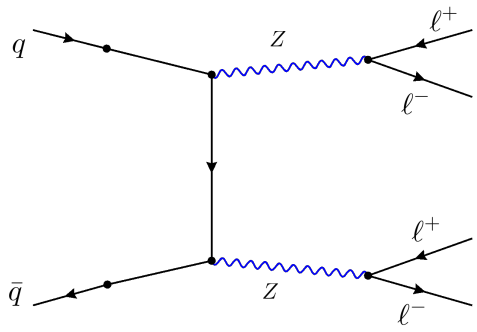

Figure 1.15: Feynman diagram showing t-channel q$\bar{\text{q}}$-initiated diboson ZZ production at tree level.

1.6.3 Z vs γ* Production

Since the Z boson couples to all SM fermions, the photon in any QED process can be replaced by a Z to obtain another valid SM process. In the diboson production processes discussed so far, the Z boson can also be replaced with a photon when no

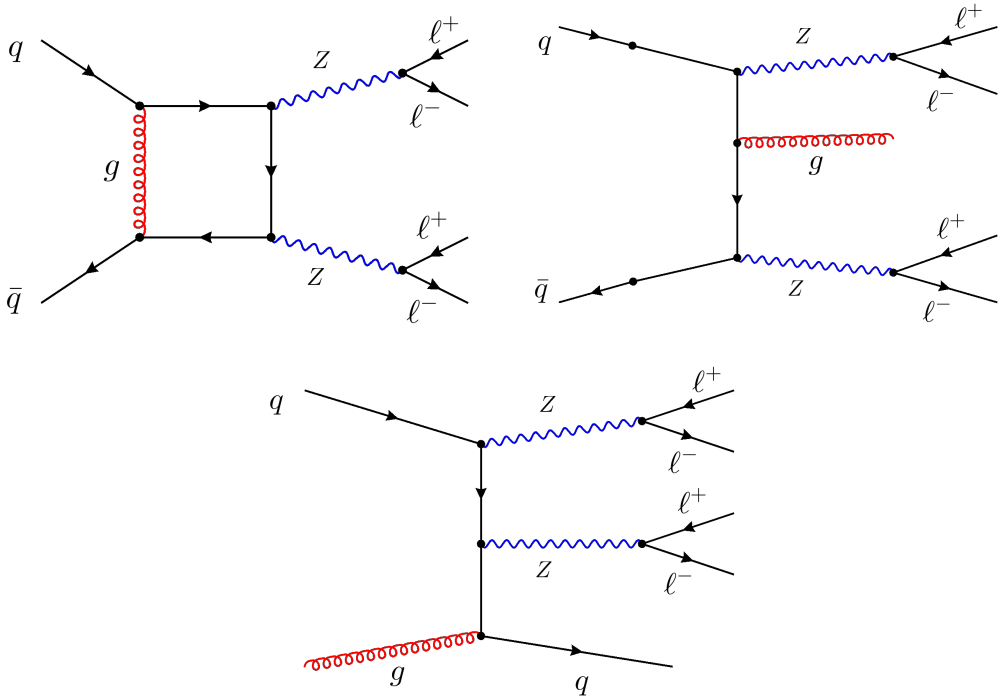

Figure 1.16: Example contributing Feynman diagrams to diboson ZZ production at NLO QCD. The loop diagrams are formally at NNLO QCD but interfere with the Born-level contributions to contribute at NLO QCD.

coupling to neutrino is involved. Therefore our selected events of dibosn ZZ production will also include a mixture of those containing Z$\gamma*$ and $\gamma*\gamma*$ production, which can be suppressed by optimizing event selections. By choosing a Z mass window of 60-120 GeV as discussed in Section 4.3 (typcially 66-116 GeV in ATLAS ZZ analyses, which causes a small difference in total ZZ cross section calculation [32]), we are able to remove most of the γ^* contamination for nonresonant ZZ production events [71], while maintaining good efficiency and sufficient statistics for the ZZ signal events.

1.6.4 Resonant ZZ*/Zγ* Production

In the ZZ+jets analysis we also look at distributions without imposing the 60-120 GeV requirement (referred to as the on-shell requirement) on the two Z bosons, there-

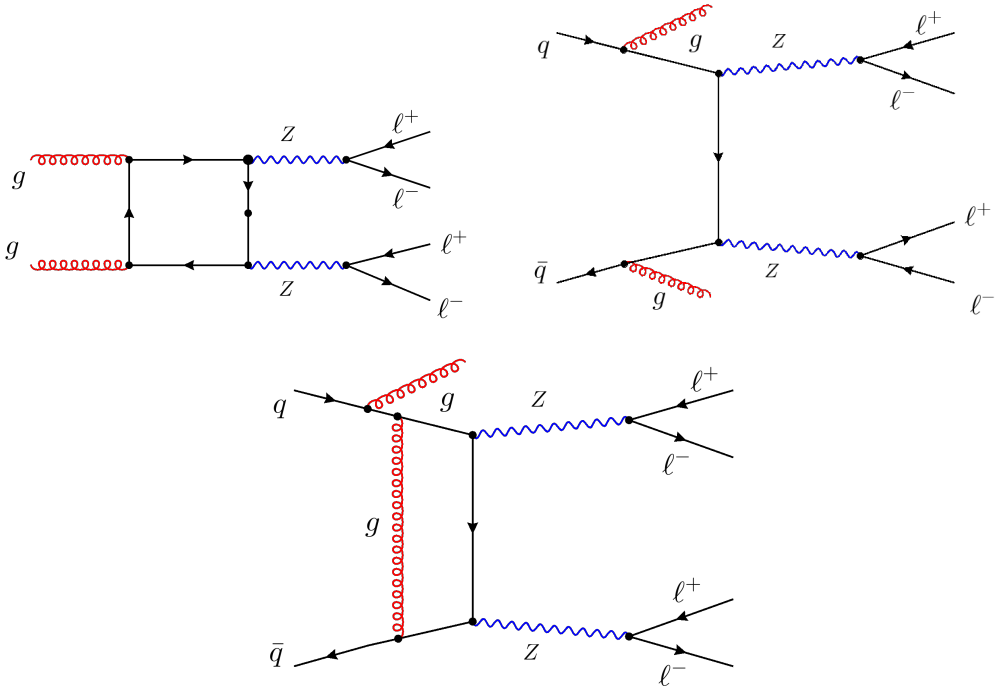

Figure 1.17: Example contributing Feynman diagrams to diboson ZZ production at NNLO QCD, where gluon-gluon fusion also contributes.

fore one of the Z bosons can have mass significantly away from the resoncane peak near 91 GeV (which we refer to as being off-shell in the experimental point of view). Loosening the requirement allows us to also look at resonant ZZ*/Zγ* production (single-resonant) processes. In Fig. 1.18 we have production of four leptons associated with a Z resonance where the second Z boson is raidated off a lepton line. This is asscoiated with a peak in the four-lepton invairant mass ($m_{4\ell}$) distribution in the $80 - 100$ GeV range as we will discuss in Section 6.1, which can be used to measure the Z $\to 4\ell$ branching ratio [72] with proper event selections. On the other hand we also have Higgs boson to ZZ production peaked at around $m_H = 125\,\text{GeV}$ in the $m_{4\ell}$ distribution, where the dominant contribution comes from the gg $\to$ H $\to$ ZZ production shown in Fig. 1.19 (top left). Other H $\to$ ZZ processes includes the VBF production, the VH production ($V = \text{W}, \text{Z}$) with Higgs boson decaying to ZZ,

and the $t\bar{t}H$ production with Higgs boson decaying to ZZ, which are also shown in Fig. 1.19. These contributions are suppressed and negligile when the $60 - 120\,\text{GeV}$ on-shell requirement is imposed.

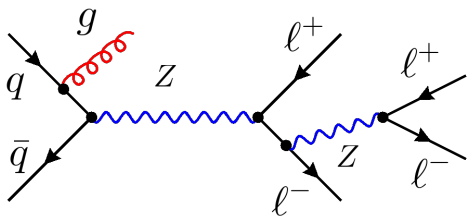

Figure 1.18: Example Feynman diagrams showing the $Z \to 4\ell$ production. The additional gluon emission in the left diagram stresses that we want to look at ZZ production assoicated with different numbers of jets.

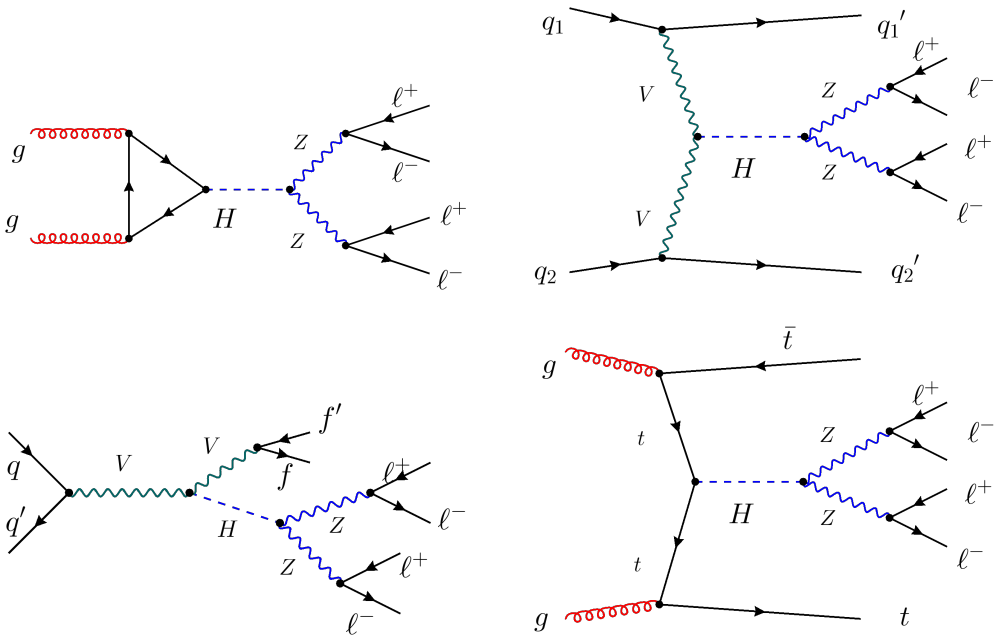

Figure 1.19: Example Feynman diagrams showing the processes containing Higgs to ZZ production, which include $gg \to H \to ZZ$ production (top left), VBF production (top right), VH production (bottom left) and $t\bar{t}H$ production (bottom right).

1.6.5 Vector Boson Scattering

Vector boson scattering (VBS) refers to processes where the vector bosons radiated off the intial-state quarks interact and scatter, resulting in pure EW production of two outgoing vector bosons, along with two jets from the incoming quarks. It provides a crucial test of the EW sector of SM. Example Fyenman diagrams of the EW production of diboson ZZ are shown in Fig. 1.20, which include scattering through two TGC vertices (top left), scattering through a quartic gauge boson coupling (QGC) vertex (bottom left), and scattering via Higgs boson in s-channel (not shown) and t-channel (top right). Without the mediation of the Higgs boson in longitudinal VBS (where both Z bosons have longitudinal polarization), the production cross section will increase with center of mass energy and violates unitarity, which indicates that the Higgs boson is an essential component of the SM theory. We consider these processes at order α^6 (for tree level) as EW production processes. The bottom right diagram in Fig. 1.20 is also at order α^6 and is an example of non-VBS EW production processes (see also [73]). As discussed in Ref. [74], this contribution needs to be included in the measurement as signal to preserve gauge invariance, but the VBS contribution can be enhanced by proper event selections.

The VBS process is characterized by two scattered jets with large rapidity separation typically emitted in the forward-backward region of the detector, originating from the two boson-emitting quarks, and this results in a region in rapidity (rapidity gap) where no additional hardronic activity from the hard scattering is expected. In addition, one or two vector bosons are produced centrally and decay into leptons, which in the case of ZZ production has low branching ratio but very clean signature. The kinematic features allow one to separate ZZ VBS events with the large background coming from the so-called QCD-induced ZZ events, which also have the ZZ+2jets signature and are at order $\alpha^4\alpha_S^2$ at tree level, with Fig. 1.17 top right

being an example Feynman diagram. Contrary to the EW ZZ+2jets analysis with signal and background defined above, in the ZZ+jets analysis, both the VBS and QCD-induced ZZ production events are treated as part of the signal.

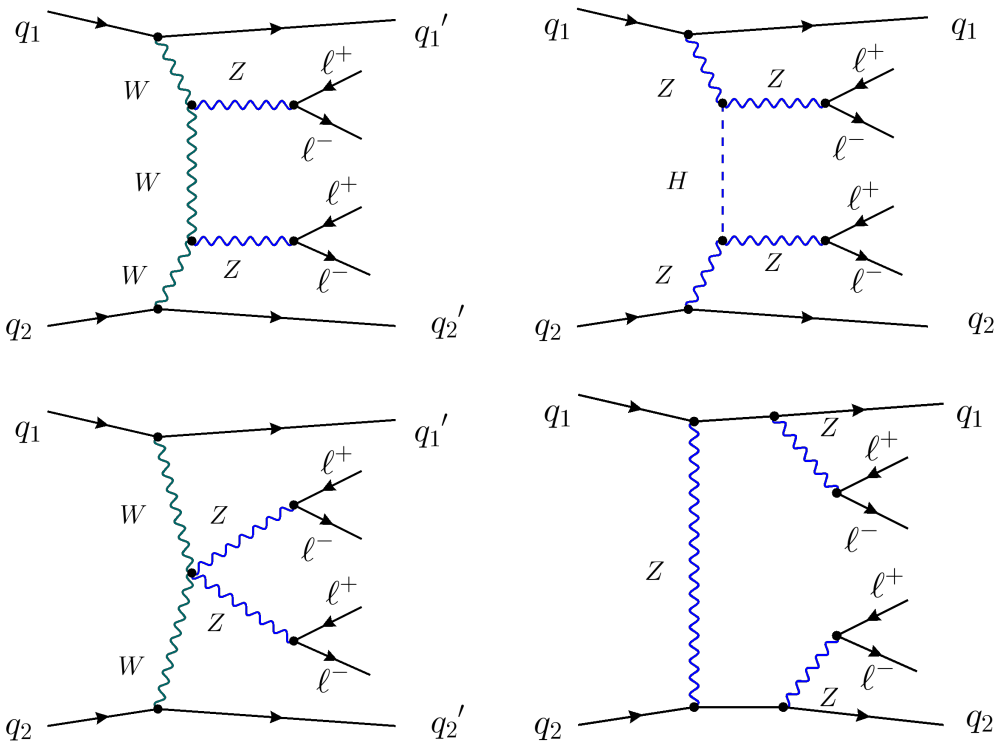

Figure 1.20: Example Feynman diagrams showing the EW production of diboson ZZ associated with two jets. The top row and the bottom left diagrams correspond to production through VBS, including diagrams containing the Higgs boson. The bottom right diagram is not associated with VBS but contributes to the EW ZZ+2jets final state.

1.6.6 Background Processes

The triboson production WWZ, WZZ, ZZZ, and in addition the $t\bar{t}Z$ production, all have decay channels that produce four genuine leptons (and additional jets if applies), therefore constituting a minor irreducible background. These processes are collectively referred to as VVV prodcution in the plots in this thesis and are mod-

eled with Monte Carlo simulation as discussed in Chapter 3. On the other hand, there are reducible background contributions due to nonprompt objects misidentified as leptons, which we discuss in detail in Section 5.1. As already mentioned previously, the QCD-induced ZZ+2jets production is the dominant background for EW ZZ+2jet production, whereas both production processes are treated as part of the signal in the ZZ+jets analysis. Note that by irreducible background, we mean that the background cannot be further reduced by using more stringent lepton identification requirements, since the four leptons in this background are genuine prompt leptons, but such background can still often be reduced by kinematic selections. The reducible background contributes to approximately 1-2% of the expected ZZ $\to 4\ell$ event yield, and the yields from the VVV processes with four prompt leptons are estimated from simulated samples to be around 1.0-1.5% of the expected ZZ $\to 4\ell$ yield.

1.7 Recent ZZ Measurements

Previously pairs of on-shell Z bosons, produced in the dilepton mass range 60–120 GeV, were studied by the CMS Collaboration using data sets with integrated luminosities of 5.1 fb^{-1}at $\sqrt{s} = 7$ TeV [75], 19.6 fb^{-1}at $\sqrt{s} = 8$ TeV [76, 77] in the ZZ $\to 2\ell 2\ell'$, ZZ $\to 2\ell 2\tau$ and ZZ $\to 2\ell 2\nu$ decay channels, where $\ell, \ell' =$ e or μ, and with integrated luminosities of 2.6 fb^{-1} [78] and 35.9 fb^{-1} [72] at $\sqrt{s} = 13$ TeV in the ZZ $\to 2\ell 2\ell'$ decay channel. The ZZ cross section was also measured at $\sqrt{s} = 5.02$ TeV based on the ZZ $\to 2\ell 2\ell'$ and ZZ $\to 2\ell 2\nu$ decay channels [79]. The differential cross sections for Z boson pair production in association with jets were measured at $\sqrt{s} = 8$ and 13 TeV with integrated luminosities of 19.7 and 35.9 fb^{-1}, respectively, using the ZZ $\to 2\ell 2\ell'$ decay channel [3], and we have used this analysis as an impor-

tant reference for analysis methods in the current ZZ+jets analysis. The most recent measurement of ZZ production cross sections with the full Run 2 data set with an integrated luminosity of $137\,\text{fb}^{-1}$at $\sqrt{s} = 13\,\text{TeV}$ performed by the CMS Collaboration was published in Ref. [2], and results on the EW production of ZZ+2jets were published in Ref. [4]. All measurements agree with SM predictions. The ATLAS Collaboration published similar results for inclusive ZZ production at $\sqrt{s} = 7$, 8, and 13 TeV [71, 80–84], which also agree with the SM, and reported observation of EW production of a Z boson pair associated with two jets [85]. These measurements are important to test predictions recently made available at next-to-next-to-leading order (NNLO) in QCD [70]. Comparing measurements at the highest collision energies to theoretical predictions tests the ability of the most advanced higher order QCD and EW calculations to predict the cross sections of complex final states with jets and multiple vector bosons, and the full Run 2 pp collision data at $\sqrt{s} = 13\,\text{TeV}$ allow diboson measurements at the highest energies and integrated luminosities to date.

The ZZ+jets analysis in this thesis is an extension to that of Ref. [2] with a focus on jet variables. The two analyses, Ref. [2] and the ZZ+jets analysis in this thesis (with results published in Ref. [1]), use the same events with a few minor differences: (i) a 0.3% update in the estimated luminosity of the data set; (ii) the simulation program MadGraph5_aMC@NLO [86] is used instead of POWHEG [87–90] as the nominal $\text{q}\bar{\text{q}} \to \text{ZZ}$ sample; and (iii) a regularized unfolding method is used instead of a simple matrix inversion. Therefore, the ZZ fiducial cross section measured in Ref. [2] is directly valid for the ZZ+jets analysis in this thesis. The results are also compared with recent nNNLO+PS predictions [5, 91], where PS stands for parton shower.

The EW ZZ+2jets analysis in this thesis serve as a cross check of the results in Ref. [4]. The ZZ+2jets analysis in this thesis uses a deep neural network appraoch

to extract the EW signal, whereas Ref. [4] uses a matrix element likelihood approach (MELA) (see references therein).

Chapter 2

The CMS Experiment and the CERN LHC

2.1 The CERN Large Hadron Collider

The Large Hadron Collider (LHC) [92–94] (we also refer to Ref. [32] for some general introduction and concepts) is currently the world's largest and most energetic particle collider. It is installed in an underground tunnel with a circumference of 27.6 km on the Franco-Swiss border near Geneva, Switzerland. A schematic plot of the CERN accelerator complex in 2019 is shown in Fig. 2.1, where LHC is the last element in the acceleration stages. The LHC has operated providing collisions to feed its physics program over three runs, with long shutdowns in between for collider and detector maintenance, and upgrades. In Run 1 from 2010 to 2012, the LHC operated at 7 TeV (2010-2011) and 8 TeV (2012) providing 6.1 fb^{-1} and 23.3 fb^{-1} of pp collision data, respectively, to the CMS experiment. In Run 2 from 2015 to 2018, the LHC increased the collision energy to 13 TeV and eventually more than doubled the peak luminosity providing 163.6 fb^{-1} of pp collision data to the CMS experiment. In Run 3, currently

in progress (since 2022), the LHC has increased the collision energy to 13.6 TeV and also increased the peak luminosity. The analyses in this thesis use data collected in the CMS experiment of LHC during 2016-2018 with an integrated luminosity of 138 fb^{-1}. We discuss further the concept and measurement of the luminosity later in this section and in Section 2.2.7.

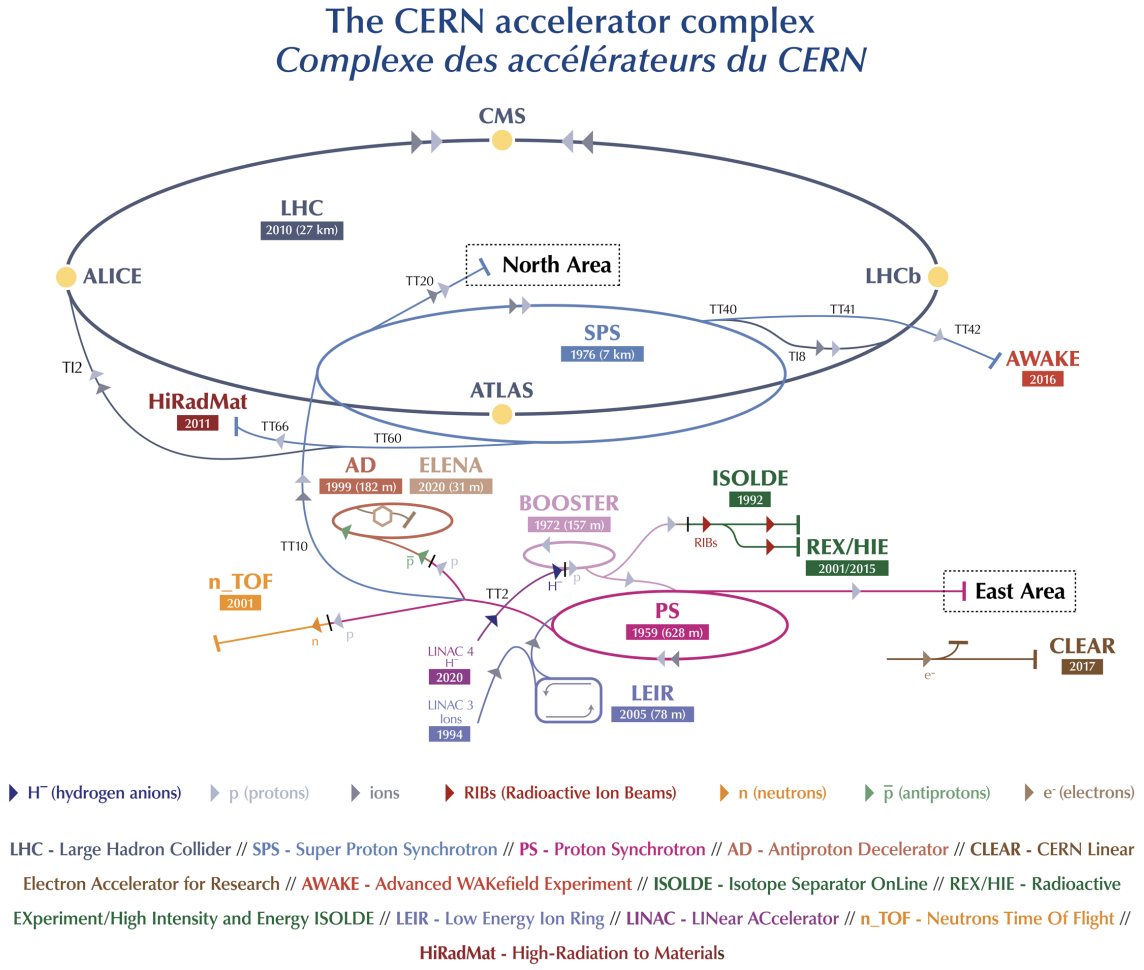

Figure 2.1: A schematic view of the CERN accelerator complex and LHC in 2019 from Ref. [95].

An overview of the machine and beam parameters during Run 2 is shown in Fig. 2.2. Filling the LHC with protons requires a range of pre-accelerators. During Run 2, protons are injected into the PS Booster (PSB) from LINAC2 (not shown in

Fig. 2.1) and accelerated by the booster to 1.4 GeV. The beam is then accelerated by the Proton Synchrotron (PS) and then by the Super Proton Synchrotron (SPS) to 450 GeV. After this the beam is finally transferred to the LHC and accelerated to 6.5 TeV. For more details related to proton injection one can refer to Ref. [94, 96].

Table 2: Overview of LHC Machine and Beam Parameters for Run 2 Compared to the Design Values

Parameter	**Design**	**2015**	**2016**		**2017**		**2018**
Beam type:	**Std**	**Std**	**Std/BCMS**	**BCMS**	**8b4e**	**8b4e-BCS**	**BCMS**
Energy [TeV]	7	6.5	6.5	6.5	6.5	6.5	6.5
Number of bunches per ring	2808	2244	2040/2076	2556	1916	1868	2556
Bunch spacing [ns]	25	25	25	25	25	25	25
Bunch population N_b [10^{11} p/b]	1.15	1.15	1.2	1.2	1.2	1.25	1.1
Transv. norm. emittance[4] ε_n [mm·mrad]	3.75	3.5	3.5/2.1	2.1	2.3	1.8	2
Betatron function at IP1 and IP5 β^* [m]	0.55	0.8	0.4	0.4	0.4/0.3	0.3	0.3/0.25[1]
Half crossing angle [µrad]	142.5	145	185/140	150	150	150/120[2]	160/130[2]
Peak luminosity [10^{34} cm^{-2}s^{-1}]	1	0.55	0.83/1.4	1.74	1.9	2.06/1.5[3]	2.1
Maximum pile up μ (per bunch crossing)	~20	~15	~20/35	~45	70/60[2]	80/60[3]	60
Stored beam energy [MJ]	360	270	345	320	240	245	320
Number days of physics operation	n.a.	88	146		140		145
Integrated luminosity per year [fb^{-1}]	n.a.	4.2	39.7		50.6		66

[1] Minimum betatron function during betatron anti-levelling
[2] Minimum crossing angle during crossing angle anti-levelling
[3] Value after luminosity-levelling by separation
[4] Value at maximum energy in stable beams

Figure 2.2: A table showing LHC machine and beam parameters from Ref. [97].

For a given process with a particular final state, the number of events produced, n, is given by:

$$n = \int \mathcal{L}\sigma \mathrm{d}t \tag{2.1}$$

where the instantaneous luminosity $\mathcal{L}$ is integrated over the time during which the events are recorded, and σ is the production cross section for the process. The luminosity of a collider is determined by the particle flux and geometry [98]. The instantaneous luminosity is given by

$$\mathcal{L} = \frac{N_1 N_2 n_b f_{\mathrm{rev}}}{A_{\mathrm{eff}}} \tag{2.2}$$

for head-on collisions as illustrated in Fig. 2.3, where N_1 and N_2 are the number of particles per bunch in beam 1 and 2, n_b is the number of bunches, f_{rev} is the revolution frequency and A_{eff} is the effective beam overlap cross section at the interaction point.

The main method to achieve high luminosity in LHC is to use a large number of bunches (as in Fig. 2.2), and to reduce the transverse beam size at the interaction points by manipulations of the magnetic focusing system (squeezing of the beams). The unit for cross section commonly used in particle physics is the barn (symbol b) defined as $10^{-24}\,\mathrm{cm}^2$. Cross sections of production processes involving heavy SM particles are typically of the order of nanobarns (nb), picobarns (pb), or femtobarns (fb).

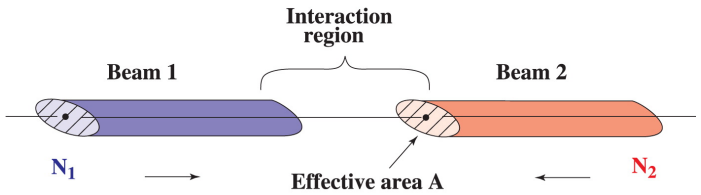

Figure 2.3: A figure from Ref. [93] illustrating luminosity from particles flux and geometry.

As discussed in Ref. [32], not all events produced are observed due to limitations in the acceptance and efficiency of the detectors, and therefore, if N signal events are observed, then the cross section is given by

$$\sigma = N/\int(\mathcal{L}A\epsilon)dt \tag{2.3}$$

The acceptance A is the fraction of events in which the kinematics of the final state particles are such that they traverse, or impact, a detector with the capability to measure them. The efficiency ϵ is the fraction of events within the acceptance that are detected. We frequently measure what is called a "fiducial" cross section, which is the part of the cross section that corresponds to a defined set of kinematic requirements on the final-state particles for which the acceptance is high. Measuring fiducial cross sections eliminates theoretical uncertainties related to the extrapolation from the fiducial phase space to the full phase space.

2.2 The Compact Muon Solenoid Detector

2.2.1 Overview

The CMS apparatus [99, 100] is a multipurpose, nearly hermetic detector, designed to trigger on [101–103] and identify electrons, muons, photons, and (charged and neutral) hadrons [104–106]. A schematic view of the CMS detector is shown in Fig. 2.4. Its central feature is a superconducting solenoid of 6 m internal diameter, providing a magnetic field of 3.8 T. Within the solenoid volume are a silicon pixel and strip tracker, a lead tungstate crystal electromagnetic calorimeter (ECAL), and a brass and scintillator hadron calorimeter (HCAL), each composed of a barrel and two endcap sections. Forward calorimeters extend the pseudorapidity coverage provided by the barrel and endcap detectors. Muons are reconstructed using gas-ionization detectors interleaved with the layers of the steel flux-return yoke outside the solenoid. More detailed descriptions of the CMS detector, together with a definition of the coordinate system used and the relevant kinematic variables, can be found in Refs. [99, 100]. See also [107] for relevant detector terminologies. In the following subsections we will discuss different parts of the CMS detector in more detail using information from references mentioned above and other references cited during the discussion.

2.2.2 Inner Tracking System

The inner tracking system (CMS tracker) is directly outside the beam pipe containing the pp collision interaction point, and is designed to precisely measure the trajectories of the charged particles from the collision, as well as precisely reconstruct the secondary vertices, which also requires fast response and hardness under high radiation. It consists of the pixel detector and strip tracker both made of silicon, and is designed to have excellent momentum resolution and tracking efficiency. The track-

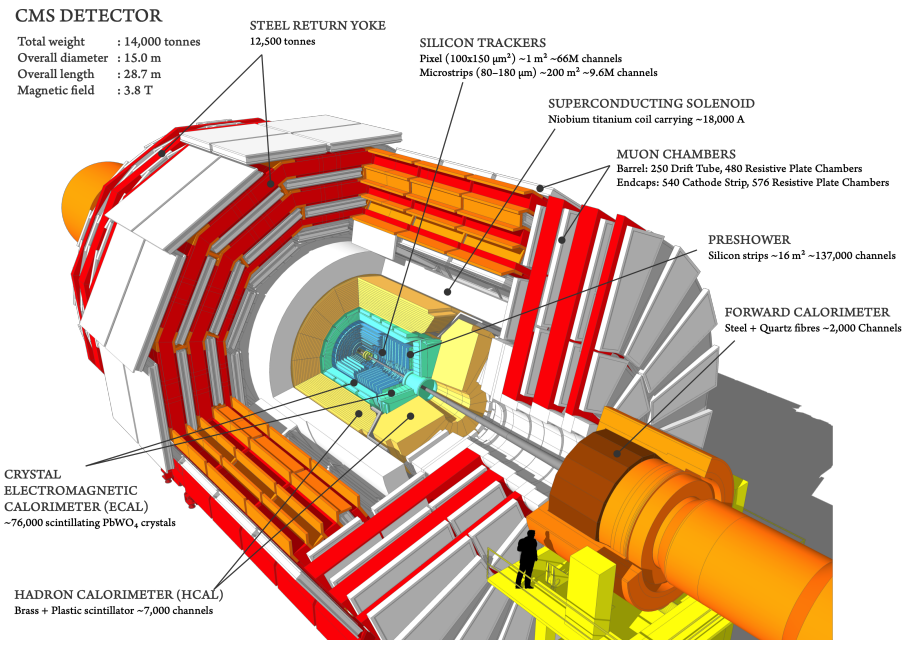

Figure 2.4: A schematic view of the CMS detector in LHC Run 1 geometry from Ref. [108].

ing system used in 2016 measured charged particles within the range $|\eta| < 2.5$. For nonisolated particles of $1 < p_{\mathrm{T}} < 10\,\mathrm{GeV}$ and $|\eta| < 1.4$, the track resolutions were typically 1.5% in p_{T} and 25–90 (45–150) μm in the transverse (longitudinal) impact parameter [106]. At the start of 2017, a new pixel detector was installed [109]; the upgraded tracker measured particles up to $|\eta| = 3.0$ with typical resolutions of 1.5% in p_{T} and 20–75 μm in the transverse impact parameter [110] for nonisolated particles of $1 < p_{\mathrm{T}} < 10\,\mathrm{GeV}$. According to simulation studies [111], similar improvements are expected in the longitudinal direction.

The pixel detector is the innermost part that provides high-precision three-dimension space points closest to the LHC collision point, with a pixel cell size of $100 \times 150\,\mu\mathrm{m}^2$. Originally during Run 1 it consisted of three barrel layers and two endcap disks on each end, providing three-point tracking for charged particles. As mentioned, its current form is an improved version installed in 2017 and referred to as the Phase 1

pixel detector. It now consists of four barrel layers (L1-L4) at radii of 29, 68, 109, and 160 mm, and three disks (D1-D3) on each end at distances of 291, 396, and 516 mm from the center of the detector, providing four-hit coverage for tracks up to $|\eta| = 3$, with a total of 124 million readout channels. The layout of the pixel detector is shown in Fig. 2.5. With the upgraded version of pixel detector, the hit efficiencies, which are the detection efficiency for a particle traversing a sensor, are measured to be [100] 97, 98, 99, and 99.5% for BPIX L1-L4, respectively, and average to 99% for the pixel forward disks (FPIX) when measured at an instantaneous luminosity of $2 \times 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. The position resolution is measured to be [100] 11.0 μm in the r-ϕ direction and 24.3 μm in the z direction for BPIX L3, and somewhat worse for BPIX L1 and L2 due to higher radiation damage. For the forward disks, the resolution is measured to be 11.9 μm in the r direction and 21.0 μm in the z direction.

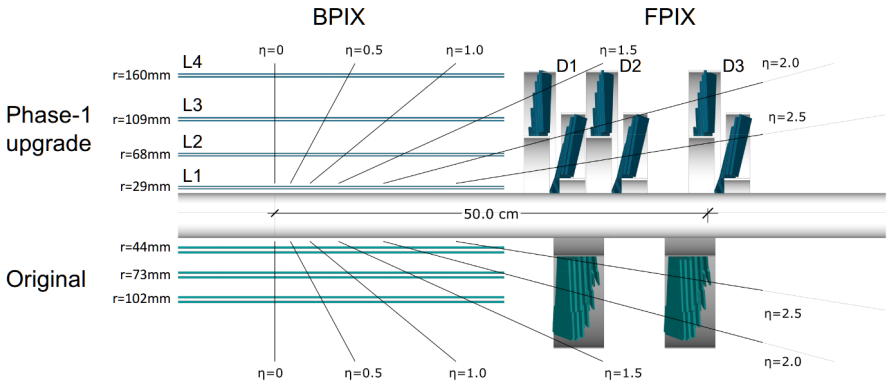

Figure 2.5: Longitudinal view of the CMS Phase-1 pixel detector layout compared to the original layout, from Ref. [109].

The silicon strip tracker (SST) contains ten layers of silicon strip sensors in the barrel, with four layers in the tracker inner barrel (TIB) and six layers in the tracker outer barrel (TOB), complemented with three plus nine disks in the endcap on each side. A schematic view is shown in Fig. 2.6. The SST has 9.3 million silicon microstrips and 198 m^2 of active silicon area distributed over 15148 modules. The strip

sensors are segmented into long, thin strips that provide a typical hit resolution values between 20 and 40 μm measured during LHC Run 2 [112]. The average hit efficiency measured under typical Run 2 conditions, at an average instantaneous luminosity of $1.11 \times 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ is about 99.5% depending on the layer [100].

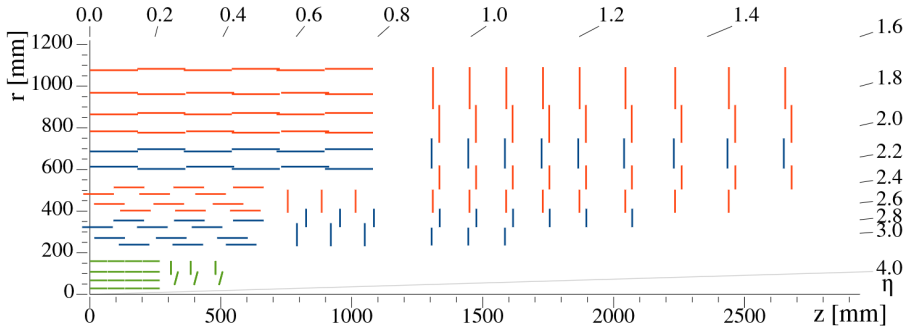

Figure 2.6: Schematic view of one quadrant in the r-z view of the CMS tracker: single-sided and double-sided strip modules are depicted as red and blue segments, respectively. The pixel detector is shown in green. The figure is from Ref. [113].

2.2.3 Electromagnetic Calorimeter

The electromagnetic calorimeter is a hermetic, homogeneous calorimeter made of 75848 lead tungstate ($PbWO_4$) crystals [100]. It is placed outside the inner tracking system and used to identify and measure the energy and position of electrons and photons, as well as measure their arrival time at the crystals which can be used to search for long-lived particles. Algorithms that constrain clusters to the size and shape expected for photons and electrons are used to reconstruct the positions and energies from energy deposits. The ECAL also has important contribution in reconstructing jets and missing transverse energy.

The high density ($8.28\,\mathrm{g/cm}^3$), short radiation length (0.89 cm) and crystal size compatible with the small Molière radius (2.2 cm) [99] result in high granularity and compact size of the calorimeter. It consists of the barrel (EB) and two endcaps (EE)

that provides a pseudorapidity coverage of $|\eta| < 3$. A preshower detector is placed in front of the endcap crystals for π^0 rejection. Avalanche photodiodes (APDs) and vacuum phototriodes (VPTs) are used as photodetectors in the barrel and in the endcaps, respectively.

The energy resolution is affected by pileup, noise, and accuracy of the calibration, in relative order of decreasing importance. The LHC Run 2 resolution is shown in Fig. 2.7 both for electrons with low bremsstrahlung emissions and for an inclusive selection of electrons. The stability of the energy scale is monitored as a function of time and found to be within 0.1% in the barrel and a few times 0.1% in the endcaps. The electron momentum is estimated by combining the energy measurement in the ECAL with the momentum measurement in the tracker. The momentum resolution for electrons with $p_{\mathrm{T}} \approx 45\,\mathrm{GeV}$ from $\mathrm{Z} \to \mathrm{ee}$ decays ranges from 1.6 to 5%. It is generally better in the barrel region than in the endcaps, and also depends on the bremsstrahlung energy emitted by the electron as it traverses the material in front of the ECAL [104, 114].

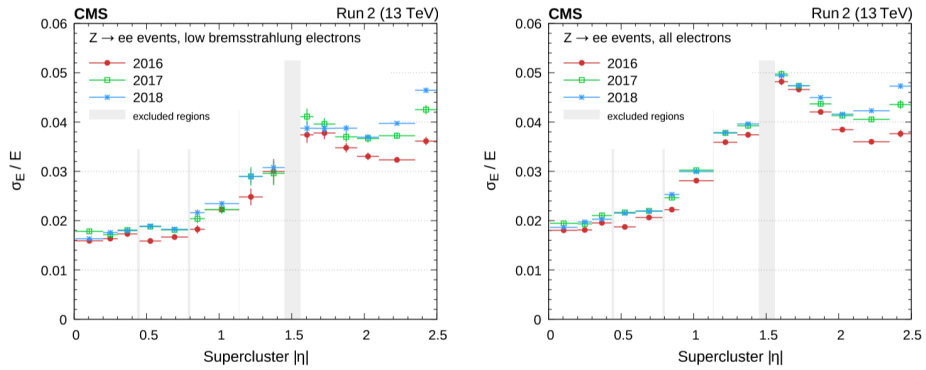

Figure 2.7: ECAL resolution for electrons having low bremsstrahlung emissions (left) and for an inclusive selection of electrons (right). The horizontal bars show the bin width. Figure is from Ref. [100].

2.2.4 Hadronic Calorimeter

The hadron calorimeter is designed to measure the energy of charged and neutral hadrons. It is important for identification of hadrons and the measurement of their properties, the reconstruction of jets and missing transverse momentum, and also contributes to the identification of electrons and photons. It is designed to have a good hermeticity, with the ability to detect hadrons in nearly the full 4π solid angle. When combining information from the entire detector, the jet energy resolution amounts typically to 15–20% at 30 GeV, 10% at 100 GeV, and 5% at 1 TeV [115].

The HCAL consists of four subdetectors: the hadron barrel (HB), hadron endcap (HE), hadron outer (HO), and hadron forward (HF) calorimeters. The HB and HE are located inside the CMS solenoid magnet and surround the ECAL. They cover the ranges $|\eta| < 1.392$ and $1.305 < |\eta| < 3.0$, respectively, and are sampling calorimeters primarily made of layers of brass plates absorbers interleaved with layers of plastic scintillating tiles. The HF is a sampling calorimeter constructed from steel absorber and quartz fibers that produces Cherenkov light detected by photomultipliers. It is located outside the solenoid at ± 11.5 m from the collision point, covering the $3.0 < |\eta| < 5.2$ range. Since the combined stopping power of the EB and HB does not provide sufficient containment for hadron showers, the HO, located outside the solenoid and utilizing the solenoid coil as an additional absorber, provides a measurement of the shower tails for $|\eta| < 1.26$.

2.2.5 Muon System

The muon system is used to identify muons and measure their momenta. As shown in Fig. 2.8, it consists of four subsystems: the drift tubes (DTs), the cathode strip chambers (CSCs), the resistive-plate chambers (RPCs), and the recently added gas

electron multiplier (GEM) detectors which were not used for LHC Run 2 data taking. The high redundancy of the muon system is a central feature of the design and is responsible for the robustness of the muon triggering and reconstruction in CMS. Muons are measured in the pseudorapidity range $|\eta| < 2.4$. In LHC Run 2, the single muon trigger efficiency exceeds 90% over the full η range, and the efficiency to reconstruct and identify muons is greater than 96%. Matching muons to tracks measured in the silicon tracker results in a relative transverse momentum resolution, for muons with p_{T} up to 100 GeV, of 1% in the barrel and 3% in the endcaps. Measurements made with cosmic ray muons show that, in the central region of the detector, the p_{T} resolution is better than 7% for muons with p_{T} up to 1 TeV [105].

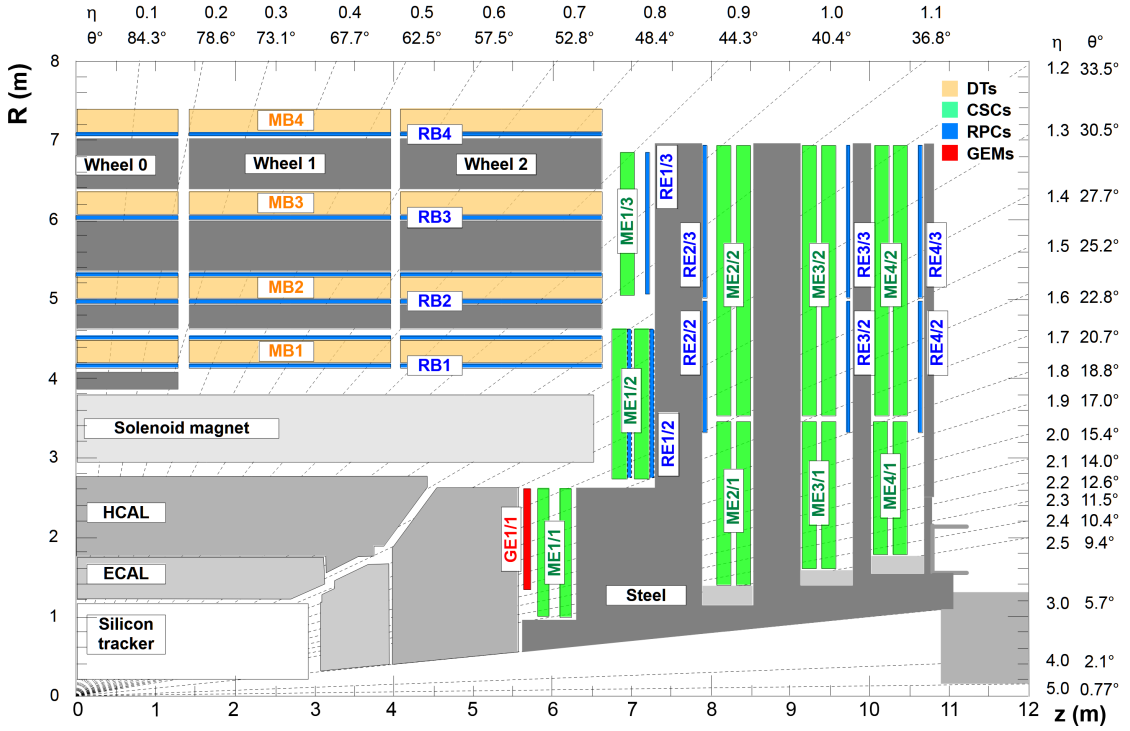

Figure 2.8: Schematic view in the r-z plane of a CMS detector quadrant at the start of LHC Run 3. Figure from Ref. [105]. As discussed in the text, the GE1/1 GEM chambers were only partially installed during LHC Run 2 [116], and were fully installed in 2020 during LHC Long Shutdown 2 period [117].

The DTs are the tracking detectors covering the range $|\eta| < 1.2$ for the barrel

muon system where there is lower expected muon rate and relatively lower local magnetic field strength [118]. They provide precise spatial and time measurements with sufficient resolution to measure transverse momentum and identify muons for both standalone triggering and offline reconstruction. The DTs consist of chambers formed by multiple layers of long rectangular tubes that are filled with a mixture of 85% Ar and 15% CO_2 gases [100]. An anode wire is located at the center of each tube, whereas cathode and field-shaping strips are positioned on its borders. This combination creates an electric field that induces an almost uniform drift of ionization electrons when charged particles traverse the gas. The charged particle's trajectory is determined from the arrival time of the currents on the anode wires of the readout. The performance of the DT system measured over LHC Run 1 and Run 2 was found to be remarkably stable and in line with the design expectations. It is documented in Refs. [105, 118], and summarized in Ref. [100]. Track segments are typically reconstructed offline with an efficiency above 99%, and they are characterized by spatial and time resolutions around 100 μm and 2 ns, respectively. The efficiency to reconstruct a standalone DT segment in the trigger and to correctly identify its bunch crossing is above 95%. The position (direction) resolution of the DT trigger segments is approximately 1 mm (5 mrad).

The CSC system in the endcap consist of multiwire proportional chambers (MWPCs) with orthogonal cathode strips, covering the range $0.9 < |\eta| < 2.4$. Signals are generated on both anode wires and cathode strips. Each CSC chamber consists of six gas layers with a mixture of about 40%Ar, 50%CO_2, and 10%CF_4, with each layer having a plane of radial cathode strips to measure the muon position in the plane perpendicular to the colliding beams axis, and a plane of anode wires running perpendicular to the strips, providing a coarse measurement in the radial direction. A precise measurement of the muon coordinate in the azimuthal direction is obtained

from charges induced on the cathode strips. Because of the higher particle flux in the endcap region, the CSCs are designed to have a faster response time than the DTs. The finely segmented cathode strips and fast readout electronics provide good timing and spatial resolution to trigger on and identify muons. The typical position and time resolutions are 50-140 μm, depending on the chamber type, and 3 ns per chamber, respectively [100].

The RPCs are double-gap chambers operated in avalanche mode to handle the high background rates with fast response. They consist of two detecting layers of high-pressure laminate (HPL) plates that are separated by a thin gap filled with a gas mixture of 95.2% $C_2H_2F_4$, 4.5% i-C_4H_{10}, and 0.3% SF_6 [100]. The electronic readout strips are located between the two layers, and a high voltage is applied to high-conductivity electrodes coated on each plate. The RPCs are located in both the barrel and endcap regions, and they complement the DTs and CSCs with a very fast response time that can be used to unambiguously identify the bunch crossing corresponding to a muon trigger candidate, with a time resolution of 1.5–2 ns [119].

As the first of three GEM stations to be installed for the HL-LHC, the GE1/1 GEM chambers were partially installed in the form of slice test in 2017 during LHC Run 2 [116], and fully installed in 2020 during LHC Long Shutdown 2 (LS2) period [117]. The second and third stations, GE 2/1 and ME0, will be installed in the third Long Shutdown (LS3) period and during the Run 4 technical stops [120]. The three stations will provide redundancy, improved resolution, and extended coverage for muon triggering during the HL-LHC. Our following discussion focuses on the GE 1/1 chambers. They are located in front of the inner ring of CSC chambers in the first endcap station, complementing the CSCs in a region of high particle flux in preparation for increased hit rate in HL-HLC. The GEM is typically operated with a gas mixture of 70% Ar and 30% CO_2 [100, 120]. The key feature of the GEM is

having three foils which consist of a perforated insulating polymer surrounded on the top and bottom by conductors, separating out four gas gaps. A voltage difference is applied on the foils to produce a strong electric field in the holes. When the gas volume is ionized, electrons are accelerated through the holes and read out on the thinly separated strips. This structure allows for high amplification factors at the order of up to several 10^5 with modest voltages that provide good timing (< 10 ns per layer) and spatial resolution (at the order of 100 μm for combined GE1/1) [100], and can be operated at high hit rates.

2.2.6 Trigger and Data Acquisition System

Events of interest are selected using a two-tiered trigger system. The first level (L1), composed of custom hardware processors, uses information from the calorimeters and muon detectors to select events at a rate of around 100 kHz within a fixed latency of 4 μs [101]. The second level, known as the high-level trigger (HLT), consists of a farm of processors running a version of the full event reconstruction software optimized for fast processing, and reduces the event rate to a few kHz before data storage [102, 103].

The L1 trigger consists of electronics responsible for making a fast selection of events based on the presence of high-energy particles in the detector. To improve the event selection in preparation for LHC Run 2, the L1 trigger hardware was entirely upgraded during the first LHC long shutdown (LS1) and the transition to the new L1 trigger was completed prior to 2016 [100, 101]. The upgraded version of the L1 trigger, shown in Fig. 2.9, was used for the Run 2 data taking during 2016–2018. It receives energy, position, timing and quality information, so-called trigger primitives (TPs), from the calorimeters and the muon detectors. The calorimeter trigger consists of two layers: Layer-1 receives, calibrates, and sorts the local energy deposits (TPs) which

are sent to the trigger by the ECAL and HCAL; Layer-2 uses these calibrated TPs to reconstruct and calibrate the physics objects such as electrons, tau leptons, jets, and energy sums. In the muon trigger system, signals from the DTs, CSCs, and RPCs are combined into TPs to reconstruct muons and measure their p_{T}. This system includes three muon track finders (MTF) which reconstruct muons in the barrel (BMTF), overlap (OMTF), and endcap (EMTF) regions of the detector, and the global muon trigger (μGMT) for final muon selection. The physics objects and candidates from the calorimeter and muon trigger systems are provided to the global trigger (μGT). At the μGT, the trigger decision is generated, based on the multiplicity and kinematic information of the various candidate trigger objects. The trigger configuration is implemented in a trigger menu comprised of several hundred seed algorithms, and the μGT executes every algorithm in the menu in parallel for the final trigger decision. Upon a positive GT decision, the full detector data are read out for further filtering in the HLT.

The HLT is a software-based system running on a farm of commercial computers, and the HLT hardware consists of a large cluster of multi-core servers, the event filter farm, that runs a Linux operating system. It further refines the purity of the collection of physics objects that are selected at L1, with an input event rate to the HLT limited to about 100 kHz by the detector read-out electronics, and targets an average output rate of about 1 kHz during Run 2 for standard pp collision events for offline storage and prompt reconstruction. Additional storage beyond the 1 kHz rate is allowed for data to be "parked", whereby the offline reconstruction is postponed until a later, non-data-taking period. The data processing of the HLT is structured around the concept of an HLT "path", which is a set of algorithmic processing steps run in a predefined order that both reconstructs physics objects and makes selections on these objects based on the physics requirements. Each HLT path is implemented as

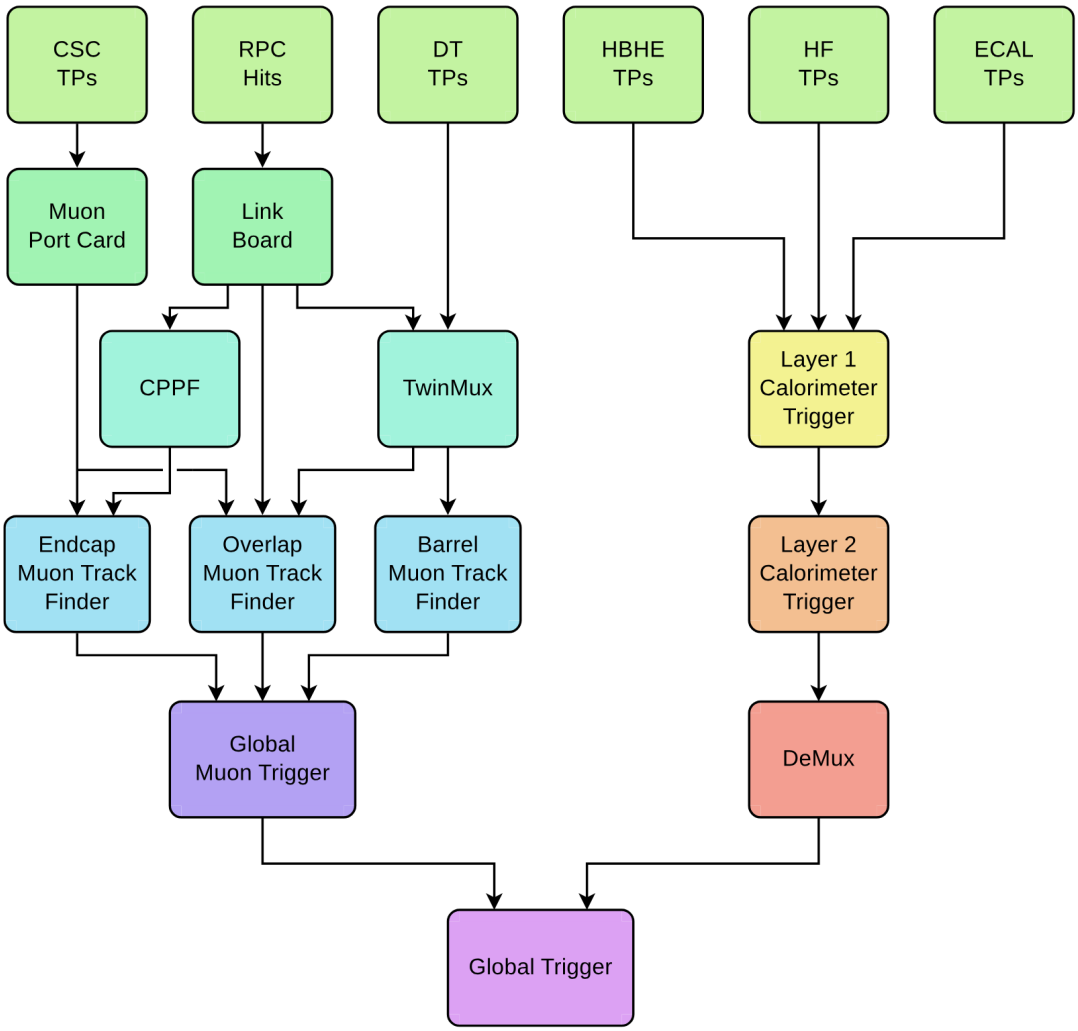

Figure 2.9: Diagram of the upgraded CMS Level-1 trigger system during Run 2. Figure from Ref. [101]. Labels in the diagram correspond to trigger primitives (TPs), concentration preprocessing and fan-out (CPPF), demultiplexing card (DeMux), and other shorthands already introduced in the previous sections.

a sequence of steps generally of increasing complexity, reconstruction refinement, and physics sophistication. Each path also requires the selections in specific L1 triggers ("L1 seeds") to have been satisfied before execution would begin. The HLT paths selecting similar physics object topologies are grouped into primary data sets that define the samples used for offline processing, which are then grouped into streams that define the outputs of the HLT processes.

The data acquisition (DAQ) system [100, 102, 121] is responsible for the readout of all detector data for events accepted by the L1 trigger, the building of complete events from subdetector event fragments, the operation of the filter farm cluster running the HLT, and the transport of event data selected by it to the permanent storage in the Tier 0 computing center. The DAQ also includes the trigger control and distribution system (TCDS) [122] newly developed in Run 2, which distributes timing and control (synchronization) data flowing to the detector front-ends and receives back status information related to the readiness of the detector systems to handle more triggers.

2.2.7 Luminosity Determination

The discussion in this subsection mainly follows from Ref. [123, 124] regarding LHC Run 2 luminosity determination. One can refer to these references for the details of detectors, measurement methods and algorithms. The luminosity $\mathcal{L}$ is a key parameter at particle colliders, and along with the available center of mass energy, it quantify the potential of delivering sufficiently large data sample and producing novel massive particles as well as events of interest. As discuss in the context for eqn. (2.1), the instantaneous luminosity $\mathcal{L}(t)$ is the process-independent ratio of the rate $R(t)$ of events produced per unit of time dt to the cross section σ for a given process:

$$R(t) = \mathcal{L}(t)\sigma \tag{2.4}$$

The precise determination of the integrated luminosity, $\int \mathcal{L}(t)dt$, has proven particularly challenging at hadron colliders, with an achieved precision typically ranging from 1 to 15% [125]. The "precision frontier" target of 1% [126] does not reflect a fundamental limitation, but rather results from a variety of uncorrelated sources of systematic uncertainty with typical magnitudes of 0.1–0.5%. A good understanding of the luminosity is critical to minimize the systematic uncertainty in important SM measurements such as cross sections for the production of EW gauge bosons or top quark pairs. The uncertainty in the luminosity measurement is often the dominant systematic uncertainty (see references in Ref. [123] for example), which also motivates an effort to improve its precision.

The knowledge of the integrated luminosity requires stability over long periods of time, and hence benefits greatly from redundant measurements whose combination can lead to an improved precision. With this consideration, A number of systems are used for luminosity measurement, some of which utilize the existing parts of the CMS detector, and several upgrades were completed during LS1. The main luminosity subdetectors (luminometers) in LHC Run 1 were the silicon pixel detector and the HF. The HF back-end electronics, which were upgraded during LS1, consist of two independent readout systems: a primary readout over optical links for physics data taking, and a secondary readout using Ethernet links, explicitly reserved for luminosity data. In addition, two other dedicated luminometers were designed, constructed, and commissioned: the Pixel Luminosity Telescope (PLT) [127] and the Fast Beam Conditions Monitor (BCM1F) [128]. The DT already discussed is also used for luminosity measurement. Finally, the Radiation Monitoring System for the Environment and Safety (RAMSES) detectors [129, 130] are part of the LHC environmental protection and monitoring systems, but also can be used to provide a luminosity measurement read out through Timber, the operational data and LHC

logging service. An overview of the position of these luminometers within CMS in Run 2 is shown in Fig. 2.10. A separate DAQ system, called BRILDAQ, was developed that is independent of the central DAQ system, so that HF, PLT, and BCM1F data, as well as LHC beam-related data, are collected and stored in a time- rather than event-based manner, and luminosity measurements can be provided regardless of the current operating state of the rest of CMS.

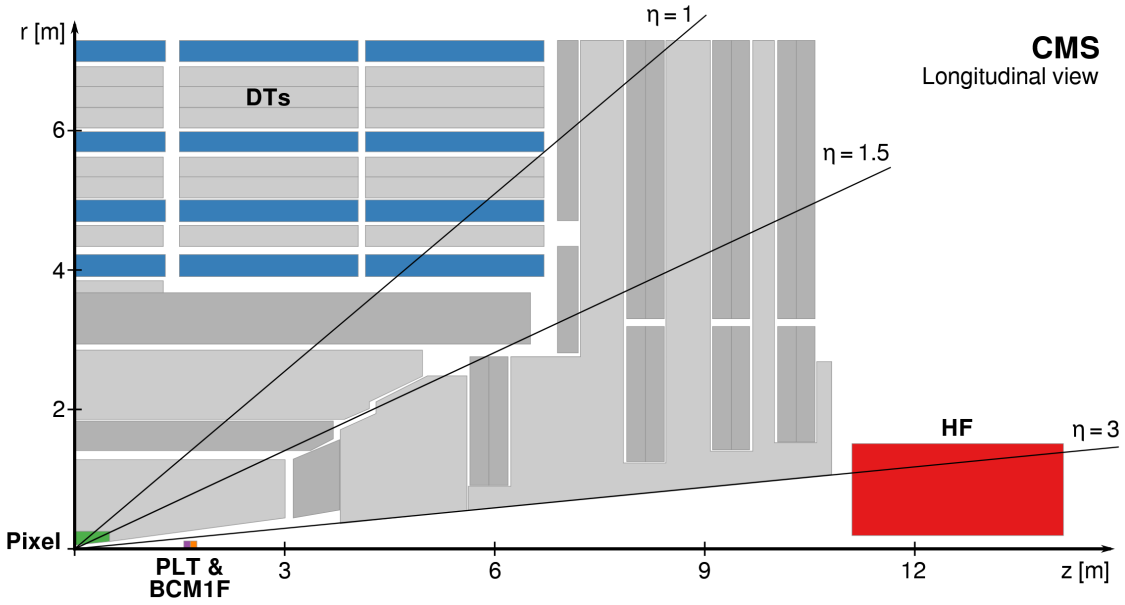

Figure 2.10: Schematic cross section through the CMS detector in the r-z plane. Figure is from Ref. [123]. The main luminometers in Run 2, as described in the text, are highlighted, showing the silicon pixel detector, PLT, BCM1F, DTs, and HF. The two RAMSES monitors used as a luminometer in Run 2 are located directly behind HF. In this view, the detector is symmetric about the horizontal and vertical axes, so only one quarter is shown here. The center of the detector, corresponding to the approximate position of the pp collision point, is located at the origin. Solid lines represent distinct η values.

Each luminometer reads out a rate of the specific quantities observed in the detector (hits, tracks, clusters, etc.). This rate, R_{read}, should be proportional to the instantaneous luminosity, $\mathcal{L}(t)$, with the constant of proportionality given by the cross section σ_{vis} in the "visible" phase space region, defined by its acceptance. Similar to

eqn. (2.4) we have:

$$R_{\text{read}}(t) = \mathcal{L}(t)\sigma_{\text{vis}} \tag{2.5}$$

The determination of σ_{vis} is carried out through van der Meer (VdM) scans [131–134] performed with a dedicated LHC machine setup, and the measured σ_{vis} can be used to determine the luminosity in regular physics fills. The integrated luminosities for the 2016, 2017, and 2018 data-taking years have 1.2, 2.3 and 2.5% individual uncertainties [123, 124, 135], while the overall uncertainty for the 2016–2018 period is 1.6%.

Chapter 3

Simulation

In this chapter we talk about various aspects of event simulation, including general Monte Carlo (MC) event generation method, matching the fixed-order calculation and parton shower that are already discussed in Section 1.5, and contents related to the detector simulation in CMS. We use the same references as at the beginning of Chapter 1. In addition we also refer to [61, 136–138] for specific NLO+PS and NNLO+PS matching methods, along with the references mentioned in the main discussion.

3.1 Monte Carlo Event Generation

As discussed in Ref. [61], in particle physics simulation, an event is referred to as the outcome of a particle collision or of isolated particle decay, which consists of a set of outgoing particles with conservation laws imposed. Due to the probabilistic nature of quantum processes, such outcome is different from event to event. An event generator is an algorithm that can produce random sequences of such simulated events. Since it can produce events before any interaction with detector (particle-level event), it is very useful for various studies such as those on detector effects by comparison

with the same set of events after full detector simulation, and for interpretation of experimental data.

Monte Carlo event generation can be thought of as sampling over the phase space using proper algorithms (that, for example, efficiently account for possible sharp peaks in the probability distribution), based on (pseudo-)random number generation. In high energy experiments like CMS, typically the hard process is generated with matrix element generators, and then interfaced into general-purpose Monte Carlo (GPMC) generators like PYTHIA8 [63] for subsequent processes such as parton shower and hadronization. An event is typically presented in the form of a list of particle properties, along with a measure of the probability for the event to occur. The relative probability of an event with respect to the whole sample is given by its weight over the total sum weight of the sample. The total cross section of the sample is also computed using multi-dimensional Monte Carlo integration, typically with a complicated integrand function, and the uncertainty decreases with the number of events generated.

Generated events are typically also unweighted so that each event has the same weight (up to a sign in the case with negative weights present), and the numbers of specific types of events generated are proportional to their cross sections, similar to how it works in nature. Unweighting can be performed using the acceptance-rejection method [18, 26].

As discussed in [61], currently the full event properties of high-energy collisions and decay processes cannot be predicted solely from first principle calculation, and within an event generator, there are many different complex phenomena (likely related) that need to be described by a proliferation of models each focusing on a limited dynamical range. Therefore, the predictions from a typical event generator can depend on a large set of parameters whose values are inferred from comparison with data. We refer to a

collection of such parameter values as a tune, and refer to the process of constraining these parameters with experimental data as generator tuning.

3.2 Matching Fixed-order Calculation and Parton Shower

Continuing our discussion in Section 1.5, matrix elements from fixed-order calculation and the parton shower that incorporates resummation can provide good description in complementary phase space regions for events with QCD emissions. Fixed-order calculation excels in describing additional emissions that are energetic and at large angles, but typically fails to describe softer and more collinear emissions due to the associated large logarithms that can eventually break the perturbative description. On the other hand, parton shower is good at describing such emissions on the soft and collinear end, but typically cannot capture the complicated patterns in hard emission. Due to unitarity, parton shower does not add higher-order corrections to the total cross section. Though the cross section can be modified with a global K-factor, this typically does not account for the difference in emission patterns. Therefore, sophisticated methods are needed to combine the good features from fixed-order calculation and parton shower. In general, we want to maintain the fixed-order accuracy of the cross section, and of the hardest emissions, but supplement these hardest emissions with the leading logarithms captured by parton shower. Other softer or more collinear emissions still need to maintain the LL accuracy from parton shower. An intrinsic problem to resolve for combining fixed-order calculation and parton shower is the double-counting of additional emissions associated with the two methods, which can have both contributions wrongly added or have both contributions missing [12].

In the following discussion, we will talk about different methods matching fixed-

order calculation and parton shower, focusing on NLO+PS and NNLO+PS methods related to the MC samples discussed in Section 3.4, which are samples used in the physics analyses in this thesis.

3.2.1 Matrix Element Correction and ME+PS

The first matching approach we discuss, the matrix element correction (MEC) method, improves the parton shower by correcting the kinematics of the hardest emission at order α_S by the exact matrix element corresponding to real emission, but doesn't correct the total cross section to NLO, which is done by the NLO+PS methods discussed in the next subsections. In this subsection we also review and introduce notations that will be used in later discussion.

First we recall eqn. (1.129) for the probability of additional emission in collinear limit,

$$\left|\mathcal{M}_{n+1}\right|^2 d\Phi_{n+1} \approx \left|\mathcal{M}_n\right|^2 d\Phi_n \frac{dt}{t} dz \frac{d\phi}{2\pi} \frac{\alpha_S}{2\pi} P_{a,bc}(z) \tag{3.1}$$

and the Sudakov form factor in eqn. (1.135) for probability of no resolvable emission (written in a form without summing over the splitting partons):

$$\Delta\left(t_m, t_n\right)_{a,bc} = \exp\left[-\int_{t_m}^{t_n} \frac{dt}{t} \frac{\alpha_S\left(t\right)}{2\pi} \int_{z_{min}(t)}^{z_{max}(t)} dz \int \frac{d\phi}{2\pi} P_{a,bc}(z)\right] \tag{3.2}$$

When stressing the general nature of the parton shower, we introduce the parton shower splitting kernel $\mathcal{K}_{ij;k}\left(t, z, \phi\right)$ to write the general form of the Sudakov form factor as

$$\Delta^{(\mathcal{K})}_{ij;k}(t_m, t_n) = \exp\left[-\int_{t_m}^{t_n} \frac{dt'}{t'} \int dz \int \frac{d\phi}{2\pi} J\left(t', z, \phi\right) \frac{\alpha_S}{2\pi} \mathcal{K}_{ij;k}\left(t', z, \phi\right)\right] \tag{3.3}$$

where $J\left(t', z, \phi\right)$ is the Jacobian that is included when necessary, and is implicit and suppressed from now on. The form of $\mathcal{K}_{ij;k}$ depends on the parton shower implementation, but in the collinear limit reduces to the splitting function $P_{a,bc}(z) = P_{(ij),ij}(z)$

presented in Chapter 1, and in the soft limit reproduces the eikonal form proportional to $\frac{1}{p_i p_j} \cdot \frac{p_i p_k}{p_j p_k}$ as in eqn. (1.141) and (1.142), with subtle details as discussed in Ref. [12]. In our notations here, the pair (ij) is taken as an emitter that splits into parton i and j, and k is referred to as the spectator parton that accounts for the recoil. We denote the one-parton emission phase space as Φ_1 with typically

$$d\Phi_1 = dtdz\frac{d\phi}{2\pi} \tag{3.4}$$

but we also use the notations $t\left(\Phi_1\right)$, $z\left(\Phi_1\right)$ etc. to indicate quantities associated with the general emission phase space Φ_1, which doesn't necessarily use (t, z, ϕ) as the variables. The combined splitting kernel and Sudakov form factor from n-body state are then given by

$$\mathcal{K}_n\left(\Phi_1\right) = \sum_{\{ij;k\}\in n} \mathcal{K}_{ij;k}\left(\Phi_1\right), \quad \Delta_n^{(\mathcal{K})}(t_m, t_n) = \prod_{\{ij;k\}\in n} \Delta_{ij;k}^{(\mathcal{K})}(t_m, t_n) \tag{3.5}$$

The MEC method was first formulated in [139] and the later developments are summarized in [140]. It is based on the observation that in many processes the differential cross section for one additional parton emission given by parton shower is larger than that given by the exact matrix element:

$$\mathcal{R}_n\left(\Phi_{\mathcal{R}}\right) \leq \mathcal{B}_n\left(\Phi_{\mathcal{B}}\right) \otimes \mathcal{K}_n\left(\Phi_1\right) \tag{3.6}$$

where we continue using our notations from Section 1.5, see eqn.(1.119). $\mathcal{R}_n$ is the real correction for process with n outgoing particles, which in other words is associated with process with $n+1$ particles at Born level. $\Phi_{\mathcal{R}}$, $\Phi_{\mathcal{B}}$, Φ_1 are the phase space for the real correction, Born level contribution and additional emission, respectively. We assume the real emission phase space can be factorized as

$$\Phi_{\mathcal{R}} = \Phi_{\mathcal{B}} \otimes \Phi_1 \tag{3.7}$$

Before applying a matching method, at LO the radiation pattern up to the first emission from parton shower is given by:

$$d\sigma_n^{(\text{Born})} = d\Phi_{\mathcal{B}}\mathcal{B}_n\left(\Phi_{\mathcal{B}}\right)\left\{\Delta_n^{(\mathcal{K})}\left(t_c,\mu_Q^2\right)+\int_{t_c}^{\mu_Q^2} d\Phi_1\left[\mathcal{K}_N\left(\Phi_1\right)\Delta_n^{(\mathcal{K})}\left(t\left(\Phi_1\right),\mu_Q^2\right)\right]\right\} \tag{3.8}$$

Here the first term in the curly bracket corresponds to the no-emission probability down to the infrared cutoff t_c, whereas the second term combines the probability of no emission above hardness t (below the upper limit μ_Q^2 associated with the process and configuration) and the probability of hardest emission at t, and therefore the integrand gives the probability of the hardest emission happening at $t < \mu_Q^2$. These two terms add up to unity, as consistent with the probabilistic interpretation that either a parton is emitted or not emitted. This is the unitarity associated with parton shower, which we mentioned before.

Based on eqn. (3.6), the trick of the MEC method is to use a modified splitting kernel given by

$$\tilde{\mathcal{K}}_n\left(\Phi_1\right) = \mathcal{R}_n\left(\Phi_{\mathcal{B}}\times\Phi_1\right)/\mathcal{B}_n\left(\Phi_{\mathcal{B}}\right) \tag{3.9}$$

and eqn. (3.8) becomes

$$d\sigma_n^{(\text{Born})} = d\Phi_{\mathcal{B}}\mathcal{B}_n\left(\Phi_{\mathcal{B}}\right)\left\{\Delta_n^{(\mathcal{R}/\mathcal{B})}\left(t_c,\mu_Q^2\right)+\int_{t_c}^{\mu_Q^2} d\Phi_1\left[\frac{\mathcal{R}_n\left(\Phi_{\mathcal{B}}\times\Phi_1\right)}{\mathcal{B}_n\left(\Phi_{\mathcal{B}}\right)}\Delta_n^{(\mathcal{R}/\mathcal{B})}\left(t\left(\Phi_1\right),\mu_Q^2\right)\right]\right\} \tag{3.10}$$

The contents in the curly bracket still add to unity, which means the cross section is not modified, but now the hardest emission is determined by the full real correction matrix element. Further emissions are still determined using the original parton shower kernel. In practice this can be implemented with simple reweighting, by using the original parton shower to generate a first test emission and accepting it with probability

$$P_{\text{MEC}} = \frac{\tilde{\mathcal{K}}_n\left(\Phi_1\right)}{\mathcal{K}_n\left(\Phi_1\right)} = \frac{\mathcal{R}_n\left(\Phi_{\mathcal{B}}\times\Phi_1\right)}{\mathcal{B}_n\left(\Phi_{\mathcal{B}}\right)\times\mathcal{K}_n\left(\Phi_1\right)} \tag{3.11}$$

Though the MEC method is straightforward, it is limited by the assumption that eqn.(3.6) is true in all emission phase space, which is not always the case since parton shower can miss part of the phase space, especially in the case of production processes at hadron colliders, where typically the emission phase space is constrained by the upper scale determined from the hard process.

3.2.2 NLO Matching with the POWHEG Method

For NLO matching, in addition to ensuring that the hardest emission pattern follows the real correction, we also want to keep the total cross section from the fixed-order NLO calculation, and at the same time maintain the intrinsic logarithmic accuracy of the parton shower. One of the commonly used NLO+PS matching method is the POWHEG method [89], where instead of improving the MEC method by simply multiplying the parton shower simulation by a global K-factor, we modify the Born-level contribution with a local K-factor.

To explain the central idea of the POWHEG method, we mainly follow a heuristic approach from Ref. [136]. Starting with eqn. (1.123) for fixed-order NLO cross section calculation with the subtraction method:

$$\begin{aligned}\sigma_n^{NLO} = &\int d\Phi_{\mathcal{B}} \left[\mathcal{B}_n\left(\Phi_{\mathcal{B}};\mu_F,\mu_R\right) + \mathcal{V}_n\left(\Phi_{\mathcal{B}};\mu_F,\mu_R\right) + \mathcal{I}_n^{(\mathcal{S})}\left(\Phi_{\mathcal{B}};\mu_F,\mu_R\right)\right] \\ &+ \int d\Phi_{\mathcal{R}} \left[\mathcal{R}_n\left(\Phi_{\mathcal{R}};\mu_F,\mu_R\right) - \mathcal{S}_n\left(\Phi_{\mathcal{R}};\mu_F,\mu_R\right)\right]\end{aligned} \tag{3.12}$$

From now on we simplify the notation and the dependence on μ_F and μ_R is implicit. Generalizing this to evaluate the expectation value of an infrared-safe observable O we have

$$\begin{aligned}\langle O\rangle_{FO} = \int O d\sigma = &\int d\Phi_{\mathcal{B}} O\left(\Phi_{\mathcal{B}}\right)\left[\mathcal{B}_n\left(\Phi_{\mathcal{B}}\right) + \mathcal{V}_n\left(\Phi_{\mathcal{B}}\right) + \mathcal{I}_n^{(\mathcal{S})}\left(\Phi_{\mathcal{B}}\right)\right] \\ &+ \int d\Phi_{\mathcal{R}} \left[\mathcal{R}_n\left(\Phi_{\mathcal{R}}\right) - \mathcal{S}_n\left(\Phi_{\mathcal{R}}\right)\right] O\left(\Phi_{\mathcal{R}}\right)\end{aligned} \tag{3.13}$$

where $O = 1$ reduces back to the cross section evaluation, which is the case of our main focus in this discussion. For a comparison with parton shower expression to be made below, we can reorganize this equation by adding and subtracting the same term:

$$\begin{aligned}
\langle O\rangle_{FO} &= \int d\Phi_{\mathcal{B}} O\left(\Phi_{\mathcal{B}}\right)\left[\mathcal{B}_n\left(\Phi_{\mathcal{B}}\right)+\mathcal{V}_n\left(\Phi_{\mathcal{B}}\right)+\mathcal{I}_n^{(\mathcal{S})}\left(\Phi_{\mathcal{B}}\right)\right] \\
&\quad+\int d\Phi_{\mathcal{R}}\left[\mathcal{R}_n\left(\Phi_{\mathcal{R}}\right)-\mathcal{S}_n\left(\Phi_{\mathcal{R}}\right)\right] O\left(\Phi_{\mathcal{R}}\right) \\
&\quad+\left[\int d\Phi_{\mathcal{R}}\mathcal{R}_n\left(\Phi_{\mathcal{R}}\right) O\left(\Phi_{\mathcal{B}}\right)-\int d\Phi_{\mathcal{R}}\mathcal{R}_n\left(\Phi_{\mathcal{R}}\right) O\left(\Phi_{\mathcal{B}}\right)\right] \\
&=\int d\Phi_{\mathcal{B}} O\left(\Phi_{\mathcal{B}}\right)\left[\mathcal{B}_n\left(\Phi_{\mathcal{B}}\right)+\mathcal{V}_n\left(\Phi_{\mathcal{B}}\right)+\mathcal{I}_n^{(\mathcal{S})}\left(\Phi_{\mathcal{B}}\right)+\int d\Phi_1\left(\mathcal{R}_n\left(\Phi_{\mathcal{R}}\right)-\mathcal{S}_n\left(\Phi_{\mathcal{R}}\right)\right)\right] \\
&\quad+\int d\Phi_{\mathcal{R}}\mathcal{R}_n\left(\Phi_{\mathcal{R}}\right)\left[O\left(\Phi_{\mathcal{R}}\right)-O\left(\Phi_{\mathcal{B}}\right)\right] \\
&\equiv\int d\Phi_{\mathcal{B}} O\left(\Phi_{\mathcal{B}}\right)\bar{\mathcal{B}}_n\left(\Phi_{\mathcal{B}}\right)+\int d\Phi_{\mathcal{R}}\mathcal{R}_n\left(\Phi_{\mathcal{R}}\right)\left[O\left(\Phi_{\mathcal{R}}\right)-O\left(\Phi_{\mathcal{B}}\right)\right]
\end{aligned} \tag{3.14}$$

Here

$$\bar{\mathcal{B}}_n\left(\Phi_{\mathcal{B}}\right) \equiv \mathcal{B}_n\left(\Phi_{\mathcal{B}}\right)+\mathcal{V}_n\left(\Phi_{\mathcal{B}}\right)+\mathcal{I}_n^{(\mathcal{S})}\left(\Phi_{\mathcal{B}}\right)+\int d\Phi_1\left(\mathcal{R}_n\left(\Phi_{\mathcal{R}}\right)-\mathcal{S}_n\left(\Phi_{\mathcal{R}}\right)\right) \tag{3.15}$$

By comparing with eqn. (3.12) we see that $\bar{\mathcal{B}}_n\left(\Phi_{\mathcal{B}}\right)$ gives the fully accurate NLO cross section when integrated over $\Phi_{\mathcal{B}}$. We can therefore think of $\bar{\mathcal{B}}_n\left(\Phi_{\mathcal{B}}\right)$ as the differential cross section of Born-level configurations with a local NLO K-factor.

On the other hand, we can compute the corresponding observable given by including the first emission in the parton shower to the Born contribution, similar to eqn. (3.8):

$$\langle O\rangle_{\mathrm{PS}}=\int d\Phi_{\mathcal{B}}\mathcal{B}_n\left(\Phi_{\mathcal{B}}\right)\left\{O\left(\Phi_{\mathcal{B}}\right)\Delta\left(t_c,\mu_Q^2\right)+\int_{t_c}^{\mu_Q^2} d\Phi_1 O\left(\Phi_{\mathcal{R}}\right)\Delta\left(t(\Phi_1),\mu_Q^2\right)\frac{1}{t}\frac{\alpha_S}{2\pi}P(z)\right\} \tag{3.16}$$

Here we write the parton shower splitting kernel $\mathcal{K}_n$ explicitly as $P(z)$, and the Sudakov form factor $\Delta_n^{(\mathcal{K})}$ as Δ. The subscript n in Δ_n is implicit from now on until when we talk about multijet merging with different n in Section 3.2.4.

To truncate to an accuracy up to α_S, we can expand $\Delta(t, \mu_Q^2)$ and keep only up to the leading terms in α_S.

$$\begin{aligned}\langle O\rangle_{\mathrm{PS}} &= \int d\Phi_{\mathcal{B}} \mathcal{B}_n\left(\Phi_{\mathcal{B}}\right)\left\{O\left(\Phi_{\mathcal{B}}\right)+\int_{t_c}^{\mu_Q^2} d\Phi_1\left(O\left(\Phi_{\mathcal{R}}\right)-O\left(\Phi_{\mathcal{B}}\right)\right)\frac{1}{t}\frac{\alpha_S}{2\pi}P(z)\right\} \\ &= \int d\Phi_{\mathcal{B}} O\left(\Phi_{\mathcal{B}}\right)\mathcal{B}_n\left(\Phi_{\mathcal{B}}\right)+\int d\Phi_{\mathcal{B}}\int_{t_c}^{\mu_Q^2} d\Phi_1 \mathcal{B}_n\left(\Phi_{\mathcal{B}}\right)\frac{1}{t}\frac{\alpha_S}{2\pi}P(z)\left[O\left(\Phi_{\mathcal{R}}\right)-O\left(\Phi_{\mathcal{B}}\right)\right]\end{aligned} \tag{3.17}$$

$\langle O\rangle_{FO}$ corresponds to full NLO calculation but does not have the resummation of enhanced emissions, whereas $\langle O\rangle_{PS}$ is not accurate at NLO, but is derived from truncating resumed calculation. If we compare eqn. (3.14) with eqn. (3.17), we find the correspondence:

$$\bar{\mathcal{B}}_n\left(\Phi_{\mathcal{B}}\right) \leftrightarrow \mathcal{B}_n\left(\Phi_{\mathcal{B}}\right), \quad \mathcal{R}_n\left(\Phi_{\mathcal{R}}\right) \leftrightarrow \mathcal{B}_n\left(\Phi_{\mathcal{B}}\right)\frac{1}{t}\frac{\alpha_S}{2\pi}P(z) \tag{3.18}$$

This motivates us to replace the terms in parton shower calculation with the corresponding ones in fixed-order calculation, inside eqn. (3.16) where the resummed result is encoded:

$$\begin{aligned}\langle O\rangle_{\mathrm{NLO+PS}} = \int d\Phi_{\mathcal{B}}\bar{\mathcal{B}}_n\left(\Phi_{\mathcal{B}}\right)\Big\{&O\left(\Phi_{\mathcal{B}}\right)\Delta\left(t_c, \mu_Q^2\right) \\ &+\int_{t_c}^{\mu_Q^2} d\Phi_1 O\left(\Phi_{\mathcal{R}}\right)\Delta^{(\mathcal{R}/\mathcal{B})}\left(t(\Phi_1), \mu_Q^2\right)\frac{\mathcal{R}_n\left(\Phi_{\mathcal{R}}\right)}{\mathcal{B}_n\left(\Phi_{\mathcal{B}}\right)}\Big\}\end{aligned} \tag{3.19}$$

where similar to the MEC method, the Sudakov form factor has been modified to

$$\Delta^{(\mathcal{R}/\mathcal{B})}\left(t_m, t_n\right) = \exp\left[-\int_{t_m}^{t_n} d\Phi_1 \frac{\mathcal{R}_n\left(\Phi_{\mathcal{R}}\right)}{\mathcal{B}_n\left(\Phi_{\mathcal{B}}\right)}\right] \tag{3.20}$$

As already discussed, $\bar{\mathcal{B}}_n\left(\Phi_{\mathcal{B}}\right)$ leads to NLO accurate cross section. Modifying it by parton shower does not affect this cross section due to unitarity. Therefore we maintain the NLO accuracy of the inclusive cross section from fixed-order calculation. Meanwhile, by replacing the kernel in parton shower with $\frac{\mathcal{R}_n(\Phi_{\mathcal{R}})}{\mathcal{B}_n(\Phi_{\mathcal{B}})}$ which contains the full real-emission matrix element, we also ensure that the hardest parton emission is

accurate up to first order in α_S. This is the underlying idea in the POWHEG method, which stands for "POsitive-Weight Hardest Emission Generator", and as we can see from the integral, there is no negative weight involved when $\bar{\mathcal{B}}_n(\Phi_{\mathcal{B}})$ is nonnegative.

As with the MEC method, there is the issue of parton shower not covering the full emission phase space. Using processes in hadron colliders, such as Higgs boson production through gluon fusion as an example, we typically choose $\mu_Q \approx m_H$ as suggested by standard resummation technology. If we look at eqn. (3.19), we see that the hard emission is constrained to the phase space below m_H, whereas we want to maintain the accuracy from the real emission matrix element in the full emission phase space. If the transverse momentum k_T^2 of the emission is used as the hardness t, then we cannot describe well the Higgs boson transverse momentum in the large p_T tail. We can attempt to fix this issue by increasing μ_Q to as high as the center of mass energy of the hadron collision, so that the full emission phase space is reachable, but this raises a new issue. We have modified the Born-level contribution $\mathcal{B}_n$ to $\bar{\mathcal{B}}_n$ with a local K-factor, and this modification applies to all events. However, for events with hardest emission at a scale higher than the typical hardness associated with n-particle state, a K-factor associated with (n+1)-particle state should be applied, which may have non-negligible difference with the n-particle K-factor modifying $\mathcal{B}_n$. To handle this issue, the generalized POWHEG method is discussed in the next subsection in the context of the MC@NLO method.

The POWHEG BOX [87–90] is a general computer code framework for implementing NLO calculations in shower Monte Carlo programs according to the POWHEG method. In section 3.4 we also discuss certain samples generated with the POWHEG BOX (or POWHEG in shorthand) used in the ZZ analyses in this thesis.

3.2.3 NLO Matching with the MC@NLO Method

As we have discussed, the POWHEG method has the issue [138] of including hard real emission configuration in the exponential and extending the resummation beyond its region of validity. Therefore we are motivated to split the real emission matrix element $\mathcal{R}_n(\Phi_\mathcal{R})$ into an infrared-singular (soft) part $\mathcal{D}_n^A$ and an infrared-finite (hard) part $\mathcal{H}_n^A$:

$$\mathcal{R}_n = \mathcal{D}_n^A + \mathcal{H}_n^A \tag{3.21}$$

Then eqn. (3.12) becomes

$$\begin{aligned}\sigma_n^{NLO} = &\int d\Phi_\mathcal{B}\left[\mathcal{B}_n(\Phi_\mathcal{B}) + \mathcal{V}_n(\Phi_\mathcal{B}) + \mathcal{I}_n^{(\mathcal{S})}(\Phi_\mathcal{B})\right] \\ &+ \int d\Phi_\mathcal{R}\left[\mathcal{D}_n^A(\Phi_\mathcal{R}) - \mathcal{S}_n(\Phi_\mathcal{R})\right] + \int d\Phi_\mathcal{R}\mathcal{H}_n^A(\Phi_\mathcal{R})\end{aligned} \tag{3.22}$$

and eqn. (3.19) can be modified to

$$\begin{aligned}\langle O\rangle_{\text{NLO+PS}} = &\int d\Phi_\mathcal{B}\bar{\mathcal{B}}_n^A(\Phi_\mathcal{B})\left\{O(\Phi_\mathcal{B})\,\Delta^A\left(t_c,\mu_Q^2\right)\right. \\ &\left.+\int_{t_c}^{\mu_Q^2} d\Phi_1 O(\Phi_\mathcal{R})\,\Delta^A\left(t(\Phi_1),\mu_Q^2\right)\frac{\mathcal{D}_n^A(\Phi_\mathcal{R})}{\mathcal{B}_n(\Phi_\mathcal{B})}\right\} \\ &+\int d\Phi_\mathcal{R}\mathcal{H}_n^A(\Phi_\mathcal{R})\end{aligned} \tag{3.23}$$

with

$$\bar{\mathcal{B}}_n^A(\Phi_\mathcal{B}) = \mathcal{B}_n(\Phi_\mathcal{B}) + \mathcal{V}_n(\Phi_\mathcal{B}) + \mathcal{I}_n^{(\mathcal{S})}(\Phi_\mathcal{B}) + \int d\Phi_1\left(\mathcal{D}_n^A(\Phi_\mathcal{R}) - \mathcal{S}_n(\Phi_\mathcal{R})\right) \tag{3.24}$$

and

$$\Delta^A(t_m,t_n) = \exp\left[-\int_{t_m}^{t_n} d\Phi_1 \frac{\mathcal{D}_n^A(\Phi_\mathcal{R})}{\mathcal{B}_n(\Phi_\mathcal{B})}\right] \tag{3.25}$$

In this way, we have the flexibility of describing emission in the infrared-singular regime with parton shower which is modified by the local K-factor, and emission in the hard regime where the K-factor is not applied. This is the underlying idea in the

MC@NLO method originally presented in Ref. [141], where we have identified the soft real emission term with the parton shower splitting kernel:

$$\mathcal{D}_n^A = \sum_{ijk} \mathcal{B}_n\left(\Phi_{\mathcal{B}}\right) \otimes \mathcal{K}_{ij;k}\left(\Phi_1\right) = \mathcal{B}_n\left(\Phi_{\mathcal{B}}\right) \otimes \mathcal{K}_n\left(\Phi_1\right) \tag{3.26}$$

The hard emission term at the third line of eqn. (3.23) gives the correct NLO cross section when combined with the soft emission term at the first two lines, and at the same time it fills the emission phase space region inaccessible by the parton shower with the correct pattern from real emission matrix element. The two terms in eqn. (3.23) are not necessarily both positive, therefore there can be negative weights and unweighting can only be done up to a sign.

MadGraph5_aMC@NLO [86] is a fully automated and public computer code that can be used to perform NLO+PS calculation with the MC@NLO formalism and includes capacities related to multijet merging. Samples generated with MadGraph5_aMC@NLO used in the ZZ analyses are discussed in Section 3.4.

3.2.4 Multijet Merging

Since matrix element provides a good description of additional hard jets at large angles, we can improve the predictive power of simulation in a different direction by combining matrix elements with different jet multiplicities, and leave the softer or more collinear emission to parton shower description. This is called the "ME+PS" approach. We first talk about merging LO matrix elements, which was first formulated in Ref. [142], known as the CKKW method, and has several variants including the CKKW-L, MLM, pseudoshower and MESS methods, as discussed in Refs. [140, 143, 144] and the references therein.

An important concern again arises from double counting between fixed-order calculation and parton shower. The central idea of merging LO matrix elements with

the CKKW-type method is to decompose the emission phase space into a hard and a soft regime that complements each other, in general by introducing a cutoff scale $Q_{\rm cut}$ that is typically related to the transverse momentum of emission. Instead of modifying the parton shower side to incorporate corrections from matrix elements as in MEC and NLO+PS methods, the matrix elements are reweighted with suitable Sudakov form factors, which transform the matrix elements from describing inclusive production of N-jet system plus all other emissions, to describing exclusive production with N jets only.

The differential cross section of merging LO matrix elements associated with $N, N+1, \cdots, N_{\rm max}$ jets is given by:

$$\begin{aligned} d\sigma = & \sum_{n=N}^{N_{\rm max}-1} d\Phi_n \mathcal{B}_n \Theta\left(Q_n - Q_{\rm cut}\right) \mathcal{E}_n^{(\mathcal{K},<Q_{\rm cut})}\left(\mu_n^2, t_c\right) \\ & + d\Phi_{N_{\rm max}} \mathcal{B}_{N_{\rm max}} \Theta\left(Q_{N_{\rm max}} - Q_{\rm cut}\right) \mathcal{E}_{N_{\rm max}}^{(\mathcal{K},<Q_{N_{\rm max}})}\left(\mu_{N_{\rm max}}^2, t_c\right) \end{aligned} \tag{3.27}$$

We start by generating matrix elements associated with $\leq N_{max}$ additional partons, then require the relative transverse momentum between any of these hard partons to be above $Q_{\rm cut}$, as in the step functions appearing in the first and second line, where Q_n is the lowest hardness among the n emissions from matrix element. Then parton shower is performed to generate the remaining parton emission below $Q_{\rm cut}$ for the matrix elements up to $N_{max} - 1$ additional emissions, whereas for the last matrix element with N_{max} additional emissions, the parton shower is not constrained by $Q_{\rm cut}$, but instead only required to be below the lowest hardness from matrix element emissions. $\mathcal{E}_n^{(\mathcal{K},<Q_{\rm cut})}\left(\mu_n^2, t_c\right)$ is the parton shower emission operator that generates all subsequent parton shower emissions off an n-particle configuration, and is recursively

defined as:

$$\mathcal{E}_n^{(\mathcal{K},<Q)}\left(\mu_Q^2, t_{\mathrm{c}}\right) = \Delta_n^{(\mathcal{K})}\left(\mu_Q^2, t_c\right) + \int^{\mu_Q^2} d\Phi_1 \left[\mathcal{K}_n^{<Q}\left(\Phi_1\right) \Delta_n^{(\mathcal{K})}\left(\mu_Q^2, t\left(\Phi_1\right)\right) \otimes \mathcal{E}_{n+1}^{(\mathcal{K},<Q)}\left(t\left(\Phi_1\right), t_{\mathrm{c}}\right)\right] \tag{3.28}$$

with

$$\mathcal{K}_N^{<Q}\left(\Phi_1\right) = \mathcal{K}_N\left(\Phi_1\right)\Theta\left(Q - Q_1\right) \tag{3.29}$$

extending the ideas presented in eqn. (3.8). Q_{cut} need to be chosen properly so that it is large enough for perturbation theory to hold for the fixed-order calculation, but small enough so that the parton shower description below this scale is accurate. Suitable scale choices of α_S also need to be used to evaluate the matrix elements. Note that in the case of $N = 0$ and $N_{max} = 1$, ME+PS and MEC have equivalent accuracy [18], but the latter does not need the parameter Q_{cut}. Alternative to the CKKW-type approaches, one can apply the MLM method [143, 145], which vetos unwanted emissions only after the parton shower evolution has finished.

When we want to describe more than one hard jet at large angle associated with the main process, ME+PS has advantage over NLO+PS. Naturally we also want to improve the NLO generators so that the associated additional jet is also described with NLO accuracy, and this is achieved with the FxFx [146], UNLOPS [147], MiNLO [148] and MEPS@NLO [149] methods. In particular the FxFx method is an NLO multijet merging method built upon the matching achieved from the MC@NLO method, and has been applied in our nominal $\mathrm{q\bar{q}} \to \mathrm{ZZ}$ sample as discussed in Section 3.4 which merges two processes with 0- and 1-jet emission both at NLO. On the other hand, the MiNLO method is a procedure that achieves NLO accuracy of multijet calculation by suitable choice of the scale of α_S and inclusion of the Sudakov form factor, and is particularly suited [150] to be implemented with the parton shower generators using the POWHEG method. The MiNLO' method [151] modifies

the MiNLO method in such a way that, for the production of a massive color singlet system plus an extra jet, NLO accuracy is also formally retained for inclusive observables, and this method does not require a merging scale. With the MiNLO' method, NNLO+PS accuracy can be achieved through reweighting, and it can be extended to the MiNNLO$_{\rm PS}$ discussed in the next subsection.

3.2.5 NNLO+PS Matching

The current experimental precision at LHC has reached a level that requires NNLO accuracy along with parton shower to describe the experimental data. Therefore NNLO+PS generators are needed, which not only have NLO accuracy for the production of an associated jet, but also have NNLO accuracy for fully inclusive observables. In the following discussion we focus on the MiNNLO$_{\rm PS}$ method which is applied in one of the ZZ MC samples used in the ZZ+jets analysis. Other NNLO+PS methods include the UN^{2}LOPS method [152] and the GENEVA method [153]. The MiNNLO$_{\rm PS}$ method [5, 150, 154–158] extends the MiNLO' method to NNLO accuracy at the fully differential level for inclusive jet multiplicity and has NLO accuracy for 1-jet configuration. With this method no merging scale is required to separate different jet multiplicities, and the LL structure of the parton shower simulation is preserved. Different from the MiNLO' method, it calculates NNLO corrections during event generation and no multidimensional reweighting is needed.

The MiNNLO$_{\rm PS}$ approach generates the first two hardest emissions and leave the remaining ones to parton shower. It works on a system of produced color singlet particles F having invariant mass Q, and can also be extended and applied to the production of a colored final state (top-quark pair production) [157]. We use the notation FJ to denote the production of F plus one light parton, and FJJ for plus two light partons. The fully differential cross section can be expressed using the

formula associated with the POWHEG method (see Refs. [5, 154] for example):

$$\frac{d\sigma}{d\Phi_{FJ}} = \bar{B}^{\mathrm{MiNNLO_{PS}}}\left(\Phi_{FJ}\right) \times \left\{ \Delta_{pwg}\left(\Lambda_{pwg}\right) + \int d\Phi_{rad} \Delta_{pwg}\left(p_{T,rad}\right) \frac{R\left(\Phi_{FJ}, \Phi_{rad}\right)}{B\left(\Phi_{FJ}\right)} \right\} \tag{3.30}$$

Here Δ_{pwg} is the POWHEG Sudakov form factor, with Λ_{pwg} the infrared cutoff having a default value of 0.89 GeV. Φ_{rad} and $p_{T,rad}$ are the phase space and transverse momentum of the second emission, respectively. B and R are the squared Born and real emission matrix elements for the FJ production, respectively, and in other words R is the Born-level squared matrix element for the FJJ production. The first emission (FJ production) is generated at NLO using $\bar{B}^{\mathrm{MiNNLO_{PS}}}$, which is the crucial element of the MiNNLO$_{\mathrm{PS}}$ method, and is modified to obtain NNLO accuracy for inclusive F production. The second emission is generated according to the POWHEG method using the content inside the curly bracket to preserve the NLO accuracy of the FJ cross section. Then additional emissions are included through parton shower. In the MiNNLO$_{\mathrm{PS}}$ method, the renormalization and factorization scales are set to the transverse momentum p_{T} of F:

$$\mu_R \sim \mu_F \sim p_{\mathrm{T}} \tag{3.31}$$

The expression of $\bar{B}^{\mathrm{MiNNLO_{PS}}}$ is given by:

$$\begin{aligned}
\bar{B}\left(\Phi_{FJ}\right)^{\mathrm{MiNNLO_{PS}}} = & \exp\left[-\tilde{S}\left(p_{\mathrm{T}}\right)\right] \left\{ \frac{\alpha_S\left(p_{\mathrm{T}}\right)}{2\pi} \left[\frac{d\sigma_{FJ}}{d\Phi_{FJ}}\right]^{(1)} \left(1 + \frac{\alpha_S\left(p_{\mathrm{T}}\right)}{2\pi} \left[\tilde{S}\left(p_{\mathrm{T}}\right)\right]^{(1)}\right) \right. \\
& + \left(\frac{\alpha_S\left(p_{\mathrm{T}}\right)}{2\pi}\right)^2 \left[\frac{d\sigma_{FJ}}{d\Phi_{FJ}}\right]^{(2)} \\
& + \left(D\left(p_{\mathrm{T}}\right) - \frac{\alpha_S\left(p_{\mathrm{T}}\right)}{2\pi} \left[D\left(p_{\mathrm{T}}\right)\right]^{(1)} - \left(\frac{\alpha_S\left(p_{\mathrm{T}}\right)}{2\pi}\right)^2 \left[D\left(p_{\mathrm{T}}\right)\right]^{(2)} \right) \\
& \left. \times F^{\mathrm{corr}}\left(\Phi_{FJ}\right) \right\}
\end{aligned} \tag{3.32}$$

where the $[X]^{(i)}$ notation stands for the coefficient of th i-th term in the perturbative expansion of quantity X. $\exp\left[-\tilde{S}\left(p_{\mathrm{T}}\right)\right]$ is the Sudakov form factor with cutoff at p_{T}.

The last term in the $\bar{B}\left(\Phi_{FJ}\right)^{\mathrm{MiNNLO_{PS}}}$ expression is of order $\alpha_S\left(p_{\mathrm{T}}\right)^3$, and it adds the relevant contributions so that integration over p_{T} gives NNLO accurate results. D is a function as derived in [150], and F^{corr} spreads the NNLO correction to the full FJ phase space.

3.3 Pileup and Detector Simulation

As discussed in Ref. [32], in typical CMS simulated samples, after simulation associated with the matrix element and parton shower, subsequent hadronization is also simulated, and underlying events with tuned parameters are added so that the simulation matches observed features of data particularly sensitive to the contribution of the underlying events. This is discussed for example in Ref. [68]. In addition, the simulated samples include additional interactions per bunch crossing, referred to as pileup. Simulated events were weighted so that the pileup distribution reproduces that observed in the data. To produce simulated events comparable to observed data, the particles are tracked through the detector, modeling their interactions with the detector elements, followed by simulation of the generation of electrical signals and their digitization to form a recorded event. The detector response is simulated using a detailed description of the CMS detector implemented with the GEANT4 package [159].

3.4 Samples Used in this Analysis

The MC simulation used for this analysis can be divided into signal and background samples. The ZZ signal production via quark-antiquark annihilation is simulated at next-to-leading order (NLO) in QCD with MADGRAPH5_aMC@NLO v2.4.2 [86] and

POWHEG 2.0 [87–90]. The MADGRAPH5_aMC@NLO sample is used as the nominal $q\bar{q} \to ZZ$ sample in reconstruction-level distributions and for unfolding, because this sample is expected to describe data better than POWHEG since it merges the 0-jet and 1-jet NLO processes, whereas the POWHEG sample is simulated at NLO accuracy for 0-jet and LO accuracy for 1-jet processes. The $gg \to ZZ$ process is simulated at leading order (LO) with MCFM v7.0 [160]. The cross sections of these samples are normalized to the cross sections calculated at NNLO in QCD for $q\bar{q} \to ZZ$ (K factor of 1.1) [70] and at NLO in QCD for $gg \to ZZ$ (K factor of 1.7) [161]. The production processes via SM Higgs boson production and decay (specifically $gg \to H \to ZZ$) are simulated with POWHEG at NLO. Electroweak ZZ production in association with two jets is simulated with MADGRAPH5_aMC@NLO [86] at LO. The nominal SM signal predictions are derived from the MADGRAPH5_aMC@NLO $q\bar{q} \to ZZ$ sample, the MCFM $gg \to ZZ$ sample, and the MADGRAPH5_aMC@NLO EW production sample, which includes vector boson fusion Higgs boson contributions and their interference with non-Higgs boson EW production, and the POWHEG $H \to ZZ$ sample.

Simulated events for the irreducible background processes containing four prompt leptons in the final state, such as $t\bar{t}Z$, WWZ, WZZ, and ZZZ, where the last three are combined and denoted as VVV, are simulated with MADGRAPH5_aMC@NLO at LO ($t\bar{t}Z$) and NLO (VVV). In the plots in this thesis we use VVV to represent all these four processes.

Parton showering, hadronization and fragmentation are simulated in all samples with PYTHIA8 8.226 and 8.230 [63], with parameters set by the CUETP8M1 [67] (CP5 [68]) tune for the 2016 (2017 and 2018) data-taking period as mentioned in Section 1.5.6, and the NNPDF3.0 (3.1) set of parton distribution functions, PDFs, [162] is used.

Results are also compared with the very recent nNNLO+PS predictions [5, 91]

(we use samples from the authors of [5]), which consist of NNLO predictions for the quark-initiated production combined with parton showers using the $\text{MiNNLO}_{\text{PS}}$ method, and NLO predictions for the loop-induced gluon fusion production matched to parton showers, with event generators for the two channels implemented in the POWHEG framework. Spin correlations, interferences, and off-shell effects are included by calculating the full process $pp \to \ell^+\ell^-\ell'^+\ell'^-$ and considering all contributions to the four-lepton final state. The contribution mediated by a Higgs boson is included in the gluon fusion production mode.

As part of the nNNLO+PS predictions, the $q\bar{q} \to ZZ$ predictions from the $\text{MiNNLO}_{\text{PS}}$ method are accurate at NNLO for inclusive production and accurate at NLO for Z+1-jet production. The combination of the two jet multiplicities does not require any unphysical merging scale [150]. These predictions are expected to be more accurate at high jet multiplicities than: (i) the POWHEG $q\bar{q} \to ZZ$ predictions, which are accurate at NLO in inclusive production; (ii) the MADGRAPH5_aMC@NLO predictions, which are simulated at NLO with the 0- and 1-jet processes, and merged using the FxFx scheme [146].

Chapter 4

Reconstruction and Selection

In this chapter, we discuss the reconstruction, identification and selection of events and different physics objects within the CMS experiment related to and associated with the ZZ analyses in this thesis. Many of the methods, concepts, configurations and estimates discussed in this chapter and also presented in Ref. [1], follow from those used in the analysis presented in Ref. [2].

4.1 Global Event Reconstruction

The global event reconstruction (also called particle-flow event reconstruction [163]) aims to reconstruct and identify each individual particle in an event, with an optimized combination of all subdetector information. In this process, the identification (ID) of the particle type (photon, electron, muon, charged hadron, neutral hadron) plays an important role in the determination of the particle direction and energy. Photons (e.g. coming from π^0 decays or from electron bremsstrahlung) are identified as ECAL energy clusters not linked to the extrapolation of any charged particle trajectory to the ECAL. Electrons (e.g. coming from photon conversions in the tracker

material or from B hadron semileptonic decays) are identified as a primary charged particle track and potentially many ECAL energy clusters corresponding to this track extrapolation to the ECAL and to possible bremsstrahlung photons emitted along the way through the tracker material. Muons (e.g. from B hadron semileptonic decays) are identified as tracks in the central tracker consistent with either a track or several hits in the muon system, and associated with calorimeter deposits compatible with the muon hypothesis. Charged hadrons are identified as charged particle tracks neither identified as electrons, nor as muons. Finally, neutral hadrons are identified as HCAL energy clusters not linked to any charged hadron trajectory, or as a combined ECAL and HCAL energy excess with respect to the expected charged hadron energy deposit.

The energy of photons is obtained from the ECAL measurement. The energy of electrons is determined from a combination of the track momentum at the main interaction vertex, the corresponding ECAL cluster energy, and the energy sum of all bremsstrahlung photons attached to the track. The energy of muons is obtained from the corresponding track momentum. The energy of charged hadrons is determined from a combination of the track momentum and the corresponding ECAL and HCAL energies, corrected for the response function of the calorimeters to hadronic showers. Finally, the energy of neutral hadrons is obtained from the corresponding corrected ECAL and HCAL energies.

For each event, hadronic jets are clustered from these reconstructed particles using the infrared and collinear safe anti-k_{T} algorithm [164, 165] with a distance parameter of 0.4. Jet momentum is determined as the vectorial sum of all particle momenta in the jet, and is found from simulation to be, on average, within 5 to 10% of the true momentum over the whole p_{T} spectrum and detector acceptance. Additional proton-proton interactions within the same or nearby bunch crossings (pileup) can

contribute extra tracks and calorimetric energy depositions to the jet momentum. To mitigate this effect, charged particles identified to be originating from pileup vertices are discarded and an offset correction is applied to correct for remaining contributions. Jet energy corrections are derived from simulation to bring the measured response of jets to that of particle level jets on average. In situ measurements of the momentum balance in dijet, photon + jet, Z + jet, and multijet events are used to account for any residual differences in the jet energy scale between data and simulation [115]. The jet energy resolution amounts typically to 15–20% at 30 GeV, 10% at 100 GeV, and 5% at 1 TeV [115]. Additional selection criteria are applied to each jet to remove jets potentially dominated by anomalous contributions from various subdetector components or reconstruction failures.

4.2 Object Identification and Selection

Electrons and muons are considered candidates for the reconstruction of ZZ final states ("signal leptons") if their $p_{\mathrm{T}}^{\ell} > 7\,(5)$ GeV and their $\left|\eta^{\ell}\right| < 2.5\,(2.4)$ for electrons (muons). Signal leptons are required to originate from the primary interaction vertex of the event, which is taken to be the vertex corresponding to the hardest scattering in the event, evaluated using tracking information alone, as described in Section 9.4.1 of Ref. [166]. The distance of the lepton track origin from the primary vertex (PV) is required to be <1 cm along the beam line, and <0.5 cm in the transverse plane. Furthermore, the significance of the three-dimensional impact parameter relative to the event vertex, $\mathrm{SIP}_{3\mathrm{D}}$, is required to satisfy $\mathrm{SIP}_{3\mathrm{D}} \equiv \left|\frac{\mathrm{IP}}{\sigma_{\mathrm{IP}}}\right| < 4$ for each lepton, where IP is the distance of closest approach of the lepton track to the primary vertex and σ_{IP} is its associated uncertainty.

Loose and tight ID requirements are defined for each lepton. The tight IDs are

used for signal leptons, whereas the loose IDs are used in control regions to define objects that might be spuriously identified as a signal lepton. An electron satisfies the loose ID if it satisfies the p_{T}, η, and vertex requirements above. It satisfies the tight ID if it satisfies the loose ID and the multivariate discriminator described in Ref. [167]. A muon satisfies the loose ID if it satisfies the above p_{T}, η, and vertex requirements, and provides a good track-matching between the tracker and the muon detectors. It satisfies the tight ID if it satisfies the loose ID, and either is tagged as a muon by the PF algorithm for the years 2016–2017 (satisfies a multivariate discriminator for the year 2018 [168]), or is a high p_{T} ($> 200\,\mathrm{GeV}$) muon and satisfies a set of requirements on the quality of the associated track.

Signal leptons are required to be isolated from other particles in the event. The relative isolation is defined as

$$R_{\mathrm{Iso}} = \left(\sum_{\mathrm{charged}} p_{\mathrm{T}} + \max\Big[0, \sum_{\mathrm{neutral}} p_{\mathrm{T}} + \sum_{\mathrm{photons}} p_{\mathrm{T}} - p_{\mathrm{T}}^{\mathrm{PU}}\left(\ell\right)\Big] \right) \Big/ p_{\mathrm{T}}^{\ell} \tag{4.1}$$

where the sums run over the p_{T} of hadrons and photons in a cone of size $\Delta R \equiv \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} = 0.3$ around the lepton momentum direction, where ϕ is the azimuthal angle in radians. To mitigate the contribution of pileup interactions to the isolation, charged hadrons are included only if they originate from the primary vertex [169]. The estimated neutral contribution to the isolation from pileup, $p_{\mathrm{T}}^{\mathrm{PU}}(\ell)$, is defined differently for electrons and muons. For electrons, $p_{\mathrm{T}}^{\mathrm{PU}}(\mathrm{e}) \equiv \rho\, A_{\mathrm{eff}}$ where the average p_{T} flow density ρ is calculated for each event using a "jet area" method [170], and is defined as the median of the $p_{\mathrm{T}}^{\mathrm{jet}}/A_{\mathrm{jet}}$ distribution for all pileup jets in the event. The effective area A_{eff} is the geometric area of the isolation cone projection on the face of the calorimeter multiplied by an η-dependent correction factor that accounts for the residual dependence of the isolation value on pileup. For muons, $p_{\mathrm{T}}^{\mathrm{PU}}(\mu) \equiv 0.5 \sum_i p_{\mathrm{T}}^{\mathrm{PU},i}$, where i runs over the momenta of the charged hadron PF can-

didates originating from pileup vertices, and the factor of 0.5 corrects for the expected ratio of charged to neutral particles in the isolation cone. For the years 2016–2017, muons are considered isolated if their $R_{\text{Iso}} < 0.35$, whereas for 2018 and for electrons the isolation requirement is included in the multivariate discriminator used for the selection.

The efficiencies for the reconstruction, identification, and isolation of signal leptons are measured in data and simulation using a tag-and-probe technique [171] based on inclusive samples of Z boson events, with an additional sample of J/ψ events for low-p_{T} muons. The measurements are performed in bins of p_{T}^{ℓ} and $\left|\eta^{\ell}\right|$, where for electrons the supercluster η is used. The electron selection efficiency in the ECAL barrel (endcaps) varies from $\sim$85 (77)% at $p_{\text{T}}^{\text{e}} \approx 10\,\text{GeV}$, to $\sim$95 (89)% for $p_{\text{T}}^{\text{e}} \geq 20\,\text{GeV}$, and is $\sim$85% in the barrel-endcap transition region. The muons are reconstructed and identified with efficiencies above $\sim$98% within $|\eta^{\mu}| < 2.4$. The ratio between the data and simulation efficiencies in each p_{T}-$|\eta|$ bin is applied as a correction factor to leptons in simulated events. If the correction factor for a given lepton is $f_{\text{eff}}^{\ell}({p_{\text{T}}}^{\ell}, \eta^{\ell})$, the efficiency correction for each event is $\prod_{\ell} f_{\text{eff}}^{\ell}({p_{\text{T}}}^{\ell}, \eta^{\ell})$, where the product index runs over the four leptons of the ZZ candidate.

As discussed in Section 4.1, jets are reconstructed based on PF candidates, rejecting the charged hadrons associated to a pileup vertex, with the anti-k_{T} clustering algorithm [164, 165] using a distance parameter $R = 0.4$. To reduce the instrumental background, tight identification criteria based on the multiplicities and energy fractions carried by charged and neutral hadrons are imposed on jets [172]. Jets from pileup are rejected using pileup jet identification criteria based on the compatibility of the associated tracks with the primary vertex, when inside the tracker acceptance, and on the jet topology [169]. Jet energy corrections are applied to the reconstructed jets [115, 173].

4.3 ZZ Candidate and Event Selection

The data samples used in the ZZ analyses in this thesis are selected by the trigger system that requires the presence of a pair of loosely isolated leptons or a triplet of leptons, with minimum-p_{T} thresholds for leptons depending on the lepton combination. Further triggers include a set of single-electron and single-muon triggers, and triggers on leptons of different flavors. The trigger efficiency within the acceptance is greater than 98%.

Events are required to have at least four leptons. Each event should contain at least one lepton with $p_{\mathrm{T}} > 20\,\mathrm{GeV}$, two leptons with $p_{\mathrm{T}} > 10\,\mathrm{GeV}$, and four leptons with $p_{\mathrm{T}} > 7\,(5)\,\mathrm{GeV}$ for electrons (muons). All leptons must pass the "tight" lepton identification and isolation requirements described in Section 4.2.

The leptons are required to be separated by $\Delta R(\ell_1, \ell_2) > 0.02$, and electrons are required to be separated from muons by $\Delta R(\mathrm{e}, \mu) > 0.05$, to remove spurious "ghost" leptons arising from ambiguities in track reconstruction. Lepton pairs originating from hadronic decays are removed by requiring that all oppositely charged lepton pairs in the ZZ candidate have $m_{\ell_1\ell_2} > 4\,\mathrm{GeV}$ regardless of lepton flavor.

Z candidates are built from two oppositely-charged leptons of the same flavor. The pair is retained if it satisfies $4 < m_{\ell^+\ell^-} < 120\,\mathrm{GeV}$. All possible four-lepton candidates in an event are then considered. For each ZZ candidate, the dilepton pair with invariant mass closest to the nominal Z mass m_{Z} (91.1876 GeV [174]) is designated Z_1, and the other is designated Z_2. The event is kept if $40 < m_{\mathrm{Z}_1} < 120\,\mathrm{GeV}$ and $4 < m_{\mathrm{Z}_2} < 120\,\mathrm{GeV}$.

In the case of multiple ZZ candidates satisfying all requirements, the ambiguity is resolved by selecting the candidate where m_{Z_1} is closest to the nominal Z mass. If more than one lepton combination is still possible, the Z_2 candidate is chosen as the

one that maximizes the scalar p_{T} sum of the leptons.

In the ZZ analyses, the jets are required to have a $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ and $|\eta| < 4.7$. In addition, jets are required to be well separated from any isolated lepton by requiring $\Delta R(\text{jet,lepton}) > 0.4$, where the lepton here satisfies all the tight requirements except for the lower p_{T} requirement ($> 5\,\mathrm{GeV}$ instead of $> 7\,\mathrm{GeV}$ for electrons) and a relaxed $\mathrm{SIP}_{\mathrm{3D}}$ requirement (<10 instead of <4 for electrons).

For EW ZZjj signal extraction, we look at the phase space region with additional requirements: $m_{jj} > 100\,\mathrm{GeV}$ for VBS enrichment and suppression of triboson production associated with a W boson, and $m_{4\ell} > 180\,\mathrm{GeV}$ to comply with MC samples that do not describe well offshell ZZ *.

4.4 Gradual Shift in the Timing of ECAL L1 Triggers

During the 2016 and 2017 data-taking periods, a gradual shift in the timing of the inputs of the ECAL L1 trigger in the $|\eta| > 2.0$ region caused a trigger inefficiency, denoted as "prefiring" [101]. The trigger missed between 10% and 20% of the events containing an electron (of p_{T} greater than or close to $50\,\mathrm{GeV}$) or a jet (of p_{T} greater than or close to $100\,\mathrm{GeV}$) in the $2.5 < |\eta| < 3.0$ range, the exact loss depending on the p_{T} and η of the trigger object. Correction factors were computed from data and applied to the acceptance evaluated by simulation. In the ZZ analyses in this thesis the correction is applied to the event weights of the simulated events.

Chapter 5

Analysis Strategies

In this chapter we discuss the analysis strategies used for the ZZ+jets and EW ZZ+2jets analyses. We have used some of the text directly from Ref. [1].

5.1 Background Estimate

Continuing our discussion in Section 1.6.6, the ZZ $\rightarrow 4\ell$ production processes have very small background (other than the QCD-induced ZZ+2jets production in the EW ZZ+2jets analysis) due to the requirement of having four well-reconstructed and isolated leptons. The background can be categorized into reducible and nonreducible backgrounds. A common irreducible background comes from processes with four genuine leptons, such as $t\bar{t}Z$, WWZ, WZZ, and ZZZ (collectively referred to as VVV in the plots). These irreducible backgrounds are estimated with MC simulation using the samples described in Section 3.4.

Reducible background, denoted as "Z+X", comes from processes where jets or nonprompt leptons are misidentified as prompt leptons. The major processes are WZ+jets, Z+jets and $t\bar{t}$. The reducible background is estimated using a data-driven

method detailed in Ref. [175], and the reducible background estimation in this analysis is based on Ref. [2]. The probability of misidentification is estimated with a sample of Z+$\ell_{\text{candidate}}$ events from the experimental data, where the Z here denotes a pair of oppositely charged, same-flavor leptons that pass the selection requirements (defined in Section 4.2 and 4.3) and satisfy $|m_{\ell^+\ell^-} - m_{\text{Z}}| < 7\,\text{GeV}$. Each event of this sample must have exactly one additional lepton candidate $\ell_{\text{candidate}}$ that satisfies the loose ID requirement without isolation requirement applied. The misidentification rate is defined as the fraction of $\ell_{\text{candidate}}$ that satisfies the full lepton ID and isolation requirements. It is measured for each lepton flavor in bins of p_{T} and η. Events from WZ and events from ZZ production with one lepton not reconstructed have a third genuine lepton that can pass the tight ID and isolation requirements, and they are regarded as contamination to the Z+$\ell_{\text{candidate}}$ events. The WZ contamination is suppressed by requiring the missing transverse momentum $p_{\text{T}}^{miss} < 25\,\text{GeV}$. The p_{T}^{miss} is defined as the magnitude of the missing transverse momentum vector $\vec{p}_{\text{T}}^{\,\text{miss}}$, the projection onto the plane transverse to the beams of the negative vector momentum sum of all reconstructed PF candidates in the event, corrected for the jet energy scale (JES). The transverse mass, calculated as $m_{\text{T}} \equiv \sqrt{(p_{\text{T}}^{\ell} + p_{\text{T}}^{miss})^2 - (\vec{p}_{\text{T}}^{\,\ell} + \vec{p}_{\text{T}}^{\,\text{miss}})^2}$, is required to be <30 GeV. The residual contamination is estimated with MC simulation and subtracted from the Z+$\ell_{\text{candidate}}$ events.

After estimation of the misidentification rate, it is applied to two control samples from the experimental data to estimate the number of reducible background events in the signal region. Both samples consist of Z+X events containing a pair of leptons that pass the full selections, and two additional leptons that pass only the loose ID requirement. In one sample enriched in WZ+jets events, (called "3P1F" meaning "3 prompt 1 fake"), one of the additional leptons is required to pass the lepton tight ID and isolation requirements, while the other lepton is required to fail at least one

of these requirements. In the other control sample enriched in Z+jets and $t\bar{t}$ events ("2P2F"), both additional leptons must fail the full lepton requirements. The ratio between loose leptons that pass and fail the full requirement,

$$F_\ell\left(p_{\mathrm{T}}^{\ell}, \eta^{\ell}\right) = \frac{f_\ell\left(p_{\mathrm{T}}^{\ell}, \eta^{\ell}\right)}{1 - f_\ell\left(p_{\mathrm{T}}^{\ell}, \eta^{\ell}\right)} \tag{5.1}$$

is applied per lepton to extrapolate from the control region to the signal region. The number of background events in the signal region is estimated to be

$$N_{\mathrm{bkg}} = \sum_{\ell \in \mathrm{3P1F}} F_\ell\left(p_{\mathrm{T}}^{\ell}, \eta^{\ell}\right) - \sum_{(\ell_1,\ell_2)\in 2P2F} F_{\ell_1}\left(p_{\mathrm{T}}^{\ell_1}, \eta^{\ell_1}\right) F_{\ell_2}\left(p_{\mathrm{T}}^{\ell_2}, \eta^{\ell_2}\right). \tag{5.2}$$

The first term is the number of events in the signal region estimated from the 3P1F sample. The 3P1F events also include 2P2F events where one lepton is misidentified as a prompt lepton (with fraction given by F_ℓ), therefore the second term in the equation is included to remove double counting. The control samples can also contain contamination from ZZ signal events due to selection inefficiency, and the contamination is estimated with MC and subtracted.

5.2 ZZ+jets Differential Cross Section Measurement

To obtain ZZ+jets differential cross sections normalized to the ZZ fiducial cross sections (for the on-shell ZZ region and for the full four-lepton invariant mass range as defined by the kinematic requirements) and compare experimental data to theoretical predictions, the data are "unfolded" to remove detector resolution, efficiency, and acceptance effects, and compared with the theoretical predictions from MadGraph-5_aMC@NLO $q\bar{q} \to ZZ$ and POWHEG $q\bar{q} \to ZZ$, where MCFM $gg \to ZZ$, POWHEG $H \to ZZ$, and MadGraph5_aMC@NLO EW ZZ predictions are also added to these

two sets of predictions. The unfolded results are also compared with the nNNLO+PS predictions.

The measured on-shell ZZ fiducial cross section from Ref. [2] is $40.5 \pm 0.7\,(\text{stat}) \pm 1.1\,(\text{syst}) \pm 0.7\,(\text{lumi})$ fb, which agrees well with the expected value of $39.3^{+0.8}_{-0.7} \pm 0.6$ fb. As explained in Chapter 1, this fiducial cross section is valid for the current analysis. The fiducial phase space selections are similar to the reconstruction-level selections and detailed in Table 5.1. We use the notation $m_{\text{Z}_1,\text{Z}_2}$ to refer to both m_{Z_1} and m_{Z_2}. The MC truth-level distributions use generator-level leptons "dressed" by adding the momenta of generator-level photons within $\Delta R(\ell, \gamma) < 0.1$ from the direction of the lepton. For EW ZZ+2jets analysis, we additionally require $m_{jj} > 100\,\text{GeV}$ and $m_{4\ell} > 180\,\text{GeV}$ for the fiducial region.

Table 5.1: Particle-level selections used to define the fiducial phase space

Particle type	Selection
	ZZ base selection
Leptons	$p_{\text{T}}(\ell_1) > 20\,\text{GeV}$ $p_{\text{T}}(\ell_2) > 10\,\text{GeV}$ $p_{\text{T}}(\ell) > 5\,\text{GeV}$ $\lvert\eta(\ell)\rvert < 2.5$
Z and ZZ	$40 < m_{\text{Z}_1} < 120\,\text{GeV}$, $4 < m_{\text{Z}_2} < 120\,\text{GeV}$ $m_{\ell\ell} > 4\,\text{GeV}$ (any oppositely charged same-flavor pair)
Jets	$p_{\text{T}}(\text{j}) > 30\,\text{GeV}$ $\lvert\eta(\text{j})\rvert < 4.7$ $\Delta R(\ell, \text{j}) > 0.4$ for each ℓ, j
	On-shell ZZ region
Z and ZZ	ZZ base selection + $60 < m_{\text{Z}_1,\text{Z}_2} < 120\,\text{GeV}$
	Full $m_{4\ell}$ range
Z and ZZ	ZZ base selection + $m_{4\ell} > 80\,\text{GeV}$

5.2.1 Unfolding

The data are unfolded with the iterative D'Agostini's method [176] including correction for background contributions, using the `RooUnfold` toolkit as described in Ref. [177]. In the notation of iterative D'Agostini's method, consider a set of causes $(C_1, C_2, ..., C_{n_C})$ corresponding to truth-level events, and a set of effects $(E_1, E_2, ..., E_{n_E})$ corresponding to measured events. If we observe the number of events in each effect to be $(n(E_1), n(E_2), ..., n(E_{n_E}))$, then the number of events in each cause $\hat{n}(C_i)$ can be estimated by

$$\hat{n}(C_i) = \frac{1}{\epsilon_i} \sum_{j=1}^{n_E} P(C_i|E_j) n(E_j) \equiv \sum_{j=1}^{n_E} M_{ij} n(E_j), \tag{5.3}$$

where $P(C_i|E_j)$ is the conditional probability of the cause being C_i given the effect is E_j, and $\epsilon_i \equiv \sum_{j=1}^{n_E} P(E_j|C_i)$ is the efficiency of C_i causing detection in any of the possible effects. $P(E_j|C_i)$ is the conditional probability of C_i causing detection of E_j, and is referred to as the normalized response matrix. The response matrix represents the correlation map between the distributions obtained after the full detector simulation, reconstruction, and selection, and the generated distributions they originate from. It can be evaluated from simulation with

$$P(E_j|C_i) = \frac{P(E_j \cap C_i)}{P(C_i)} = \frac{n_{ij}^{MC}}{n_i^{MC}} \tag{5.4}$$

where n_{ij}^{MC} (the response matrix without normalization) is the number of signal MC events that are detected in E_j and originate from C_i, and n_i^{MC} is the number of signal MC events in C_i. M_{ij} is called the unfolding matrix. A sample of n_{ij}^{MC} for 2018 results are shown in Figures 5.1–5.4, where these plots are validation plots that enter the CMS internal analysis note for the ZZ+jets analysis. The `MadGraph5_aMC@NLO` $q\bar{q} \to ZZ$ and MCFM $gg \to ZZ$ sample is used in the creation of these response

matrices. We can see that they are close to being diagonal and we expect good unfolding performance.

Using Bayes' theorem, $P(C_i|E_j)$ can be written as

$$P(C_i|E_j) = \frac{P(E_j|C_i)P_0(C_i)}{\sum_{l=1}^{n_C} P(E_j|C_l)P_0(C_l)} = \frac{P(E_j|C_i)n_0(C_i)}{\sum_{l=1}^{n_C} P(E_j|C_l)n_0(C_l)}, \tag{5.5}$$

where $P_0(C_i)$ and $n_0(C_i)$ are the prior probability and event distribution, respectively, and are evaluated from signal MC truth-level event distribution as the initial input. After evaluating $\hat{n}\,(C_i)$ with equations (5.3–5.5), we can calculate

$$\hat{N}_{\text{true}} = \sum_{i=1}^{n_C} \hat{n}(C_i) \tag{5.6}$$

$$\hat{P}(C_i) = \frac{\hat{n}(C_i)}{\hat{N}_{\text{true}}} \tag{5.7}$$

and replace $P_0(C_i)$ with $\hat{P}(C_i)$, $n_0(C_i)$ with $\hat{n}(C_i)$, and then enter the next iteration to improve the evaluation of $\hat{n}(C_i)$. In this analysis we decide to stop at the 4th iteration. On one hand, we want to have enough number of iterations to avoid bias towards the input $P_0(C_i)$ distribution, and 4 iterations will be sufficient. On the other hand, we want to stop the iteration as soon as a new iteration only introduces little variation to the unfolded results, since more iteration is expected to eventually include larger fluctuation from the data. To estimate the compatibility of unfolded distributions between iterations, a χ^2 is calculated within the RooUnfold package,

$$\chi^2 = \sum_{i=1}^{n_C} \frac{[(\hat{P}(C_i)^{k+1} - \hat{P}(C_i)^k)\hat{N}_{\text{true}}^{k+1}]^2}{\text{Max}[(\hat{P}(C_i)^{k+1} + \hat{P}(C_i)^k)\hat{N}_{\text{true}}^{k+1}, 1]} \tag{5.8}$$

where k is the number of iterations. In studying 2016 unfolded distributions, $\chi^2/D.O.F = \chi^2/n_C$ is at most 0.1 at the 4th iteration for all the unfolded distributions, therefore

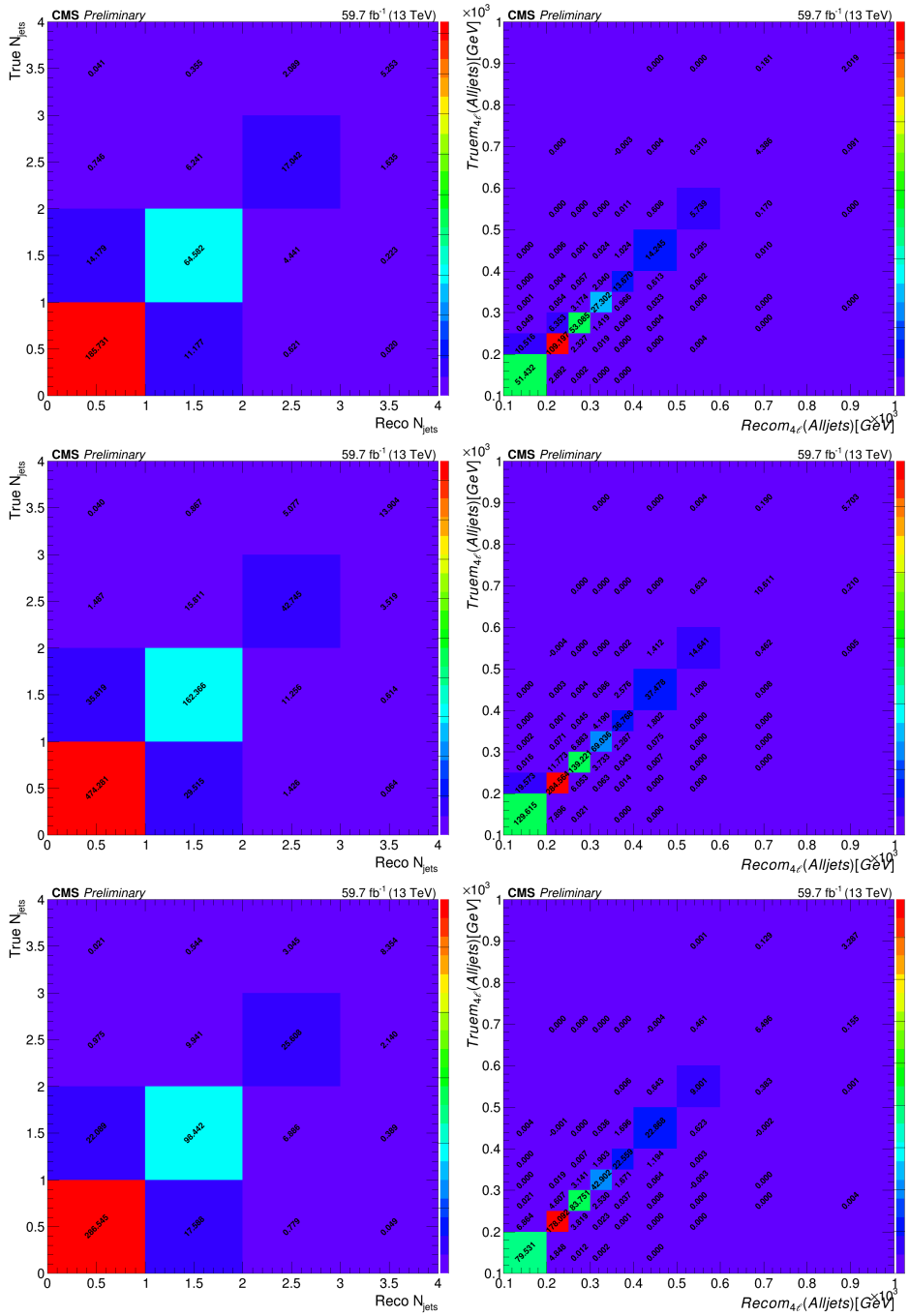

Figure 5.1: Response matrices for number of jets and m_{4l} with 59.7 fb^{-1} in the 4e channel (top), 2e2μ channel (middle), and 4μ channel (bottom). The x-axis is for the reconstruction-level quantity, the y-axis is for the true-level quantity.

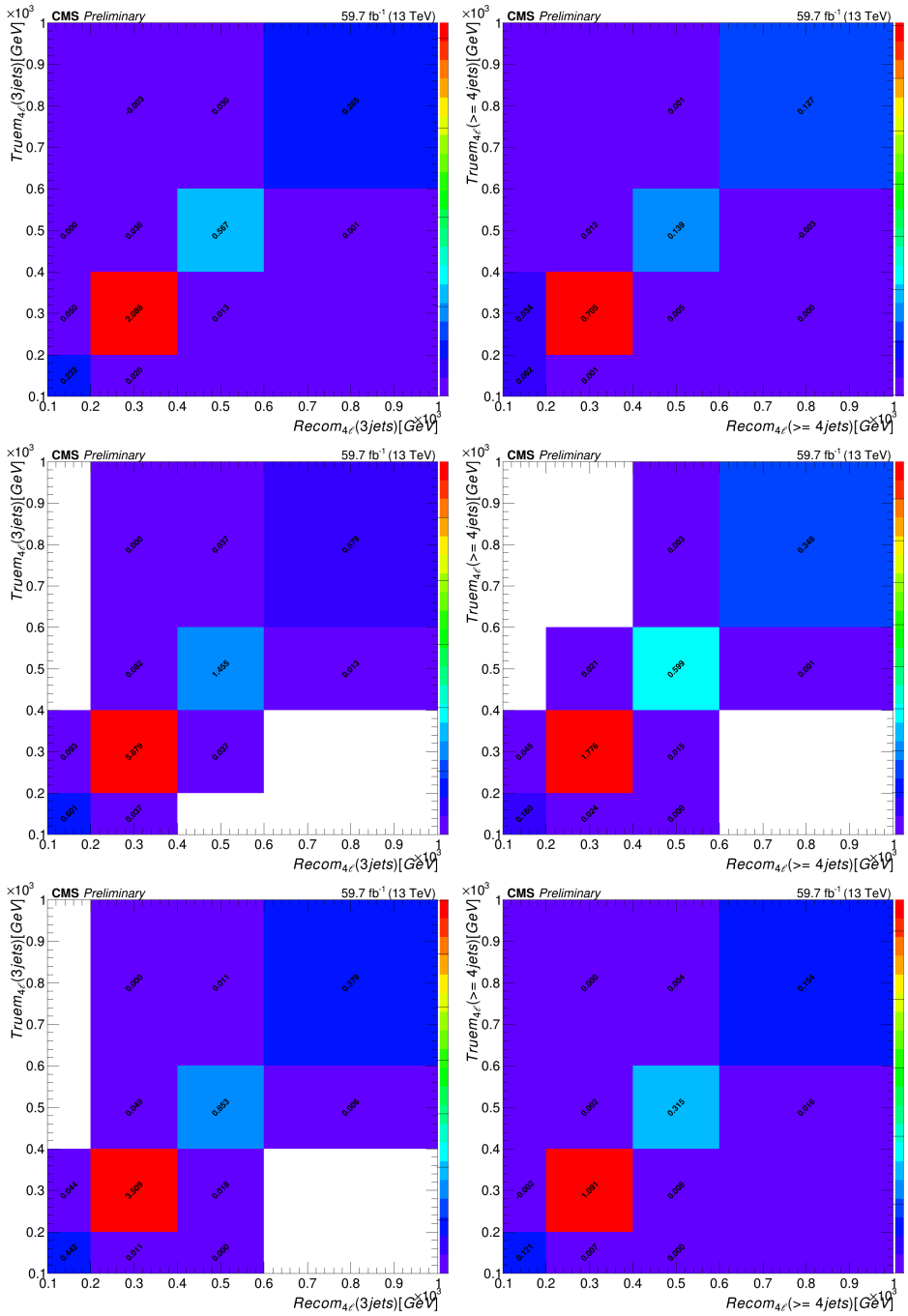

Figure 5.2: Response matrices for m_{4l} in events with 3 and at least 4 jets respectively with 59.7 fb^{-1} in the 4e channel (top), 2e2μ channel (middle), and 4μ channel (bottom). The x-axis is for the reconstruction-level quantity, the y-axis is for the true-level quantity.

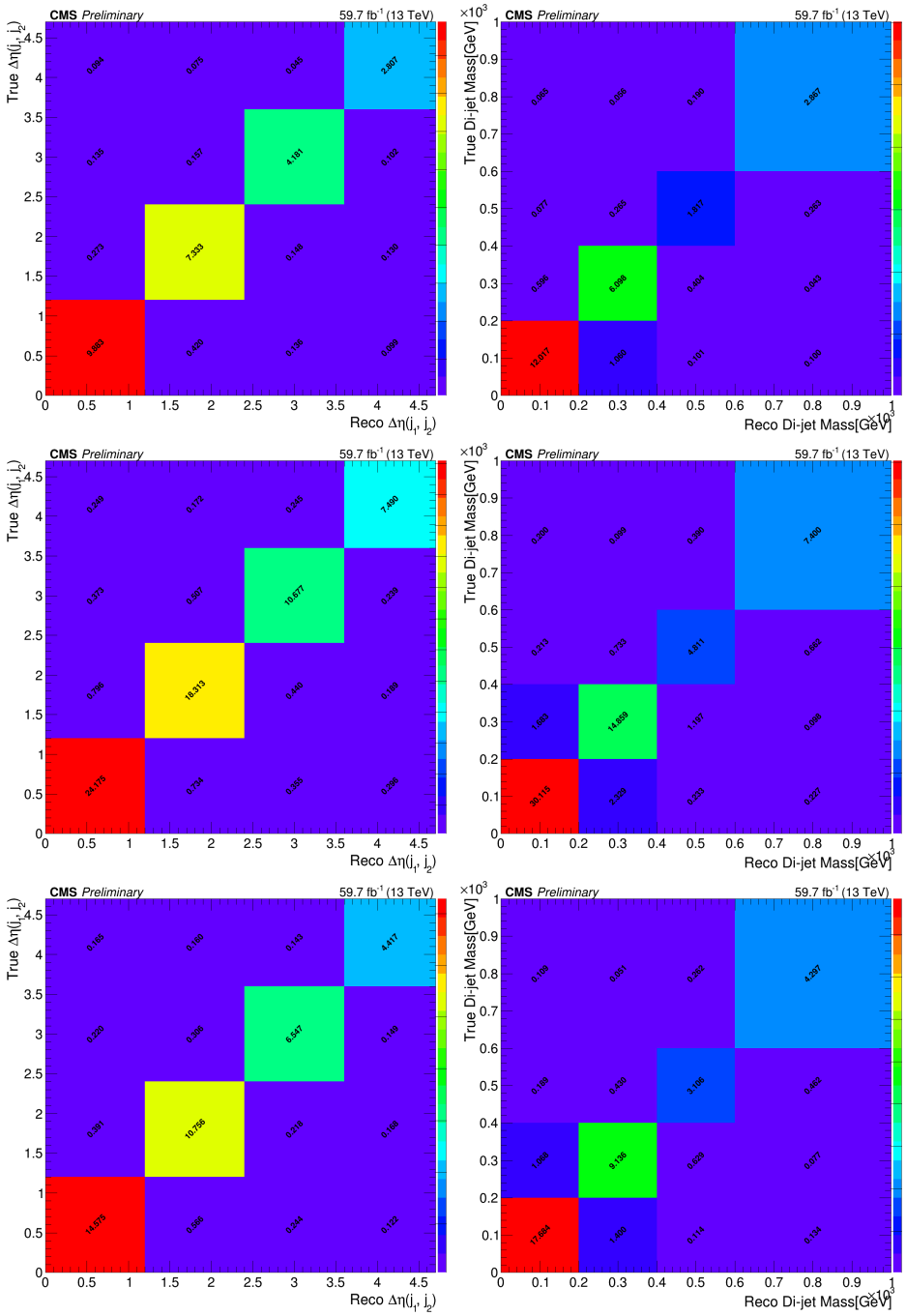

Figure 5.3: Response matrices for $\Delta\eta$ and Di-jet mass between leading$-p_{\mathrm{T}}$ and subleading$-p_{\mathrm{T}}$ jets with 59.7 fb^{-1} in the 4e channel (top), 2e2μ channel (middle), and 4μ channel (bottom). The x-axis is for the reconstruction-level quantity, the y-axis is for the true-level quantity.

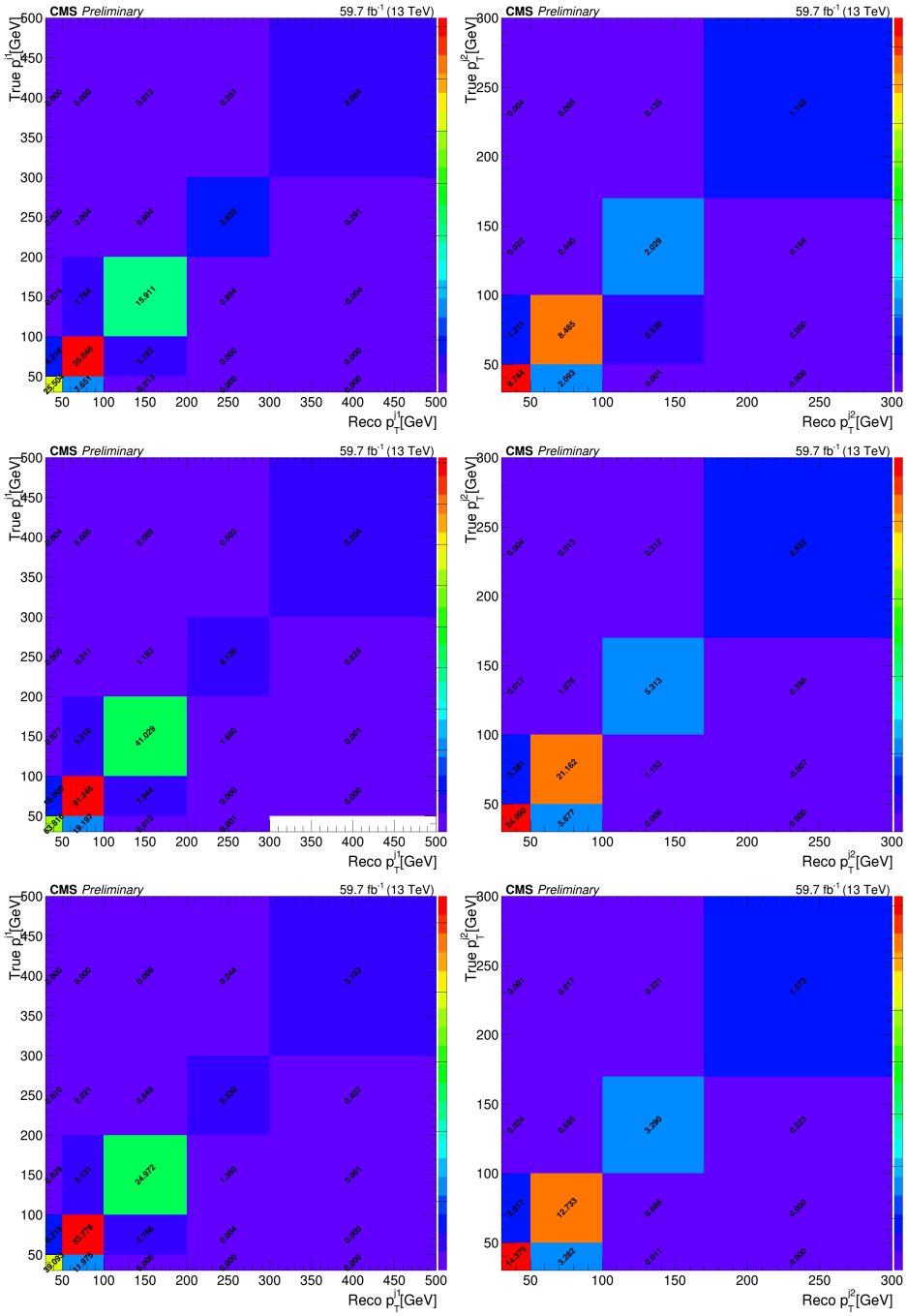

Figure 5.4: Response matrices for p_{T} of leading-p_{T} and subleading-p_{T} jets with 59.7 fb^{-1} in the 4e channel (top), 2e2μ channel (middle), and 4μ channel (bottom). The x-axis is for the reconstruction-level quantity, the y-axis is for the true-level quantity.

further iteration introduces little change to the results and it is reasonable to stop the iteration.

The statistical uncertainty in the unfolded distribution is propagated from the data statistical uncertainty. The covariance matrix of the unfolded distribution associated with statistical uncertainty is given by

$$V(\hat{n}(C_k),\hat{n}(C_l)) = \sum_{i,j=1}^{n_E} \frac{\partial \hat{n}(C_k)}{\partial n(E_i)} V(n(E_i),n(E_j)) \frac{\partial \hat{n}(C_l)}{\partial n(E_j)} \tag{5.9}$$

where $V(n(E_i),n(E_j))$ for experimental data is normally just a diagonal matrix, since we assume measurements in different bins to be uncorrelated. Using equation 5.3, $\frac{\partial \hat{n}(C_i)}{\partial n(E_j)}$ can be calculated as (see Ref. [177])

$$\frac{\partial \hat{n}(C_i)}{\partial n(E_j)} = M_{ij} + \sum_{k=1}^{n_E} M_{ik} n(E_k) \left(\frac{1}{n_0(C_i)} \frac{\partial n_0(C_i)}{\partial n(E_j)} - \sum_{l=1}^{n_C} \frac{\epsilon_l}{n_0(C_l)} \frac{\partial n_0(C_l)}{\partial n(E_j)} M_{lk} \right) \tag{5.10}$$

In the first iteration, $n_0(C_i)$ is independent of experimental data and $\frac{\partial n_0(C_i)}{\partial n(E_j)} = 0$, therefore $\frac{\partial \hat{n}(C_i)}{\partial n(E_j)} = M_{ij}$. This becomes the $\frac{\partial n_0(C_i)}{\partial n(E_j)}$ in the next iteration and the evaluation can continue.

In constructing the response matrix, there are MC events that pass the fiducial selections but do not have corresponding reconstruction-level events that pass reconstruction-level selections ("missed events"), which have been accounted for by the efficiency ϵ_i. There are also MC events that pass the reconstruction-level selections, but do not have corresponding events at particle level that pass the fiducial selections ("fake events"). These out-of-fiducial events are calculated by $n_{\text{fake}}(E_j) = n^{MC}(E_j) - \sum_{i=1}^{n_C} n_{ij}^{MC}$, and are accounted for in the unfolding process by adding an extra cell C_{n_C+1} to the collection of causes, with $n_0(C_{n_C+1}) = \sum_{j=1}^{n_E} n_{\text{fake}}(E_j)$, and adding an extra row to $P(E_j|C_i)$, $P(E_j|C_{n_C+1}) = \frac{n_{\text{fake}}(E_j)}{n_0(C_{n_C+1})}$. Then the unfolding procedure

as described before will include the contribution from these events. The size of the contribution of these out-of-fiducial events can be up to 15% for events with at least one jet. The distribution of the "missed" and "faked" events of several variables are shown in Fig. 5.5–Fig. 5.7, where these figures are validation plots produced during the review of the ZZ+jets analysis within the CMS Collaboration.

In addition, the nonprompt and VVV background events are also added to the out-of-fiducial events, so that these background events can also be accounted for similarly in the unfolding process. If we instead subtract these background events from experimental data directly before unfolding, we can have negative bin content for bins with low statistics, which can cause abnormal behavior in unfolding.

5.3 Systematic Uncertainties

For ZZ+jets study, the systematic uncertainties are propagated through the unfolding by reevaluating the response matrix with the sample used in building the matrix shifted or reweighted to reflect a one standard deviation variation in the quantity of interest. The resulting difference in the final normalized unfolded distributions is taken as the uncertainty related to that quantity. For EW ZZ+2jets signal extraction, which we will formally discuss in Section 5.4 but it appears to be reasonable to discuss its systematic uncertainties here along with the ZZ+jets uncertainties, the systematic uncertainties are accounted for as shape uncertainties or log-normal uncertainties associated with nuisance parameters.

The systematic uncertainty in the trigger efficiency is estimated to be 2%, and cancels out in normalized differential cross sections. To evaluate uncertainties associated with lepton efficiencies, the response matrix and template histogram are reevaluated using lepton efficiency correction factors varied up and down by the tag-

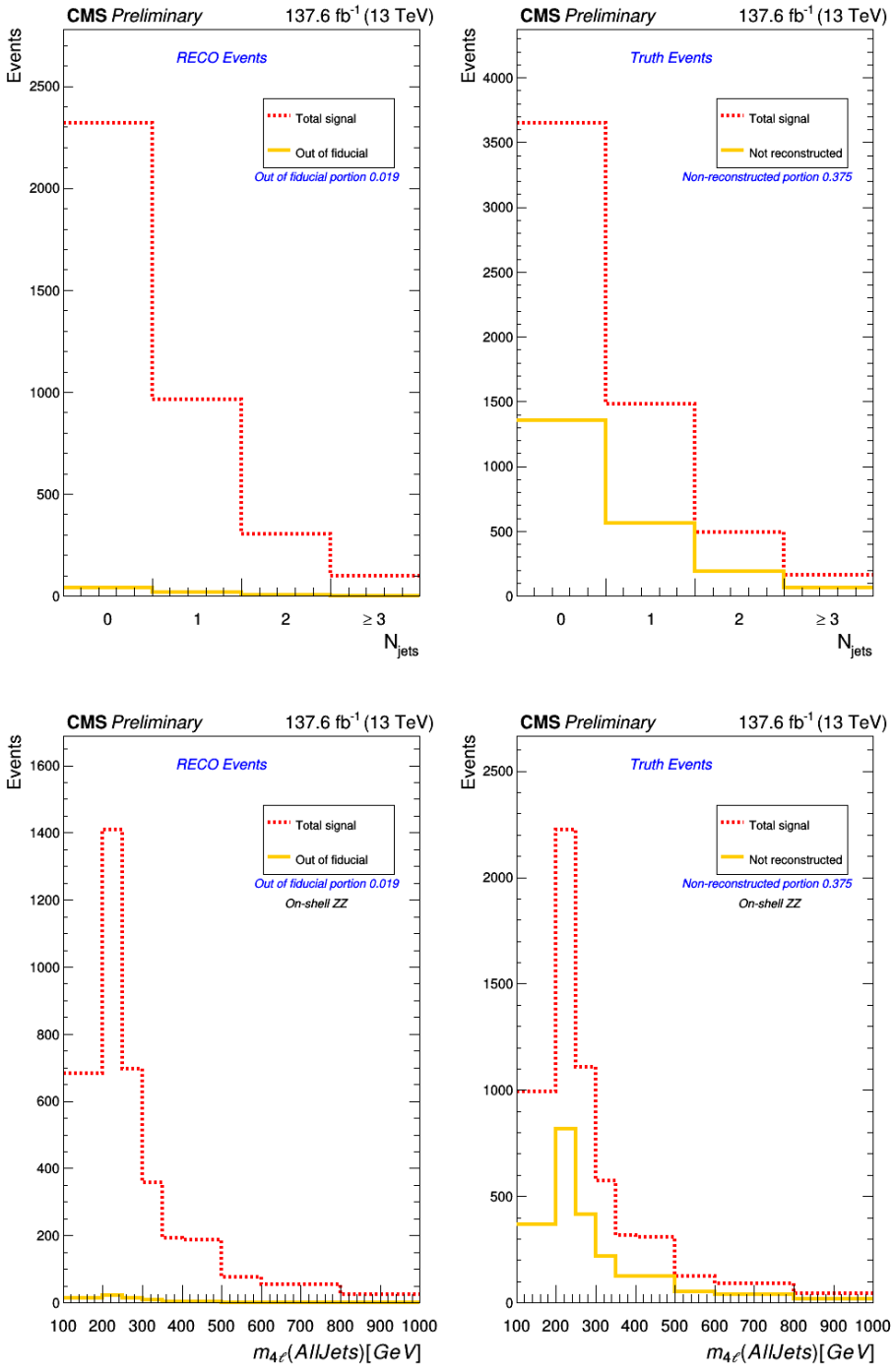

Figure 5.5: Total signal MC reconstructed events vs out-of-fiducial reconstructed events (left) and total signal MC truth-level events vs non-reconstructed truth-level events (right), for N_{jets} and $m_{4\ell}$ in the on-shell ZZ region.

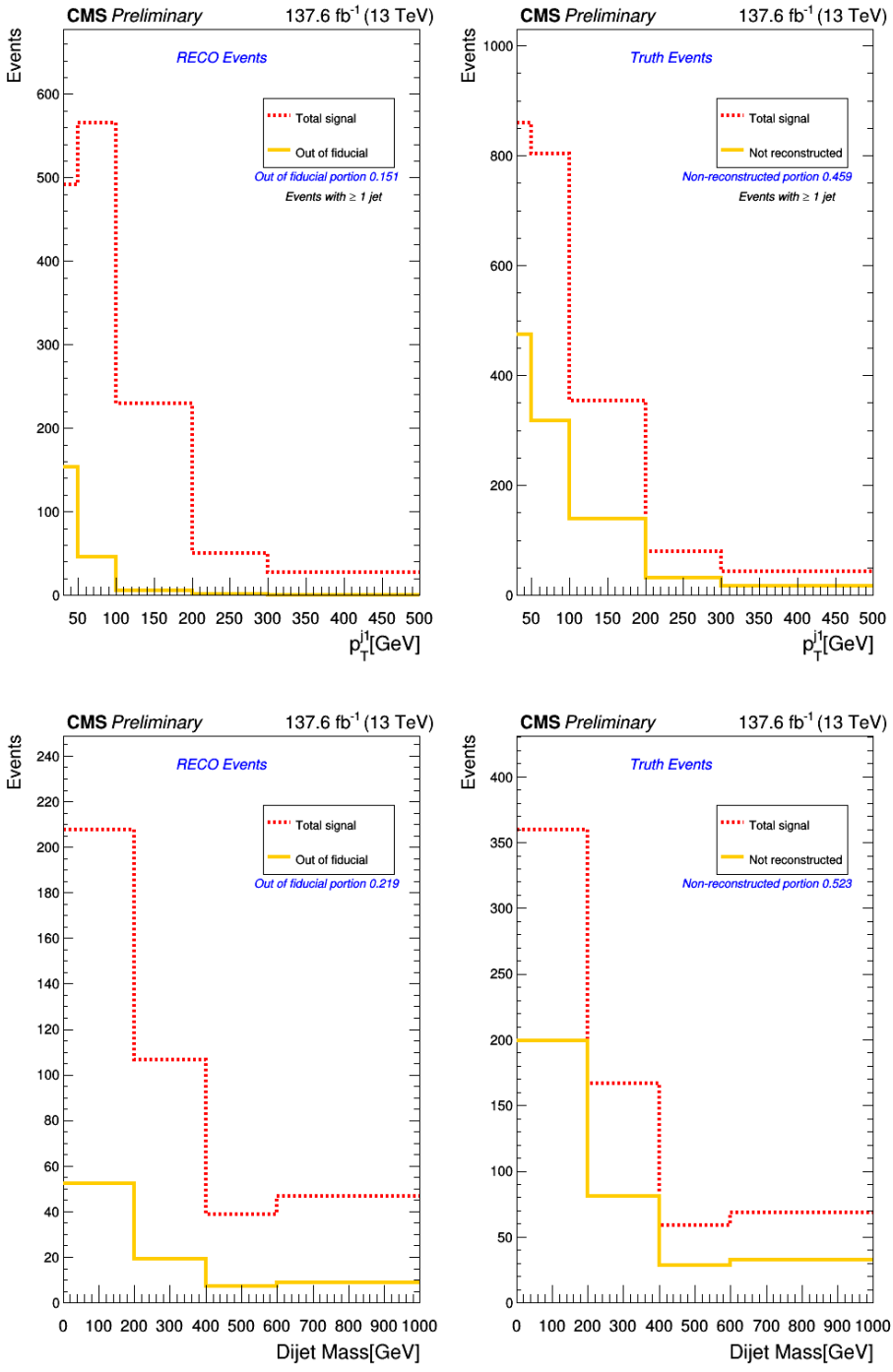

Figure 5.6: Total signal MC reconstructed events vs out-of-fiducial reconstructed events (left) and total signal MC truth-level events vs non-reconstructed truth-level events (right), for p_{T} of the highest-p_{T} jet (events with at least one jet) and dijet mass (events with at least two jets), in the on-shell ZZ region.

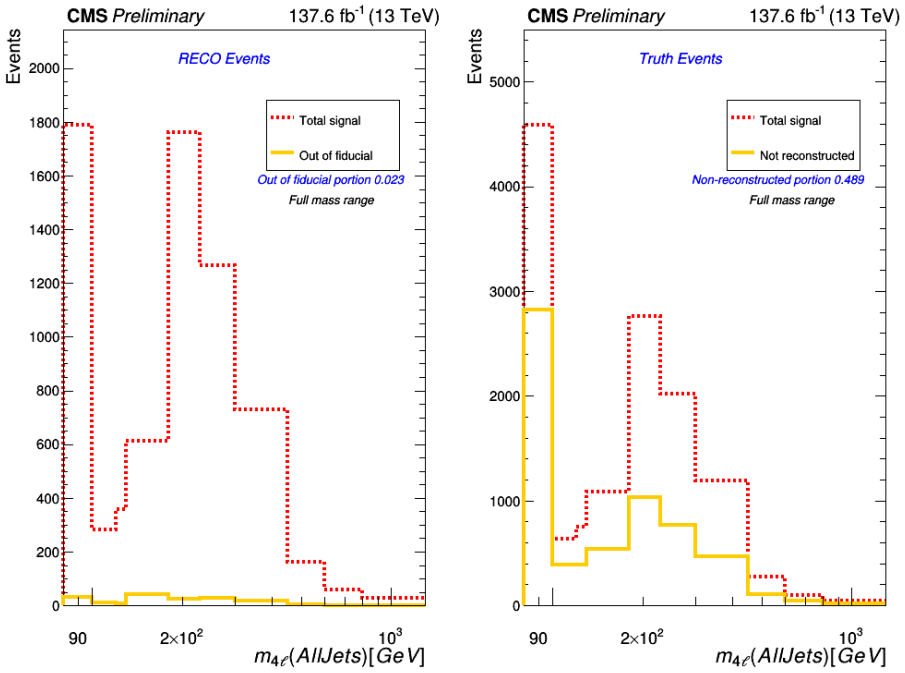

Figure 5.7: Total signal MC reconstructed events vs out-of-fiducial reconstructed events (left) and total signal MC truth-level events vs non-reconstructed truth-level events (right), for $m_{4\ell}$ in the full $m_{4\ell}$ range.

and-probe [171] fit uncertainties. Electrons and muons are treated separately, and all leptons of the same type are treated as correlated. For the uncertainties associated with the JES and jet energy resolution (JER), the jet p_{T} is varied by shifting the JES and the spreading up and down by their uncertainties, and the response matrix and template histogram are reevaluated.

The uncertainty in lepton fake rates is 40%, and is dominated by the statistical uncertainty but also includes systematic uncertainties associated with the underlying physics processes between events in the 3ℓ and 4ℓ control regions. The reducible background is varied up and down by the lepton fake rate uncertainty (40%) and the unfolding is repeated to estimate the associated uncertainty from the difference between the normalized distributions. This uncertainty is treated as a log-normal uncertainty for EW ZZ+2jets signal extraction.

The pileup uncertainty is evaluated by recomputing the response matrix and template histogram with the total inelastic cross section [178], which defines the pileup weights applied to MC, varied up and down by 4.6%. The uncertainty associated with the luminosity is evaluated by reevaluating the response matrix with the simulation normalized to the integrated luminosity varied up and down by its total uncertainty, which is 1.2, 2.3 and 2.5% for 2016, 2017 and 2018, respectively, as mentioned in Section 2.2.7. It is small for unfolding as expected due to the cancellation from the normalization by the fiducial cross section. This uncertainty is treated as a log-normal uncertainty for EW ZZ+2jets signal extraction.

The uncertainty arising from generator-specific modeling differences is evaluated from the difference between the measurements unfolded with the response matrix based on the $q\bar{q} \to ZZ$ sample simulated by MadGraph5_aMC@NLO and the POWHEG sample.

The PDF and related strong coupling (α_S) uncertainties are evaluated by reweighting the MadGraph5_aMC@NLO sample to PDF and α_S variations, and then redoing the unfolding and template histogram filling, and combining the results according to the procedure described in Ref. [179]. For the renormalization (μ_R) and factorization (μ_F) scales (QCD scales) uncertainties, the response matrix and template histogram are reevaluated with the MadGraph5_aMC@NLO $q\bar{q} \to ZZ$ sample reweighted to reflect the distribution with μ_F and μ_R independently varied up and down by a factor of 2. All combinations are considered except those in which μ_F and μ_R differ by a factor of four, and the envelope of all variations is used.

The normalization of the MCFM sample ($gg \to ZZ$) is varied by the scale and PDF uncertainties of its cross section ($^{+18\%}_{-14\%}$), and the resulting difference between the normalized distributions is used for the ZZ+jets unfolding uncertainty. For EW ZZ+2jets signal extraction, we model the associated uncertainty as a log-normal

uncertainty.

The contributions of each source of systematic uncertainty to the final ZZ+jets results are summarized in Tables 5.2–5.4, and the impact plot for EW ZZ+2jets signal extraction is shown in Fig. 5.10 (located in Section 5.4). The numbers in Tables 5.2–5.4 are only indicative. They are estimated by varying each source and obtaining the difference in the normalized unfolded distributions. Each number in the tables is not showing an estimate of uncertainty per bin, but an estimate of the portion of the uncertainty contribution per distribution, given by

$$\frac{\sum_{i=1}^{N_{\text{bins}}} |h_{\text{central}}(i) - h_{\text{varied}}(i)|}{\sum_{i=1}^{N_{\text{bins}}} h_{\text{central}}(i)} \tag{5.11}$$

where h_{central} and h_{varied} are the central and varied histograms, respectively, both with sum of bin contents normalized to 1, and N_{bins} is the total number of bins. There are, in general, two estimates from up/down variations and the larger one is used.

Table 5.2: Contributions of each source of systematic uncertainty to the normalized differential cross section measurements of jet variables for ZZ+jets. Uncertainties depend on the distributions and are listed as a range.

Systematic source	Uncertainty range
Electron efficiency	0.13–0.30%
Muon efficiency	0.02–0.08%
Jet energy resolution	1.65–3.85%
JES correction	0.93–5.32%
Reducible background	0.05–0.43%
Pileup	0.04–1.08%
Luminosity	$< 0.03\%$
$q\bar{q} \to ZZ$ MC choice	0.52–4.52%
$gg \to ZZ$ cross section	0.01–0.19%
QCD scales	0.16–0.82%
PDF	0.05–0.12%
PDF α_S	0.01–0.10%

Table 5.3: The contributions of each source of systematic uncertainty in the normalized differential cross sections measurements for ZZ+jets as a function of $m_{4\ell}$ with jet multiplicity from 0 to 3 and more, in events satisfying $60 < m_{Z_1,Z_2} < 120\,\text{GeV}$.

Systematic source	$m_{4\ell}$ with all jets	0 jet	1 jet	2 jets	3 and more jets
Electron efficiency	0.42%	0.38%	0.66%	0.36%	0.26%
Muon efficiency	0.05%	0.06%	0.07%	0.09%	0.08%
Jet energy resolution	–	0.07%	1.72%	1.65%	0.80%
JES correction	–	0.17%	1.77%	1.95%	0.97%
Reducible background	0.18%	0.18%	0.32%	0.33%	0.96%
Pileup	0.02%	0.05%	0.11%	0.13%	0.35%
Luminosity	0.01%	0.01%	0.02%	0.02%	0.05%
$q\bar{q} \to ZZ$ MC choice	0.35%	0.65%	0.94%	0.48%	0.35%
$gg \to ZZ$ cross section	0.02%	0.03%	0.09%	0.06%	0.09%
QCD scales	0.15%	0.16%	0.58%	0.54%	0.62%
PDF	0.05%	0.05%	0.15%	0.15%	0.21%
PDF α_S	0.02%	0.01%	0.05%	0.03%	0.02%

Table 5.4: The contributions of each source of systematic uncertainty in the normalized differential cross sections measurements for ZZ+jets as a function of $m_{4\ell}$ with jet multiplicity from 0 to 3 and more, in events from the full $m_{4\ell}$ range.

Systematic source	$m_{4\ell}$ with all jets	0 jet	1 jet	2 jets	3 and more jets
Electron efficiency	2.12%	2.55%	2.28%	1.77%	1.46%
Muon efficiency	0.71%	0.78%	0.92%	0.79%	0.42%
Jet energy resolution	–	0.11%	1.73%	2.63%	2.32%
JES correction	–	0.33%	1.64%	3.01%	2.02%
Reducible background	2.22%	2.19%	2.88%	3.40%	5.09%
Pileup	0.21%	0.28%	0.19%	0.32%	0.52%
Luminosity	0.12%	0.12%	0.16%	0.17%	0.25%
$q\bar{q} \to ZZ$ MC choice	0.57%	0.48%	1.22%	3.07%	4.21%
$gg \to ZZ$ cross section	0.10%	0.18%	0.61%	0.80%	0.46%
QCD scales	0.27%	0.25%	0.67%	1.25%	1.86%
PDF	0.07%	0.09%	0.20%	0.23%	0.28%
PDF α_S	0.08%	0.08%	0.15%	0.20%	0.28%

5.4 Electroweak Signal Extraction

The plots in this section regarding the EW ZZ+2jets analysis are based on CMS data and simulation, but not formally approved by the CMS Collaboration. For EW ZZ+2jets production, We want to use a deep neural network (DNN) to distinguish

EW ZZ+2jets from QCD-induced ZZ+2jets events with a set of commonly studied observables for VBS in each event. A DNN is a model consisting of multiple layers that is used to learn increasingly meaningful representations of data through successive layers (see [180] for terminologies in the discussion of this section). Typically we feed training data (which is not referring to experimental data here, as contrary to the common usage of this term in high energy physics in the other parts of this thesis) into the neural network so that it can recognize patterns inside the data, and we monitor the training process with validation data. Then we apply the trained DNN to the remaining MC events not used for training (we refer to as the testing data), to test the DNN and obtain template distribution histograms. To build and apply the neural network, we use `Keras` [181], a convenient high-level neural networks API written in Python, which in our case runs on top of TensorFlow [182], an open source machine learning platform.

With `Keras`, we build a simple neural network suitable for binary classification with 3 fully connected (dense) layers. The first two layers have 32 hidden units and use relu activation, while the third layer outputs a single scalar value (predicted probability) with sigmoid activation. For optimizer we use RMSprop with learning rate 0.001, and for the loss function we use "binary_crossentropy" which is suitable for binary classification. We preprocess the input data samples by normalizing each observable to have a mean equal to 0 and a standard deviation equal to 1 across the training and validation data, and then repeat the same operation on the testing data (using mean and standard deviation of the training and validation data for normalization, instead of the testing data). We train for 20 epochs with batch size 512.

MC samples discussed in Section 3.4 are used to prepare the training data for DNN. We mix around 90,000 EW ZZ+2jets, 70,000 MadGraph5_aMC@NLO $q\bar{q} \rightarrow$

ZZ and 70,000 gg $\rightarrow$ ZZ raw MC events from 2016 and 2017 MC samples in 4e,4μ and 2e2μ channels, with each event tagged as either signal or background event. 12 observables are used for training to separate signal and background, chosen by referring to Refs. [4, 183]: m_{jj}, $\Delta\eta_{jj}$, $m_{4\ell}$, $\Delta\phi(\mathrm{Z}_1, \mathrm{Z}_2)$, η^{j1}, η^{j2}, p_{T}^{j1}, p_{T}^{j2}, N_{jets}, $\eta^{j1}\times\eta^{j2}$, and the two Zeppenfeld variables defined as $\eta^{\mathrm{Z}_i} - \frac{1}{2}(\eta^{j1} + \eta^{j2})$, $i = 1, 2$. These MC events are put into the DNN built with `Keras` for training, with 26,000 events out of these training events used for validation. The trained DNN is then applied on the 2018 MC events to test performance and form histogram templates. The DNN diagnostic plots during training and the ROC curve from the test results are shown in Fig. 5.8. There doesn't appear to be overfitting and the accuracy is at a reasonable level. The ROC area under the curve (AUC) is 0.919, indicating good separation power of the trained DNN.

The template histograms are then input into the CMS statistical analysis tool COMBINE [184] to perform a maximum likelihood fit to the full Run 2 data distribution of the DNN predicted probability, in order to extract the signal strength of EW ZZ+2jets production and estimate the significance (see Section 6.3 for results). The pre-fit and post-fit distributions are shown in Fig 5.9, where the uncertainties associated with each parameter in the likelihood is estimated in the COMBINE tool by calculating the covariance matrix at the best-fit point (more accurate uncertainty estimation in COMBINE is used for the final signal strength value). We also want to evaluate the impact of each systematic uncertainty, which is the shift in the signal strength with respect to the best-fit value, that is induced if a given nuisance parameter is shifted by its $\pm1\sigma$ post-fit uncertainty values. The impact plot is shown in Fig. 5.10. We see that the systematic uncertainties associated with QCD scales and JES, JER have the largest impact and they are also noticeably constrained by the fit.

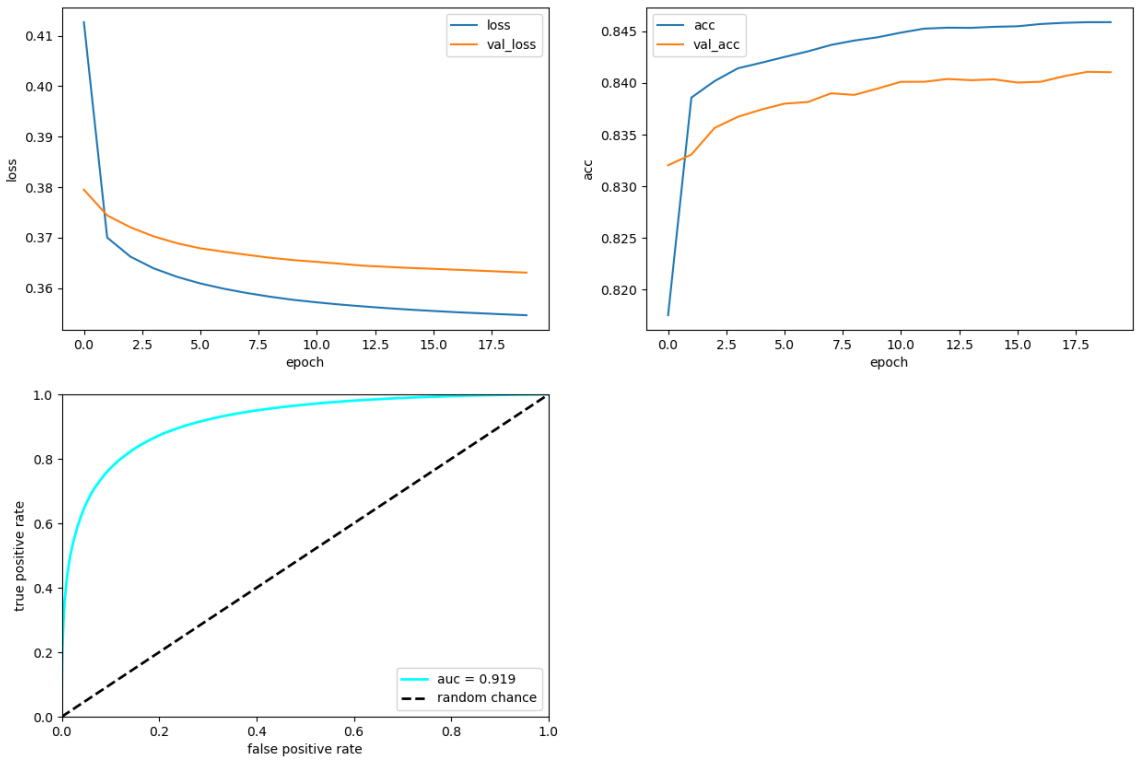

Figure 5.8: Loss function value vs epoch (top left), accuracy vs epoch (top right), and ROC curve (bottom left).

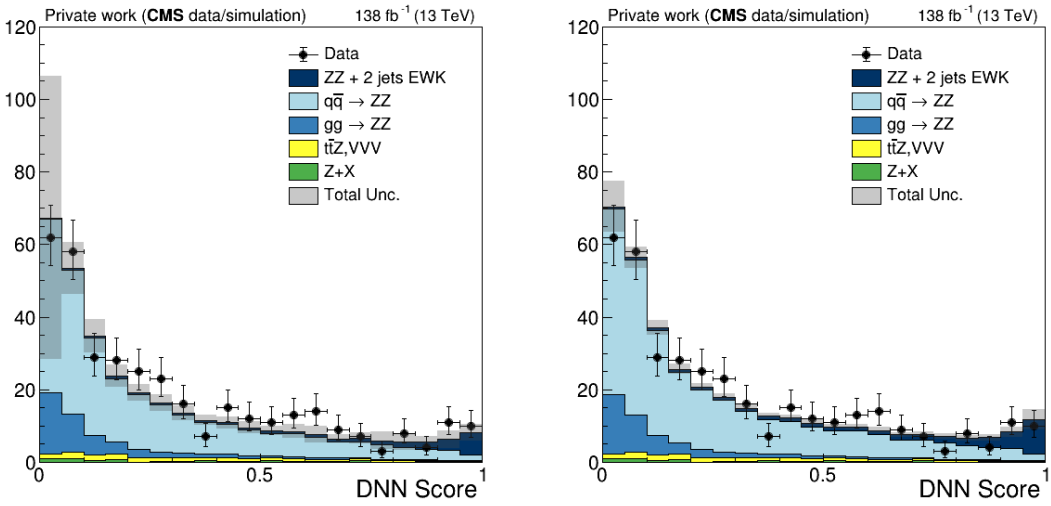

Figure 5.9: The pre-fit (left) and post-fit (right) data vs MC distribution of the DNN predicted probability. The uncertainties in these two plots are estimated with the Minuit HESSE routine used by the COMBINE tool, by calculating the covariance matrix at the best-fit point.

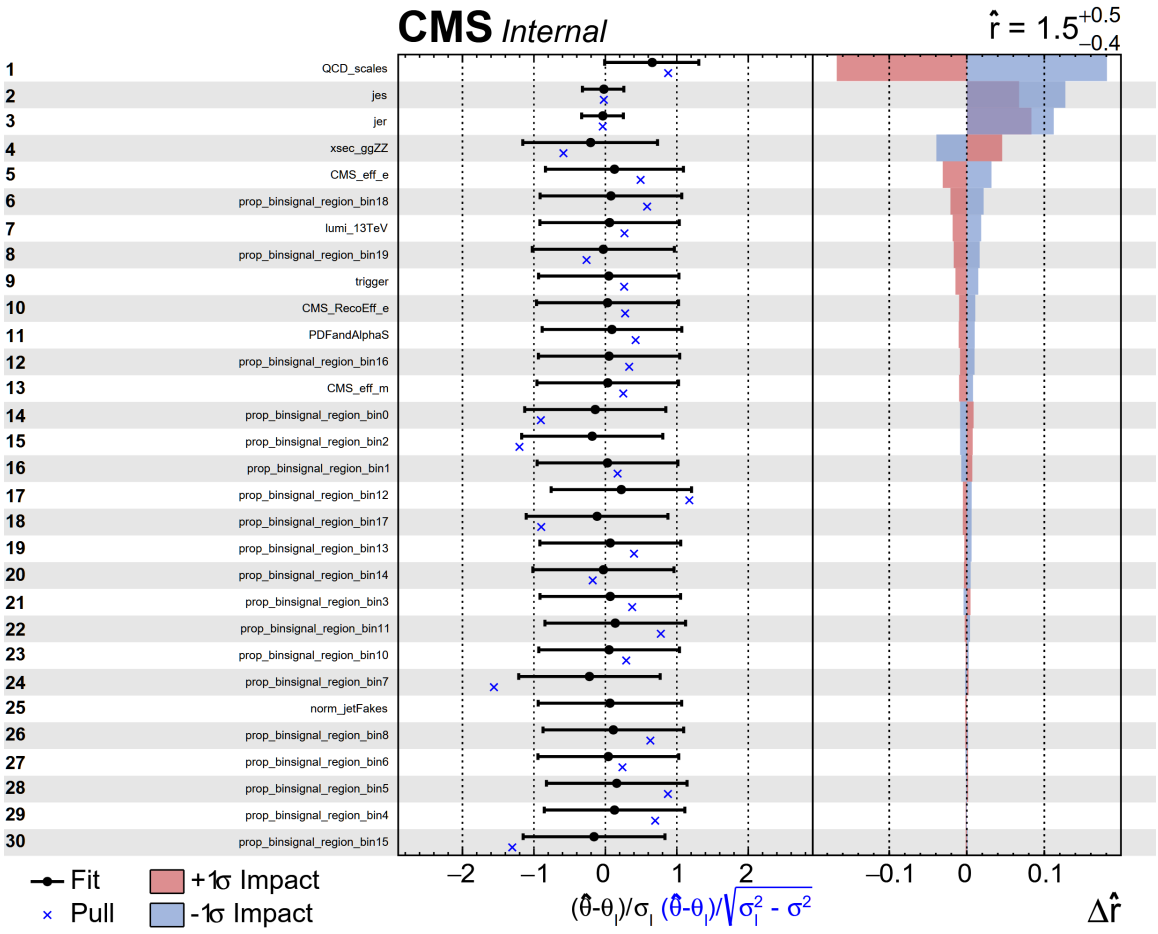

Figure 5.10: Impact plots made with COMBINE tool for uncertainties with significant contributions. The red and blue bars indicate the imapcts. The black dot indicates $\frac{\theta-\theta_I}{\sigma_I}$, where θ and θ_I are the post-fit and pre-fit values of the nuisance parameter and σ_I is the pre-fit uncertainty. The asymmetric black error bar shows the post-fit uncertainty divided by the pre-fit uncertainty, meaning that parameters with error bars smaller than ± 1 are constrained in the fit. The blue cross shows the pull, as defined at the legend.

Chapter 6

Results

In this chapter we present the results for the ZZ+jets and EW ZZ+2jets analyses. For the ZZ+jets results, we have used the text directly from Ref. [1] up to minor modifications. On the other hand, the results regarding the EW ZZ+2jets analysis presented in this chapter are based on CMS data and simulation, but not formally approved by the CMS Collaboration, and they serve as a cross check for the CMS published analysis [4].

6.1 ZZ +jets Differential distributions

Differential distributions for various reconstructed quantities are presented in this subsection. We proceed with unfolding the data to compare directly with particle-level theoretical predictions, and the results are presented in the next subsection. Figure 6.1 (left) shows the number of reconstructed jets with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ for the ZZ+jets events with $60 < m_{\mathrm{Z}_1,\mathrm{Z}_2} < 120\,\mathrm{GeV}$. The last bin includes all events with three or more jets. The description of events with three and more jets requires NNLO and even higher corrections, but there are not enough hard jets from the

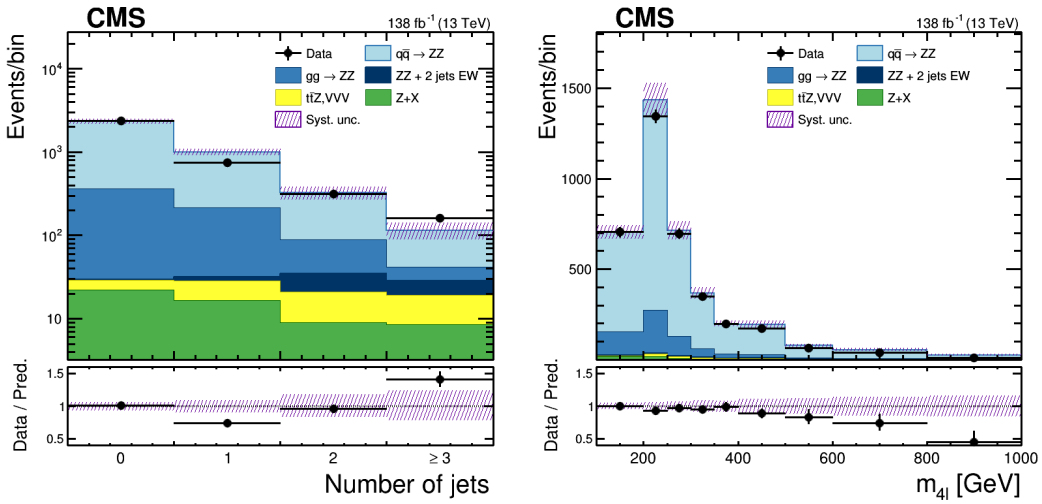

Figure 6.1: Distribution of the number of jets with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ (left) and of m_{ZZ} (right) for ZZ+jets events with $60 < m_{\mathrm{Z_1,Z_2}} < 120\,\mathrm{GeV}$ for the combined 4e, 4μ, and 2e2μ decay channels. Points represent the data, vertical bars the statistical uncertainties, and shaded histograms represent the expected standard model predictions and reducible background estimated from data. The purple band of slashes represents the systematic uncertainties in the predictions, which includes systematic uncertainties associated with trigger efficiency, lepton efficiencies, jet energy correction and jet energy resolution, pileup, luminosity, Monte Carlo generator choice, $\mathrm{gg} \to \mathrm{ZZ}$ cross section, and reducible background. The overflow is included in the last bin of the distributions.

matrix element in the MC samples used, therefore the difference between data and predictions at high jets multiplicity is expected. The 0 and 2 jet bins are well described by the predictions, whereas in the 1 jet bin the predictions significantly overestimate the measured event yield. The $m_{4\ell}$ distribution is shown in Fig. 6.1 (right), inclusive in the number of jets. This distribution is well described by the predictions, except for the increasing discrepancy between data and MC towards high masses; this can be mitigated by adding the EW corrections, as demonstrated in Ref. [5] and in the next subsection.

Figure 6.2 shows the p_{T} and $|\eta|$ distributions for the highest- and second-highest-p_{T} jet in events with at least one and two jets, respectively. As expected from the distribution of the number of jets, the predictions overestimate the measurements in

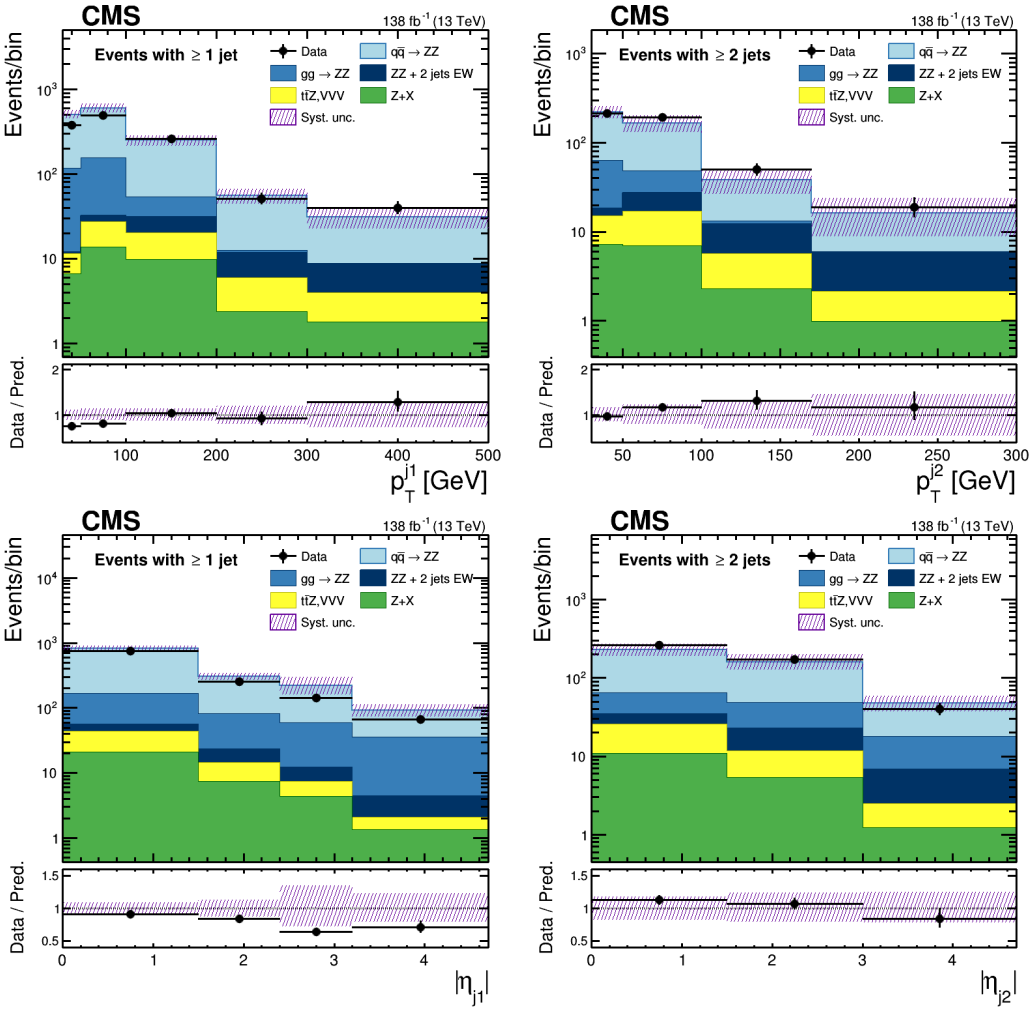

Figure 6.2: Distribution of the p_{T} of the highest-p_{T} jet (upper left) in events with at least one jet, and of the p_{T} of the second-highest-p_{T} jet (upper right) in events containing at least two jets. The $|\eta|$ distribution of the highest-p_{T} (lower left) and second-highest-p_{T} (lower right) jets. Events are required to have $60 < m_{\mathrm{Z}_1,\mathrm{Z}_2} < 120\,\mathrm{GeV}$. Other details are as in the caption of Fig. 6.1.

the highest-p_{T} jet distributions. The largest difference is observed for highest-p_{T} jets with $p_{\mathrm{T}} < 100\,\mathrm{GeV}$, whereas the second-highest-p_{T} distributions are better described. Apart from the difference in the yield, the p_{T} distributions of both the highest-p_{T} and second-highest-p_{T} jets show similar differential behavior with respect to the predictions (similar trend up to 300 GeV), which is demonstrated in the lower panels of the figures, where data-to-prediction ratios are presented. Similar conclusions are valid also for the $|\eta|$ distributions.

The invariant mass of the dijet system and $\Delta\eta$ between two jets with highest p_{T} are among the most important dijet distributions. The dijet mass distribution is well described by predictions, as shown in Fig. 6.3 (left), whereas in the $|\Delta\eta|$ distribution there is a small trend between data and predictions that can be seen in the lower panel of Fig. 6.3 (right). As expected, the contribution of the EW ZZ+2jets production is increasing towards the high jet separation and dijet mass, but still remains a small part of the total ZZ cross section.

The effect of the presence of jets in ZZ events is also studied using the $m_{4\ell}$ distribution for different jet multiplicities (Figs. 6.4, 6.5). Each distribution contains only events with exactly 0, 1, 2, 3 jets or $\geq$4 jets. The predictions describe well the normalized differential behavior, but fail to describe the event yield in the 1-jet case. With increasing jet multiplicities the predicted yields decrease much faster than the measured ones. In the case of 4 or more jets, the data yields are significantly larger than predicted. All distributions in Fig. 6.4 are presented for events with on-shell Z bosons, $60 < m_{\mathrm{Z_1,Z_2}} < 120\,\mathrm{GeV}$.

The same analysis is repeated in the full $m_{4\ell}$ range and the results are presented in Fig. 6.5. The data and MC predictions are compared in three mass regions: Z boson region, Higgs boson region, and nonresonant ZZ production region. It is important to note that the Higgs boson sample is simulated using the POWHEG NLO predictions,

whereas a similar contribution in the ZZ sample has the $gg \to ZZ$ process simulated at LO and normalized to NLO prediction (see Section 3.4 for detail). As shown in Fig. 6.5 (upper left), the predictions describe well the data that are inclusive in jet multiplicity. In Fig. 6.5 (middle left), the predictions do not describe the event yield of the ZZ nonresonant part, but agree well with data in the Z and Higgs boson production regions. With increasing jet multiplicity, the agreement between data and predictions for ZZ and Z production regions becomes worse, whereas the predictions for the Higgs boson region are compatible with the data within large statistical uncertainties.

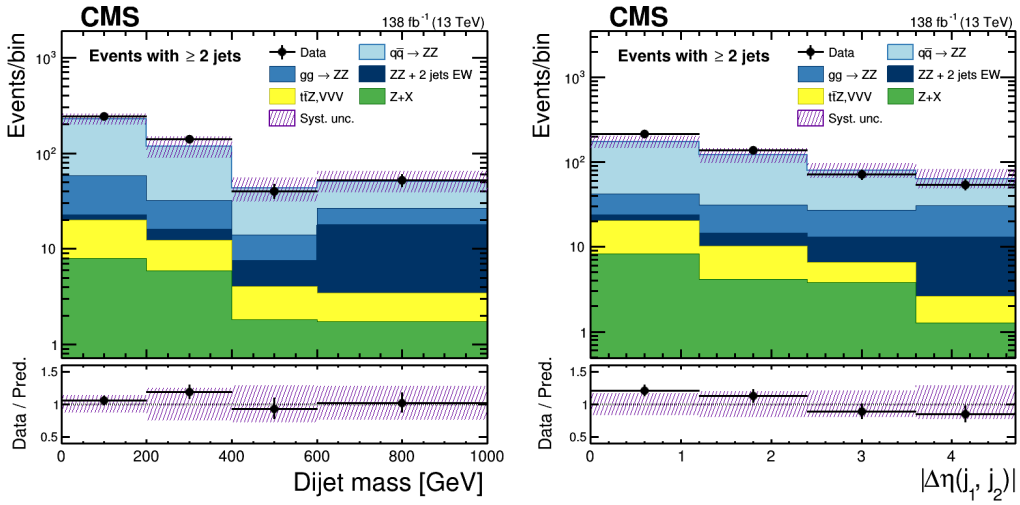

Figure 6.3: The dijet mass (left) and $|\Delta\eta|$ (right) between the two highest-p_{T} jets in events with at least two jets. Events are required to have $60 < m_{\mathrm{Z}_1,\mathrm{Z}_2} < 120\,\mathrm{GeV}$. Other details are as in the caption of Fig. 6.1.

The measured and expected event yields for all decay channels and jet multiplicities in different mass ranges are summarized in Tables 6.1 and 6.2.

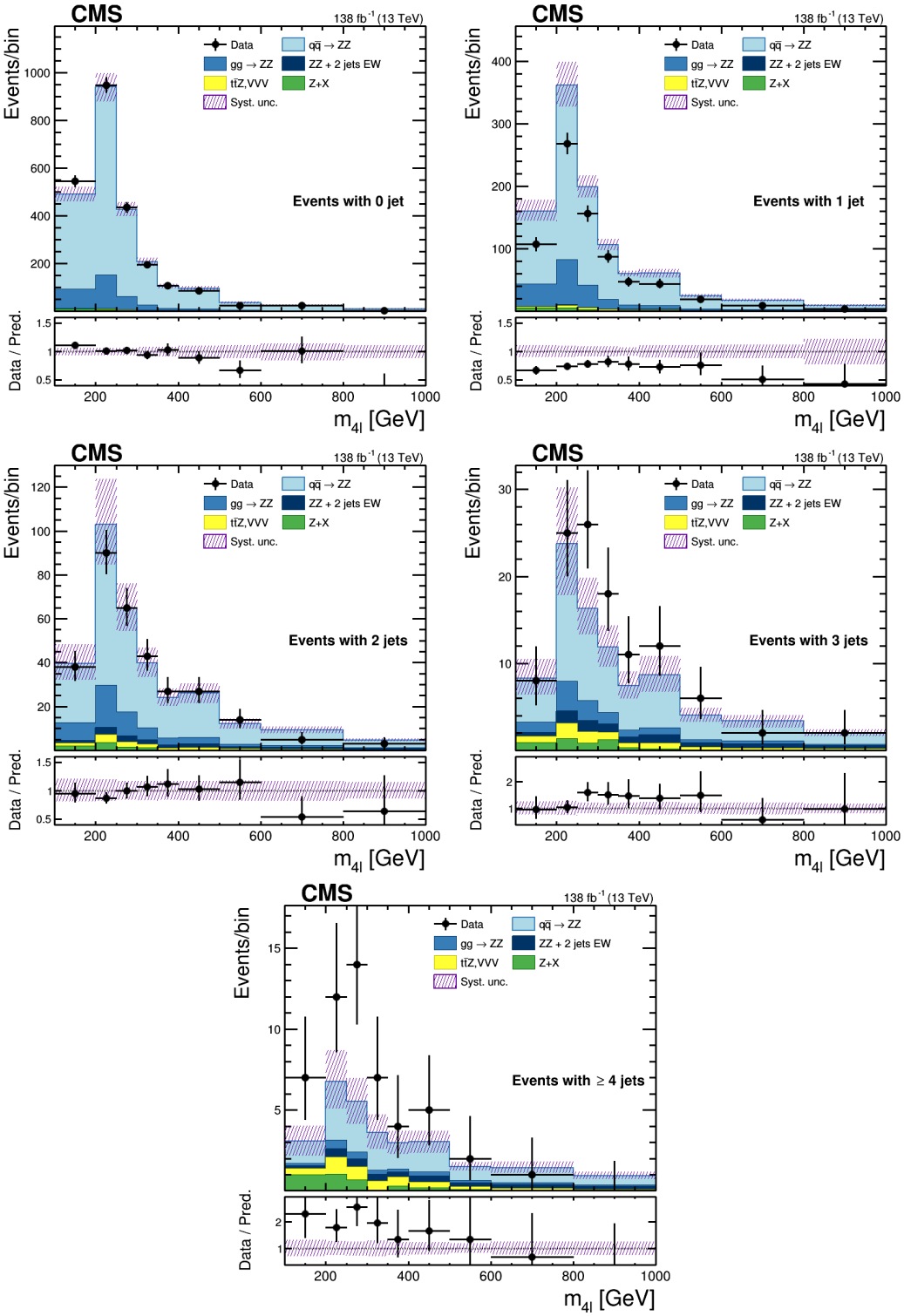

Figure 6.4: The $m_{4\ell}$ distributions for events with $60 < m_{\mathrm{Z}_1,\mathrm{Z}_2} < 120\,\mathrm{GeV}$ and different jet multiplicities. Other details are as in the caption of Fig. 6.1.

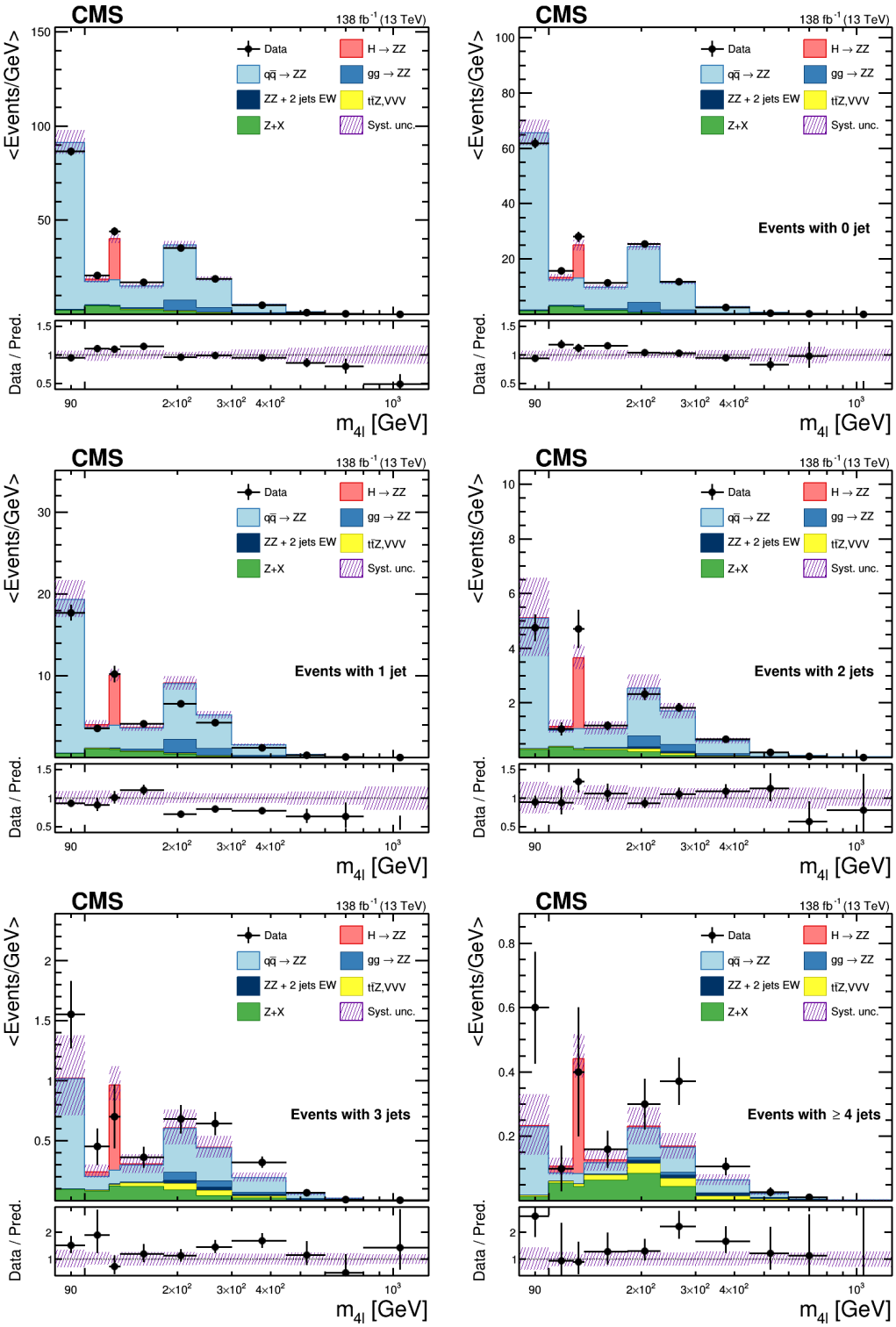

Figure 6.5: The $m_{4\ell}$ distributions in the full four-lepton invariant mass range for events with different jet multiplicities, normalized by bin width. Other details are as in the caption of Fig. 6.1.

Table 6.1: The observed and expected yields of ZZ events in different mass ranges, and estimated yields of background events, shown for each final state and for the sum. The first uncertainty is statistical, and the second one is systematic. (Due to rounding, the sum of individual entries may not match the total value shown.)

Process	eeee	ee$\mu\mu$	$\mu\mu\mu\mu$	$2\ell 2\ell'$
		$80 < m_{4\ell} < 100\,\text{GeV}$		
Background	$4.6 \pm 0.5 \pm 1.8$	$15.5 \pm 1.6 \pm 6.2$	$22.8 \pm 2.1 \pm 9.1$	$43 \pm 3 \pm 17$
Signal	$216 \pm 1^{+40}_{-36}$	$731 \pm 2^{+66}_{-64}$	$841 \pm 2^{+59}_{-57}$	$1790 \pm 3^{+140}_{-140}$
Total expected	$220 \pm 1^{+40}_{-36}$	$747 \pm 3^{+66}_{-64}$	$864 \pm 3^{+59}_{-58}$	$1830 \pm 4^{+140}_{-140}$
Data	194	698	838	1730
		$60 < m_{\text{Z}_1,\text{Z}_2} < 120\,\text{GeV}$		
Background	$22.9 \pm 0.9 \pm 5.7$	$46 \pm 2 \pm 10$	$28.9 \pm 1.3 \pm 6.5$	$98 \pm 2 \pm 23$
Signal	$716 \pm 2^{+63}_{-60}$	$1830 \pm 3^{+140}_{-140}$	$1138 \pm 3^{+85}_{-82}$	$3680 \pm 5^{+280}_{-270}$
Total expected	$739 \pm 2^{+63}_{-60}$	$1870 \pm 4^{+140}_{-140}$	$1167 \pm 3^{+85}_{-82}$	$3780 \pm 5^{+280}_{-270}$
Data	671	1805	1106	3582

Table 6.2: The observed and expected yields of ZZ events in different mass ranges, and estimated yields of background events, shown for each jet multiplicity. The first uncertainty is statistical, and the second one is systematic. (Due to rounding, the sum of individual entries may not match the total value shown.)

Process	0 jet	1 jet	2 jets	3 jets	$\geq$4 jets
			$80 < m_{4\ell} < 100\,\text{GeV}$		
Background	$25 \pm 2 \pm 10$	$9.1 \pm 1.3 \pm 3.6$	$6.1 \pm 1.0 \pm 2.4$	$1.9 \pm 0.6 \pm 0.8$	$0.4 \pm 0.3 \pm 0.1$
Signal	$1300 \pm 3^{+100}_{-100}$	$371 \pm 2^{+48}_{-45}$	$95 \pm 1^{+29}_{-28}$	$18.7 \pm 0.4^{+7.1}_{-6.2}$	$4.5 \pm 0.2^{+1.9}_{-1.8}$
Total expected	$1320 \pm 3^{+100}_{-100}$	$381 \pm 2^{+48}_{-45}$	$101 \pm 1^{+29}_{-28}$	$20.6 \pm 0.7^{+7.1}_{-6.2}$	$4.9 \pm 0.3^{+2.0}_{-1.8}$
Data	1238	354	95	31	12
			$60 < m_{\text{Z}_1,\text{Z}_2} < 120\,\text{GeV}$		
Background	$29.3 \pm 1.4 \pm 8.9$	$28.6 \pm 1.2 \pm 6.7$	$21.2 \pm 0.9 \pm 3.7$	$11.6 \pm 0.7 \pm 2.0$	$7.6 \pm 0.5 \pm 1.5$
Signal	$2320 \pm 3^{+160}_{-170}$	$960 \pm 3^{+100}_{-90}$	$303 \pm 1^{+60}_{-56}$	$75 \pm 1^{+20}_{-19}$	$21.9 \pm 0.3^{+7.9}_{-7.2}$
Total expected	$2350 \pm 4^{+160}_{-170}$	$990 \pm 3^{+100}_{-100}$	$324 \pm 2^{+60}_{-56}$	$87 \pm 1^{+21}_{-19}$	$29.5 \pm 0.7^{+8.1}_{-7.4}$
Data	2367	741	312	110	52

6.2 ZZ +jets Differential cross sections

The unfolded differential distributions normalized to the ZZ fiducial cross section are presented in Figs. 6.6–6.9. Figure 6.6 (left) shows the normalized $\text{d}\sigma/\text{d}m_{4\ell}$ cross section. The MadGraph5_aMC@NLO, powheg and nNNLO+PS predictions demonstrate similar behavior and describe well the differential behavior at low $m_{4\ell}$, whereas they overestimate the measured values in the moderate to high $m_{4\ell}$ regions. This discrepancy can be mitigated with EW corrections as discussed in Ref. [5]. To estimate

the effect of the corrections, a differential K-factor from [5] for the NLO EW corrections was applied to the nNNLO+PS predictions as a function of $m_{4\ell}$. The EW-corrected nNNLO+PS predictions describe the measured values better than those without the corrections, although at high values of $m_{4\ell}$ only within large statistical uncertainties of the measurements. The EW corrections are only significant in $m_{4\ell}$ and have negligible effect on any other normalized distribution presented in this paper; therefore in all other non-$m_{4\ell}$ distributions only nominal nNNLO+PS predictions are shown. For the $m_{4\ell}$ distributions for various jet multiplicities the EW corrections are not available and therefore these distributions do not contain EW corrections.

Figure 6.6 (right) shows the differential cross sections as a function of the number of jets in the events. The MadGraph5_aMC@NLO and POWHEG predictions show similar distributions. Similar to the discussion in the previous section, neither of the two MC simulations describes the 1-jet cross section, and both simulations predict too low cross sections for high jet multiplicities. On the other hand, the nNNLO+PS prediction describes the high jet multiplicity bin better than the other two predictions, whereas the agreement in the 1-jet bin is also improved. In general, the nNNLO+PS prediction describes the N_{jets} distribution better than MadGraph5_aMC@NLO and POWHEG.

Figure 6.7 shows the differential cross sections in bins of p_T and $|\eta|$ for the highest- and the second highest-p_T jet in events with at least one and two jets. The p_T distributions show moderate differences between data and predictions, whereas the $|\eta|$ distributions are well described within uncertainties.

Figure 6.8 shows the differential cross sections for dijet events as a function of (left) $|\Delta\eta|$ and (right) the dijet mass between two highest-p_T jets. Within uncertainties the dijet mass is described by the predictions, whereas $|\Delta\eta|$ measurements show a small trend with respect to the predictions.

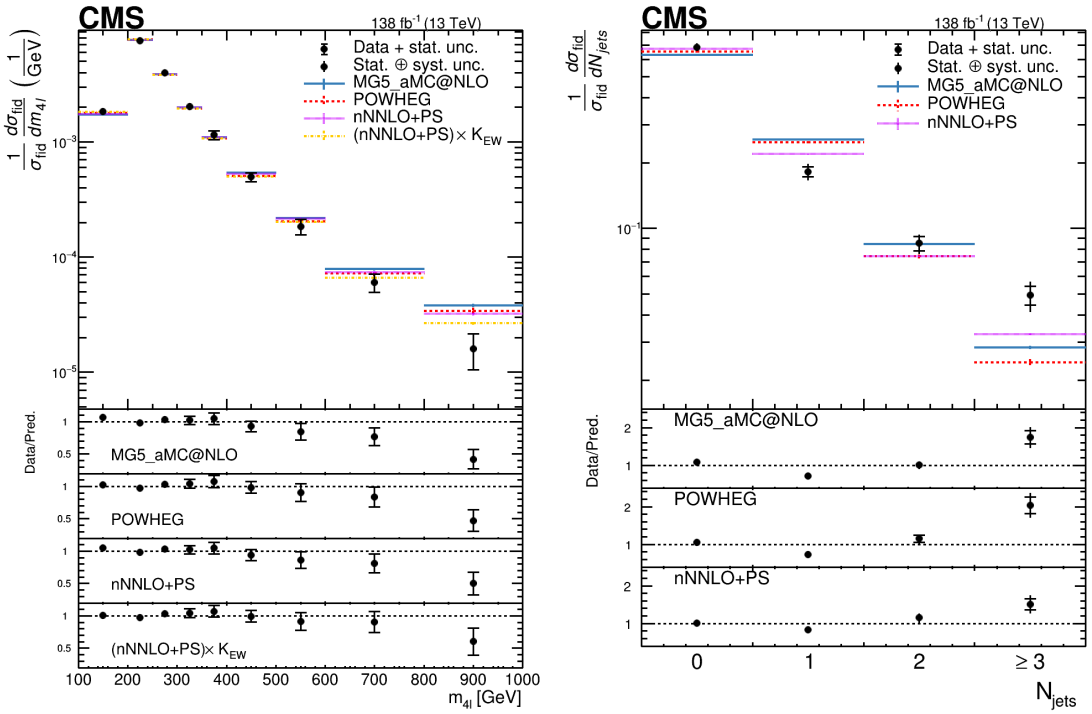

Figure 6.6: Differential cross sections normalized to the fiducial cross section as a function of (left) $m_{4\ell}$, (right) the number of jets with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$. The on-shell Z requirement $60 < m_{\mathrm{Z}_1,\mathrm{Z}_2} < 120\,\mathrm{GeV}$ is applied. Points represent the unfolded data, solid histograms the MadGraph5_aMC@NLO $\mathrm{q\bar{q}} \to \mathrm{ZZ}$ predictions, and red dashed histograms the POWHEG $\mathrm{q\bar{q}} \to \mathrm{ZZ}$ predictions. MCFM $\mathrm{gg} \to \mathrm{ZZ}$, POWHEG $\mathrm{H} \to \mathrm{ZZ}$, and MadGraph5_aMC@NLO EW ZZ predictions are included in these two sets of predictions. The purple dashed histograms represent the nNNLO+PS predictions, and the yellow dashed histogram represents the nNNLO+PS prediction with EW corrections applied. Vertical bars on both MC predictions represent the statistical uncertainties. The lower panels show the ratio of the measured to the predicted cross sections. The vertical bars on data points with horizontal lines at the ends represent the statistical uncertainties only, whereas the vertical bars without horizontal lines at the ends represent the total uncertainties calculated as the sum in quadrature of the statistical and systematic uncertainties. The overflow is included in the last bin of the distributions.

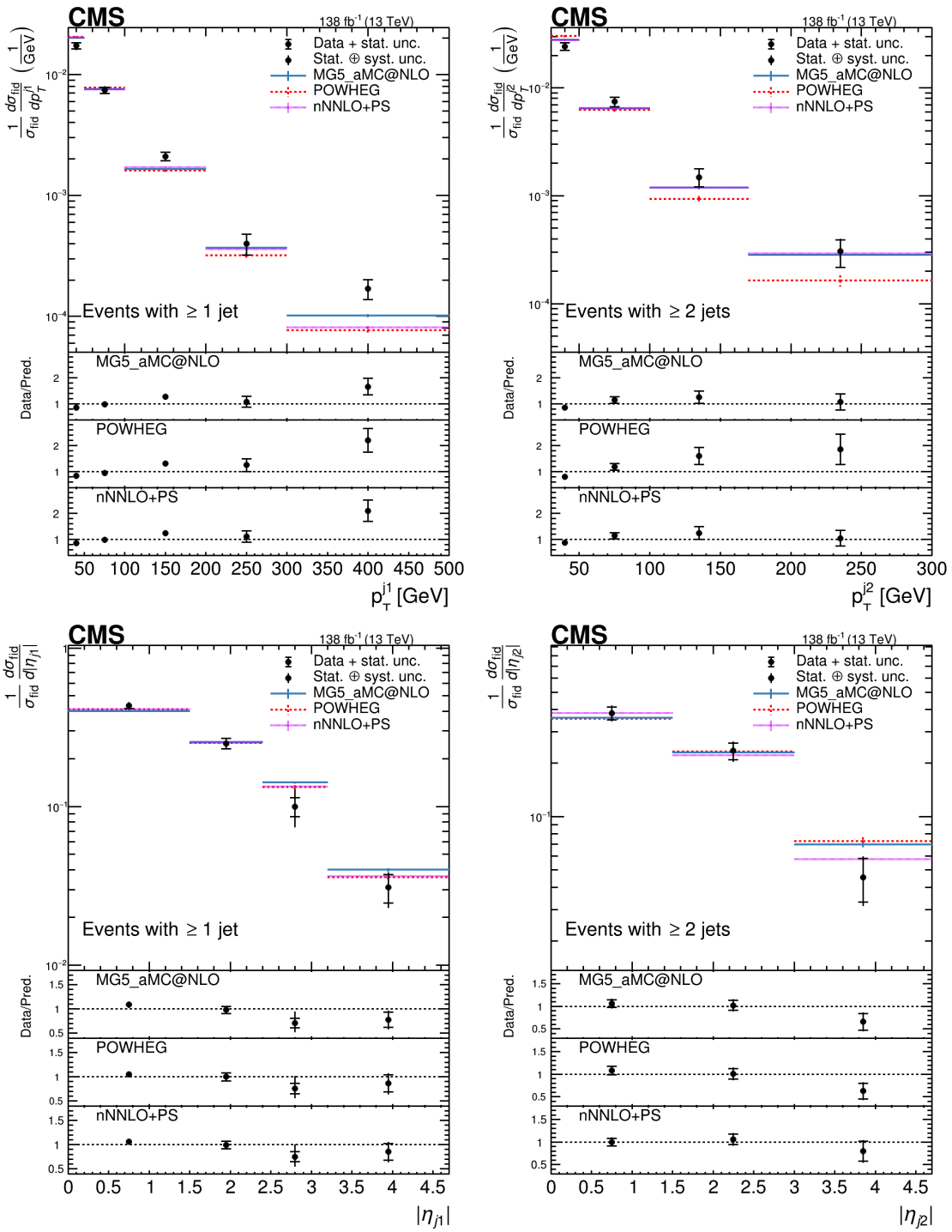

Figure 6.7: Differential cross sections normalized to the fiducial cross section as a function of p_{T} and $|\eta|$ of the highest- and the second-highest-p_{T} jet in events containing at least one or two jets, respectively. The on-shell Z requirement $60 < m_{\mathrm{Z}_1,\mathrm{Z}_2} < 120\,\mathrm{GeV}$ is applied. Other details are as in the caption of Fig. 6.6.

Finally, the $\mathrm{d}\sigma/\mathrm{d}m_{4\ell}$ differential cross section is presented in Fig. 6.9 for the full four-lepton invariant mass range and inclusive in jet multiplicity. The measured normalized differential cross section is well described by the predictions. Additional $\mathrm{d}\sigma/\mathrm{d}m_{4\ell}$ differential cross sections with different jet multiplicities for the on-shell Z bosons and for the full four-lepton invariant mass range are presented in Figs. 6.10 and 6.11. The comparison with the theoretical predictions shows the same behavior than for the measurements presented in the previous section.

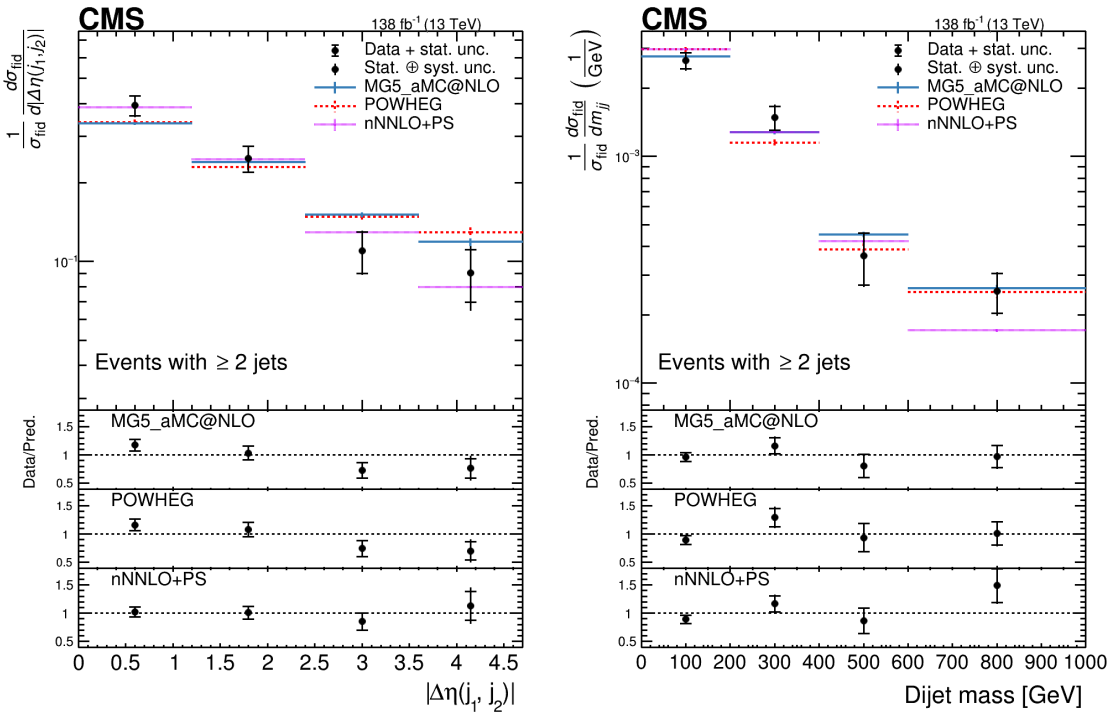

Figure 6.8: Differential cross sections normalized to the fiducial cross section as a function of (left) $|\Delta\eta|$ and (right) dijet mass between highest-p_{T} jets in events with at least two jets. Events with $60 < m_{\mathrm{Z}_1,\mathrm{Z}_2} < 120\,\mathrm{GeV}$ requirement. Other details are as in the caption of Fig. 6.6.

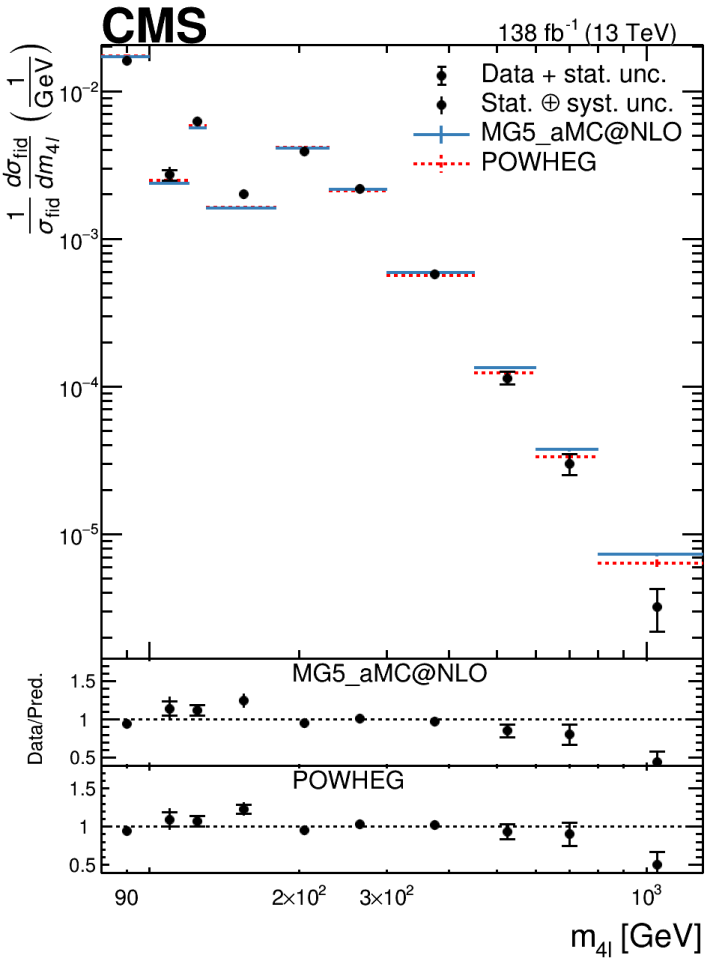

Figure 6.9: Differential cross sections normalized to the fiducial cross section as a function of $m_{4\ell}$ for the full four-lepton invariant mass range. Other details are as in the caption of Fig. 6.6.

6.3 Vector Boson Scattering

The measured signal strength in the EW ZZ+2jets fiducial region is $\mu_{EW} = 1.45^{+0.51}_{-0.44}$ (68% CL). To estimate the significance, we use asymptotic approximation [185] of the test statistics distribution with background-only hypothesis ($\mu_{EW} = 0$), and calculate the p value with COMBINE. The background-only hypothesis is excluded with a significance of 4.4 (3.4 expected) standard deviations. It is observed that the significance can vary by about 0.4 due to choice of random seeds used in `Keras` when

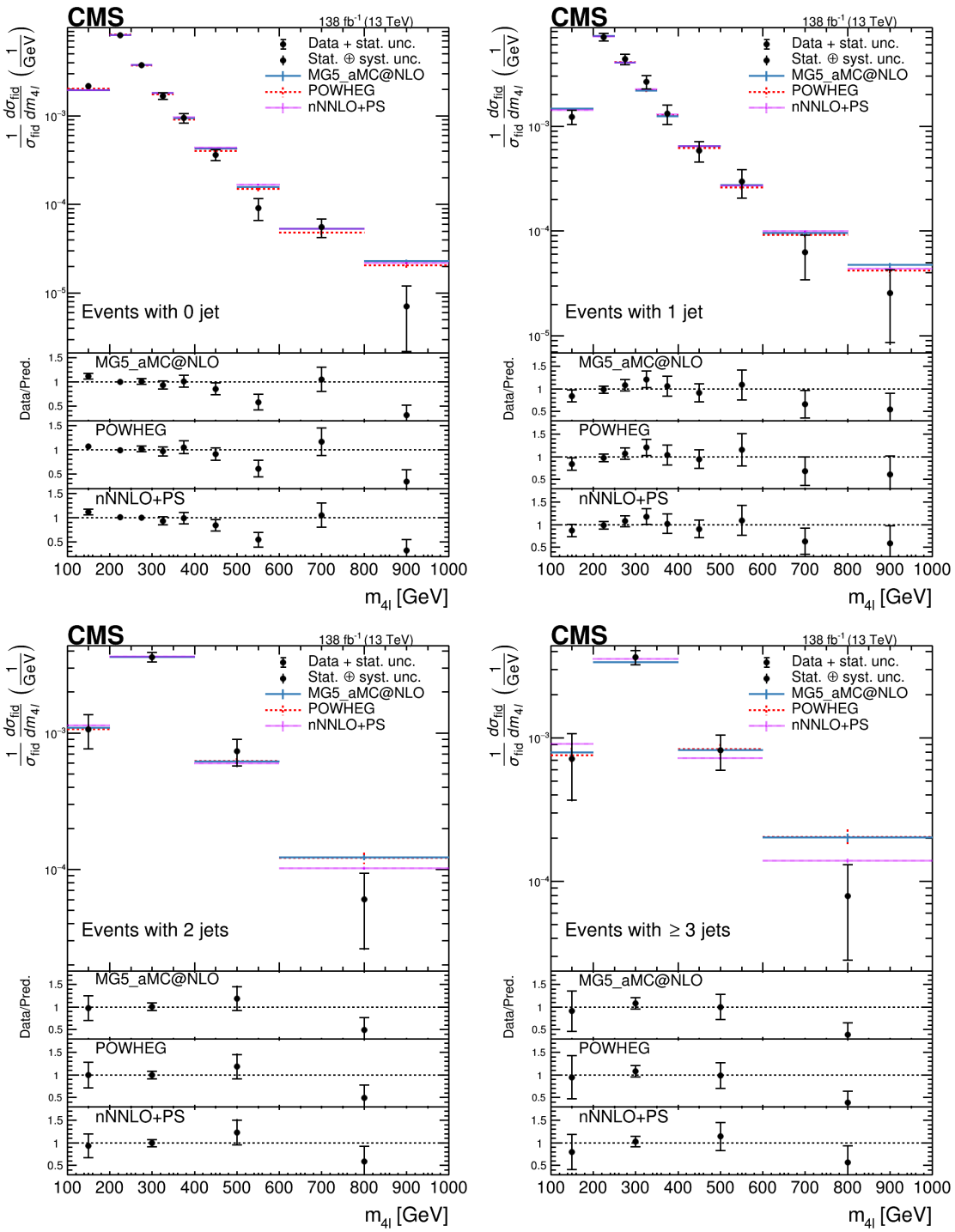

Figure 6.10: Differential cross sections normalized to the fiducial cross section as a function of $m_{4\ell}$ for $60 < m_{Z_1,Z_2} < 120\,\text{GeV}$ and for different jet multiplicities. Other details are as in the Fig. 6.6 caption.

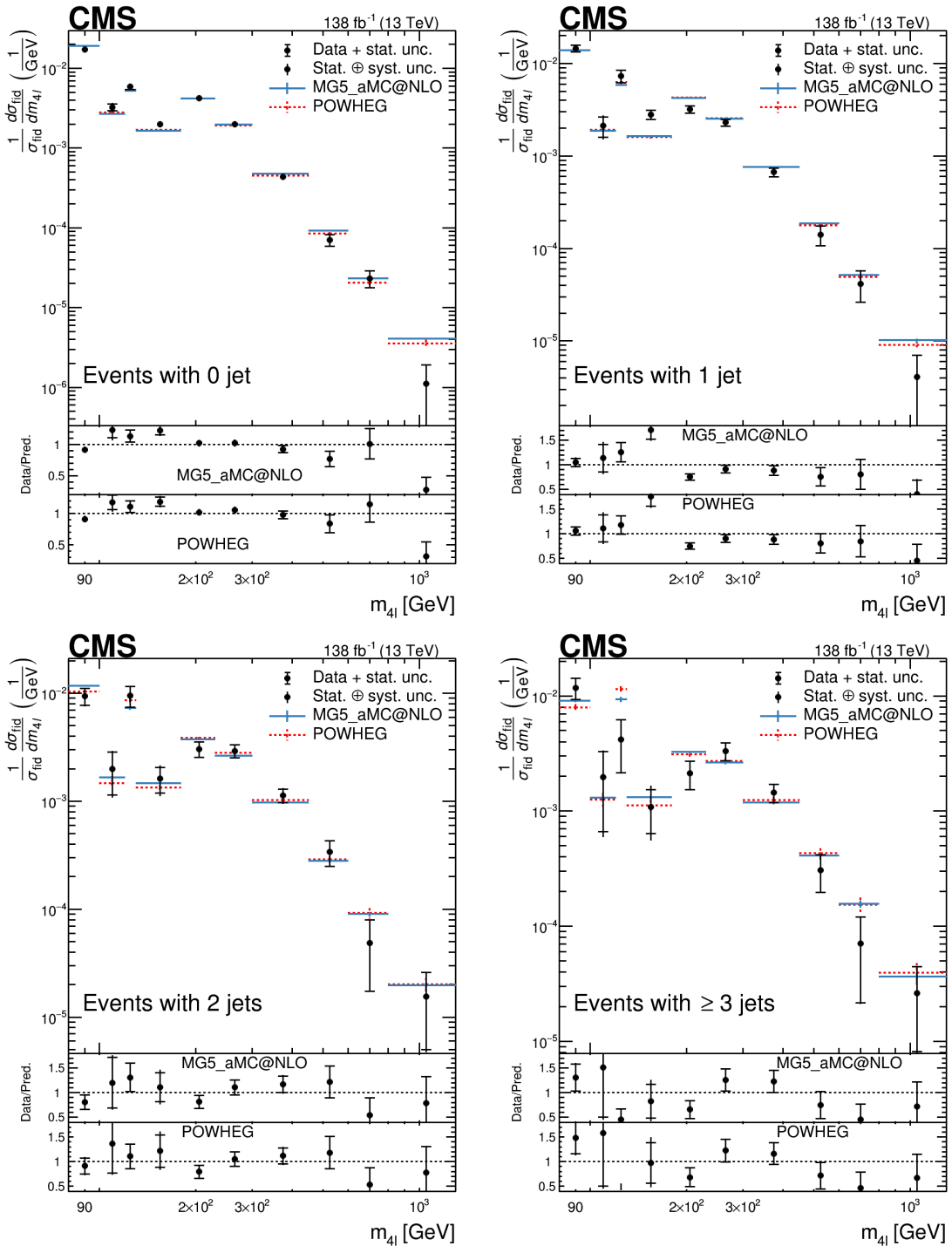

Figure 6.11: Differential cross sections normalized to the fiducial cross section as a function of $m_{4\ell}$ for the full four-lepton invariant mass range and different jet multiplicities. Other details are as in the Fig. 6.6 caption.

training the DNN. These results are consistent with the CMS results [4] obtained from using a discriminant based on a matrix element likelihood approach (MELA). The results may be improved by using more sophisticated machine learning approaches, optimizing the DNN training, and using larger experimental datasets and improved simulation.

Chapter 7

Conclusions

7.1 Summary

In the ZZ+jets analysis, the four-lepton production in association with jets, pp $\rightarrow$ $(\mathrm{Z}/\gamma^*)(\mathrm{Z}/\gamma^*) + \mathrm{jets} \rightarrow 2\ell 2\ell' + \mathrm{jets}$, where $\ell, \ell' = \mathrm{e}$ or μ, was studied in proton-proton collisions at a center-of-mass energy of 13 TeV. The data sample corresponds to an integrated luminosity of $138\,\mathrm{fb}^{-1}$collected with the CMS detector at the LHC during 2016–2018. Differential distributions and differential cross sections normalized to the ZZ fiducial cross section were measured with respect to various kinematic variables: number of jets, jet transverse momentum (p_T) and pseudorapidity (η), invariant mass of the dijet system and η difference between the highest-p_T and second-highest-p_T jets, and invariant mass of the four leptons ($m_{4\ell}$) for different jet multiplicities. Tabulated results are provided in HEPData [186]. In general, predictions of theoretical models agree with the data, but in some regions significant discrepancies between predicted and measured values were observed. The recent nNNLO+PS prediction improves the data/prediction agreement in the 1-jet and high jet multiplicity regions, and describes the distribution of jet multiplicities better than NLO samples generated

with the event generators MadGraph5_aMC@NLO and POWHEG. The inclusion of electroweak corrections improves the description of the $m_{4\ell}$ distribution. These measurements demonstrate the necessity for better Monte Carlo modeling in events with complex multiboson final states and extra jets. Further improvement of the predictions is required to describe the ZZ+jets production in the whole phase space.

In the EW ZZ+2jets analysis, a deep neural network approach is used to extract the signal of EW diboson ZZ production associated with 2 jets, again with $2\ell 2\ell'$ in the final states using the same CMS dataset. The observed (expected) significance is 4.4 (3.4) and the signal strength is measured to be $\mu_{\mathrm{EW}} = 1.45^{+0.51}_{-0.44}$, which is consistent with the published CMS results in Ref. [4].

7.2 Prospects

In presenting this discussion we have used those references cited in this section. As discussed in Ref. [32], the upgraded High-Luminosity LHC machine (HL-LHC), currently scheduled to start running in 2030, is planned to deliver an integrated luminosity of $3000\,\mathrm{fb}^{-1}$at a collision energy of $\sqrt{s} = 14\,\mathrm{TeV}$. In addition, the CMS detector, with its trigger and readout, will be substantially upgraded for HL-LHC running, resulting in important improvements in performance. The larger data set will improve the cross section measurement of the ZZ+jets production, which is currently statistically limited. Therefore, the larger data set will allow more detailed studies of the backgrounds associated with the EW ZZ and Higgs boson production, and allow tighter selection to reduce them, increasing the precision of the signal process measurements.

On the other hand, the VBS ZZ production processes, which remains unobserved as in Run 2 CMS results, is expected to be observed during LHC Run 3 [32], but dur-

ing the HL-LHC period the cross section of some VBS final states will be measured with a precision similar to that of current measurements of diboson final states (projected measurement uncertainty for EW ZZ VBS ranges from 20–100% by ATLAS; see Ref. [187]). An interesting prospect for the full HL-LHC data set is the measurement of longitudinal VBS, where both vector bosons are longitudinally polarized ($V_L V_L$). This is a crucial process in establishing the mathematical consistency of the SM, because in the SM, this process is unitarized due to the presence of Higgs boson contributions, and therefore deviations from this would indicate the presence of BSM physics. CMS projects the sensitivity of $\mathrm{Z}_L\mathrm{Z}_L$ to be 1.4σ with 3000 fb^{-1} [188] based on published analysis [183]. With the advance and optimization of analysis strategies, including improvements in the machine learning techniques, and with the extended muon coverage from the muon system upgrade, and improved electron identification performance from the CMS High-Granularity Calorimeter (HGCAL) [189], we expect further improvement of the sensitivity of these studies.

Bibliography

[1] CMS Collaboration, "Measurement of differential ZZ + jets production cross sections in pp collisions at $\sqrt{s} = 13$ TeV," *JHEP*, vol. 10, p. 209, 2024. DOI: 10.1007/JHEP10(2024)209. arXiv: 2404.02711 [hep-ex].

[2] CMS Collaboration, "Measurements of pp $\rightarrow$ ZZ production cross sections and constraints on anomalous triple gauge couplings at $\sqrt{s} = 13\,\text{TeV}$," *Eur. Phys. J. C*, vol. 81, p. 200, 2021. DOI: 10.1140/epjc/s10052-020-08817-8. arXiv: 2009.01186 [hep-ex].

[3] CMS Collaboration, "Measurement of differential cross sections for Z boson pair production in association with jets at $\sqrt{s} = 8$ and 13 TeV," *Phys. Lett. B*, vol. 789, p. 19, 2019. DOI: 10.1016/j.physletb.2018.11.007. arXiv: 1806.11073 [hep-ex].

[4] CMS Collaboration, "Evidence for electroweak production of four charged leptons and two jets in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$," *Phys. Lett. B*, vol. 812, p. 135 992, 2020. DOI: 10.1016/j.physletb.2020.135992. arXiv: 2008.07013 [hep-ex].

[5] L. Buonocore, G. Koole, D. Lombardi, L. Rottoli, M. Wiesemann, and G. Zanderighi, "ZZ production at nNNLO+PS with MiNNLO$_{PS}$," *JHEP*, vol. 01, p. 072, 2022. DOI: 10.1007/JHEP01(2022)072. arXiv: 2108.05337 [hep-ph].

[6] R. V. Harlander, S. Y. Klein, and M. Lipp, "FeynGame," *Comput. Phys. Commun.*, vol. 256, p. 107 465, 2020. DOI: 10.1016/j.cpc.2020.107465. arXiv: 2003.00896 [physics.ed-ph].

[7] R. Harlander, S. Y. Klein, and M. C. Schaaf, "FeynGame-2.1 – Feynman diagrams made easy," *PoS*, vol. EPS-HEP2023, p. 657, 2024. DOI: 10.22323/1.449.0657. arXiv: 2401.12778 [hep-ph].

[8] L. Bündgen, R. V. Harlander, S. Y. Klein, and M. C. Schaaf, "FeynGame 3.0," *Comput. Phys. Commun.*, vol. 314, p. 109 662, 2025. DOI: 10.1016/j.cpc.2025.109662. arXiv: 2501.04651 [hep-ph].

[9] M. Thomson, *Modern Particle Physics*. Cambridge University Press, 2013.

[10] G. Kane, *Modern Elementary Particle Physics: Explaining and Extending the Standard Model*, 2nd ed. Cambridge University Press, 2017.

[11] M. D. Schwartz, *Quantum Field Theory and the Standard Model*. Cambridge University Press, 2013.

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

[13] D. Goldberg, *The Standard Model in a Nutshell*. Princeton University Press, 2017.

[14] T. Lancaster and S. Blundell, *Quantum Field Theory for the Gifted Amateur*. OUP Oxford, 2014.

[15] D. Griffiths, *Introduction to Elementary Particles*, 2nd, Revised Edition. Wiley-VCH, 2008.

[16] M. E. Peskin and D. V. Schroeder, *An Introduction to quantum field theory*. CRC Press, 2018. DOI: 10.1201/9780429503559.

[17] B. Grzadkowski, M. Iskrzynski, M. Misiak, and J. Rosiek, "Dimension-Six Terms in the Standard Model Lagrangian," *JHEP*, vol. 10, p. 085, 2010. DOI: 10.1007/JHEP10(2010)085. arXiv: 1008.4884 [hep-ph].

[18] Particle Data Group Collaboration, "Review of particle physics," *Phys. Rev. D*, vol. 110, p. 030 001, 2024. DOI: 10.1103/PhysRevD.110.030001.

[19] R. K. Ellis, W. J. Stirling, and B. R. Webber, *QCD and Collider Physics* (Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology). Cambridge University Press, 1996.

[20] J. M. Campbell, J. W. Huston, and W. J. Stirling, "Hard Interactions of Quarks and Gluons: A Primer for LHC Physics," *Rept. Prog. Phys.*, vol. 70, p. 89, 2007. DOI: 10.1088/0034-4885/70/1/R02. arXiv: hep-ph/0611148.

[21] J. C. Collins, D. E. Soper, and G. F. Sterman, "Factorization of Hard Processes in QCD," *Adv. Ser. Direct. High Energy Phys.*, vol. 5, p. 1, 1989. DOI: 10.1142/9789814503266_0001. arXiv: hep-ph/0409313.

[22] Bryan Webber, *QCD Phenomenology at High Energy*, CERN Academic Training Lectures 2008. URL: https://indico.cern.ch/event/24747/, Feb. 2008.

[23] Gavin Salam, *QCD (for LHC)*, Presented at the 2009 European School of High-Energy Physics, in Bautzen, Germany. URL: https://gsalam.web.cern.ch/teaching/PhD-courses.html, Jun. 2009.

[24] Massimiliano Grazzini, *QCD and Monte Carlo tools*, Presented at the 2019 CERN-Fermilab HCP Summer School, held at CERN. URL: https://indico.cern.ch/event/795313/, Aug. 2019.

[25] Peter Skands, *QCD and Event Generators*, Presented at the 2020 Fermilab-CERN Hadron Collider Physics Summer School, held at Fermilab. URL: https://indico.fnal.gov/event/43762/, Aug. 2020.

[26] Rikkert Frederix, *Monte Carlo Generators*, Presented at the 2021 CERN-Fermilab HCP Summer School, held at CERN. URL: https://indico.cern.ch/event/1023573/, Aug. 2021.

[27] M. L. Mangano, "Introduction to QCD," 2000. DOI: 10.5170/CERN-1999-004.53. https://cds.cern.ch/record/454171.

[28] M. Mangano, "QCD and the physics of Hadronic Collisions," *CERN Yellow Rep. School Proc.*, vol. 4, pp. 27–62, 2018. DOI: 10.23730/CYRSP-2018-004.27. https://cds.cern.ch/record/2674114.

[29] B. Z. Kopeliovich and A. H. Rezaeian, "Applied High Energy QCD," *Int. J. Mod. Phys. E*, vol. 18, p. 1629, 2009. DOI: 10.1142/S0218301309013737. arXiv: 0811.2024 [hep-ph].

[30] A. D. Martin, "Proton structure, Partons, QCD, DGLAP and beyond," *Acta Phys. Polon. B*, vol. 39, R. Fiore, A. Papa, and C. Royon, Eds., p. 2025, 2008. arXiv: 0802.0161 [hep-ph].

[31] S. Höche, "Introduction to Parton-Shower Event Generators," in *Theoretical Advanced Study Institute in Elementary Particle Physics: Journeys Through the Precision Frontier: Amplitudes for Colliders*, 2015, p. 235. DOI: 10.1142/9789814678766_0005. arXiv: 1411.4085 [hep-ph].

[32] CMS Collaboration, "Stairway to discovery: A report on the CMS programme of cross section measurements from millibarns to femtobarns," *Phys. Rept.*, vol. 1115, p. 3, 2025. DOI: 10.1016/j.physrep.2024.11.005. arXiv: 2405.18661 [hep-ex].

[33] J. C. Romao and J. P. Silva, "A resource for signs and Feynman diagrams of the Standard Model," *Int. J. Mod. Phys. A*, vol. 27, p. 1 230 025, 2012. DOI: 10.1142/S0217751X12300256. arXiv: 1209.6213 [hep-ph].

[34] F. Englert and R. Brout, "Broken Symmetry and the Mass of Gauge Vector Mesons," *Phys. Rev. Lett.*, vol. 13, J. C. Taylor, Ed., p. 321, 1964. DOI: 10.1103/PhysRevLett.13.321.

[35] P. W. Higgs, "Broken symmetries, massless particles and gauge fields," *Phys. Lett.*, vol. 12, p. 132, 1964. DOI: 10.1016/0031-9163(64)91136-9.

[36] P. W. Higgs, "Broken Symmetries and the Masses of Gauge Bosons," *Phys. Rev. Lett.*, vol. 13, J. C. Taylor, Ed., p. 508, 1964. DOI: 10.1103/PhysRevLett.13.508.

[37] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, "Global Conservation Laws and Massless Particles," *Phys. Rev. Lett.*, vol. 13, J. C. Taylor, Ed., p. 585, 1964. DOI: 10.1103/PhysRevLett.13.585.

[38] P. W. Higgs, "Spontaneous Symmetry Breakdown without Massless Bosons," *Phys. Rev.*, vol. 145, p. 1156, 1966. DOI: 10.1103/PhysRev.145.1156.

[39] T. W. B. Kibble, "Symmetry breaking in nonAbelian gauge theories," *Phys. Rev.*, vol. 155, J. C. Taylor, Ed., p. 1554, 1967. DOI: 10.1103/PhysRev.155.1554.

[40] CMS Collaboration, "Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC," *Phys. Lett. B*, vol. 716, p. 30, 2012. DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235 [hep-ex].

[41] CMS Collaboration, "Observation of a New Boson with Mass Near 125 GeV in pp Collisions at $\sqrt{s} = 7$ and 8 TeV," *JHEP*, vol. 06, p. 081, 2013. DOI: 10.1007/JHEP06(2013)081. arXiv: 1303.4571 [hep-ex].

[42] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC," *Phys. Lett. B*, vol. 716, p. 1, 2012. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex].

[43] G. Golub and C. Van Loan, *Matrix Computations*. Johns Hopkins University Press, 1996.

[44] E. Eichten, K. Gottfried, T. Kinoshita, J. B. Kogut, K. D. Lane, and T.-M. Yan, "Spectrum of Charmed Quark-Antiquark Bound States," *Phys. Rev. Lett.*, vol. 34, pp. 369–372, 1975, [Erratum: Phys.Rev.Lett. 36, 1276 (1976)]. DOI: 10.1103/PhysRevLett.34.369.

[45] G. S. Bali, "QCD forces and heavy quark bound states," *Phys. Rept.*, vol. 343, p. 1, 2001. DOI: 10.1016/S0370-1573(00)00079-X. arXiv: hep-ph/0001312.

[46] V. N. Gribov and L. N. Lipatov, "Deep inelastic ep scattering in perturbation theory," *Sov. J. Nucl. Phys.*, vol. 15, p. 438, 1972.

[47] L. N. Lipatov, "The parton model and perturbation theory," *Yad. Fiz.*, vol. 20, p. 181, 1974.

[48] 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*, vol. 46, p. 641, 1977.

[49] G. Altarelli and G. Parisi, "Asymptotic Freedom in Parton Language," *Nucl. Phys. B*, vol. 126, p. 298, 1977. DOI: 10.1016/0550-3213(77)90384-4.

[50] G. Curci, W. Furmanski, and R. Petronzio, "Evolution of Parton Densities Beyond Leading Order: The Nonsinglet Case," *Nucl. Phys. B*, vol. 175, p. 27, 1980. DOI: 10.1016/0550-3213(80)90003-6.

[51] W. Furmanski and R. Petronzio, "Singlet Parton Densities Beyond Leading Order," *Phys. Lett. B*, vol. 97, p. 437, 1980. DOI: 10.1016/0370-2693(80)90636-X.

[52] S. Moch, J. A. M. Vermaseren, and A. Vogt, "The Three loop splitting functions in QCD: The Nonsinglet case," *Nucl. Phys. B*, vol. 688, p. 101, 2004. DOI: 10.1016/j.nuclphysb.2004.03.030. arXiv: hep-ph/0403192.

[53] A. Vogt, S. Moch, and J. A. M. Vermaseren, "The Three-loop splitting functions in QCD: The Singlet case," *Nucl. Phys. B*, vol. 691, p. 129, 2004. DOI: 10.1016/j.nuclphysb.2004.04.024. arXiv: hep-ph/0404111.

[54] T. Kinoshita, "Mass singularities of Feynman amplitudes," *J. Math. Phys.*, vol. 3, p. 650, 1962. DOI: 10.1063/1.1724268.

[55] T. D. Lee and M. Nauenberg, "Degenerate Systems and Mass Singularities," *Phys. Rev.*, vol. 133, G. Feinberg, Ed., B1549, 1964. DOI: 10.1103/PhysRev.133.B1549.

[56] H1, ZEUS Collaboration, "Combined Measurement and QCD Analysis of the Inclusive $e^{\pm}p$ Scattering Cross Sections at HERA," *JHEP*, vol. 01, p. 109, 2010. DOI: 10.1007/JHEP01(2010)109. arXiv: 0911.0884 [hep-ex].

[57] H1, ZEUS Collaboration, "Combination of measurements of inclusive deep inelastic $e^{\pm}p$ scattering cross sections and QCD analysis of HERA data," *Eur. Phys. J. C*, vol. 75, no. 12, p. 580, 2015. DOI: 10.1140/epjc/s10052-015-3710-4. arXiv: 1506.06042 [hep-ex].

[58] NNPDF Collaboration, "The path to N^3LO parton distributions," *Eur. Phys. J. C*, vol. 84, no. 7, p. 659, 2024. DOI: 10.1140/epjc/s10052-024-12891-7. arXiv: 2402.18635 [hep-ph].

[59] S. Kallweit and M. Wiesemann, "ZZ production at the LHC: NNLO predictions for $2\ell 2\nu$ and 4ℓ signatures," *Phys. Lett. B*, vol. 786, p. 382, 2018. DOI: 10.1016/j.physletb.2018.10.016. arXiv: 1806.05941 [hep-ph].

[60] M. Schmelling, "QCD results from the study of hadronic Z decays," *Phys. Scripta*, vol. 51, p. 683, 1995. DOI: 10.1088/0031-8949/51/6/004.

[61] C. Bierlich *et al.*, "A comprehensive guide to the physics and usage of PYTHIA 8.3," *SciPost Phys. Codeb.*, vol. 2022, p. 8, 2022. DOI: 10.21468/SciPostPhysCodeb.8. arXiv: 2203.11601 [hep-ph].

[62] S. Gieseke, P. Stephens, and B. Webber, "New formalism for QCD parton showers," *JHEP*, vol. 12, p. 045, 2003. DOI: 10.1088/1126-6708/2003/12/045. arXiv: hep-ph/0310083.

[63] T. Sjöstrand *et al.*, "An introduction to PYTHIA 8.2," *Comput. Phys. Commun.*, vol. 191, p. 159, 2015. DOI: 10.1016/j.cpc.2015.01.024. arXiv: 1410.3012 [hep-ph].

[64] T. Sjostrand and P. Z. Skands, "Transverse-momentum-ordered showers and interleaved multiple interactions," *Eur. Phys. J. C*, vol. 39, p. 129, 2005. DOI: 10.1140/epjc/s2004-02084-y. arXiv: hep-ph/0408302.

[65] M. Bahr, S. Gieseke, and M. H. Seymour, "Underlying events in Herwig++," pp. 641–645, 2009. DOI: 10.3204/DESY-PROC-2009-02/30. arXiv: 0809.2669. https://cds.cern.ch/record/1127194.

[66] CMS Collaboration, "CMS PYTHIA8 colour reconnection tunes based on underlying-event data," *Eur. Phys. J. C*, vol. 83, no. 7, p. 587, 2023. DOI: 10.1140/epjc/s10052-023-11630-8. arXiv: 2205.02905 [hep-ex].

[67] CMS Collaboration, "Event generator tunes obtained from underlying event and multiparton scattering measurements," *Eur. Phys. J. C*, vol. 76, p. 155, 2016. DOI: 10.1140/epjc/s10052-016-3988-x. arXiv: 1512.00815 [hep-ex].

[68] CMS Collaboration, "Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements," *Eur. Phys. J. C*, vol. 80, p. 4, 2020. DOI: 10.1140/epjc/s10052-019-7499-4. arXiv: 1903.12179 [hep-ex].

[69] CMS Collaboration, "Combination of CMS searches for heavy vector boson resonances at $\sqrt{s} = 13$ TeV," CERN, Geneva, Tech. Rep., 2025. https://cds.cern.ch/record/2945027.

[70] F. Cascioli *et al.*, "ZZ production at hadron colliders in NNLO QCD," *Phys. Lett. B*, vol. 735, p. 311, 2014. DOI: 10.1016/j.physletb.2014.06.056. arXiv: 1405.2219 [hep-ph].

[71] ATLAS Collaboration, "Measurement of ZZ production in pp collisions at $\sqrt{s} = 7$ TeV and limits on anomalous ZZZ and ZZγ couplings with the ATLAS detector," *JHEP*, vol. 03, p. 128, 2013. DOI: 10.1007/JHEP03(2013)128. arXiv: 1211.6096 [hep-ex].

[72] CMS Collaboration, "Measurements of the pp $\to$ ZZ production cross section and the Z $\to 4\ell$ branching fraction, and constraints on anomalous triple gauge couplings at $\sqrt{s} = 13$ TeV," *Eur. Phys. J. C*, vol. 78, p. 165, 2018, [Erratum: DOI 10.1140/epjc/s10052-018-5769-1]. DOI: 10.1140/epjc/s10052-018-5567-9. arXiv: 1709.08601 [hep-ex].

[73] Qibin LIU, *$Z\gamma$ VBS measurements with ATLAS full Run2*, Presented at Multi-Boson Interactions 2022 held at Shanghai, China. URL: https://indico.global/event/12230/contributions/108431/, Aug. 2022.

[74] ATLAS Collaboration, "Vector boson scattering measurements in ATLAS," CERN, Geneva, Tech. Rep., 2023. https://cds.cern.ch/record/2872673.

[75] CMS Collaboration, "Measurement of the ZZ production cross section and search for anomalous couplings in $2\ell 2\ell'$ final states in pp collisions at $\sqrt{s} = 7$ TeV," *JHEP*, vol. 01, p. 063, 2013. DOI: 10.1007/JHEP01(2013)063. arXiv: 1211.4890 [hep-ex].

[76] CMS Collaboration, "Measurement of the pp → ZZ production cross section and constraints on anomalous triple gauge couplings in four-lepton final states at $\sqrt{s} = 8\,\text{TeV}$," *Phys. Lett. B*, vol. 740, p. 250, 2015, [Corrigendum: DOI 10.1016/j.physletb.2016.04.010]. DOI: 10.1016/j.physletb.2014.11.059. arXiv: 1406.0113 [hep-ex].

[77] CMS Collaboration, "Measurements of the ZZ production cross sections in the $2\ell 2\nu$ channel in proton-proton collisions at $\sqrt{s} = 7$ and 8 TeV and combined constraints on triple gauge couplings," *Eur. Phys. J. C*, vol. 75, p. 511, 2015. DOI: 10.1140/epjc/s10052-015-3706-0. arXiv: 1503.05467 [hep-ex].

[78] CMS Collaboration, "Measurement of the ZZ production cross section and $Z \to \ell^+\ell^-\ell'^+\ell'^-$ branching fraction in pp collisions at $\sqrt{s} = 13\,\text{TeV}$," *Phys. Lett. B*, vol. 763, p. 280, 2016. DOI: 10.1016/j.physletb.2016.10.054. arXiv: 1607.08834 [hep-ex].

[79] CMS Collaboration, "Measurements of the electroweak diboson production cross sections in proton-proton collisions at $\sqrt{s} = 5.02$ TeV using leptonic decays," *Phys. Rev. Lett.*, vol. 127, p. 191 801, 2021. DOI: 10.1103/PhysRevLett.127.191801. arXiv: 2107.01137 [hep-ex].

[80] ATLAS Collaboration, "Measurements of four-lepton production in pp collisions at $\sqrt{s} = 8\,\text{TeV}$ with the ATLAS detector," *Phys. Lett. B*, vol. 753, p. 552, 2016. DOI: 10.1016/j.physletb.2015.12.048. arXiv: 1509.07844 [hep-ex].

[81] ATLAS Collaboration, "Measurement of the ZZ production cross section in pp collisions at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector," *Phys. Rev. Lett.*, vol. 116, p. 101 801, 2016. DOI: 10.1103/PhysRevLett.116.101801. arXiv: 1512.05314 [hep-ex].

[82] ATLAS Collaboration, "Measurement of the four-lepton invariant mass spectrum in 13 TeV proton-proton collisions with the ATLAS detector," *JHEP*, vol. 04, p. 048, 2019. DOI: 10.1007/JHEP04(2019)048. arXiv: 1902.05892 [hep-ex].

[83] ATLAS Collaboration, "Measurement of ZZ production in the $\ell\ell\nu\nu$ final state with the ATLAS detector in pp collisions at $\sqrt{s} = 13$ TeV," *JHEP*, vol. 10, p. 127, 2019. DOI: 10.1007/JHEP10(2019)127. arXiv: 1905.07163 [hep-ex].

[84] ATLAS Collaboration, "$ZZ \to \ell^+\ell^-\ell'^+\ell'^-$ cross-section measurements and search for anomalous triple gauge couplings in 13 TeV *pp* collisions with the ATLAS detector," *Phys. Rev. D*, vol. 97, p. 032 005, 2018. DOI: 10.1103/PhysRevD.97.032005. arXiv: 1709.07703 [hep-ex].

[85] ATLAS Collaboration, "Observation of electroweak production of two jets and a Z-boson pair," *Nature Phys.*, vol. 19, no. 2, p. 237, 2023. DOI: 10.1038/s41567-022-01757-y. arXiv: 2004.10612 [hep-ex].

[86] J. Alwall *et al.*, "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations," *JHEP*, vol. 07, p. 079, 2014. DOI: 10.1007/JHEP07(2014)079. arXiv: 1405.0301 [hep-ph].

[87] S. Alioli, P. Nason, C. Oleari, and E. Re, "A general framework for implementing NLO calculations in shower Monte Carlo programs: The POWHEG BOX," *JHEP*, vol. 06, p. 043, 2010. DOI: 10.1007/JHEP06(2010)043. arXiv: 1002.2581 [hep-ph].

[88] S. Alioli, P. Nason, C. Oleari, and E. Re, "NLO vector-boson production matched with shower in POWHEG," *JHEP*, vol. 07, p. 060, 2008. DOI: 10.1088/1126-6708/2008/07/060. arXiv: 0805.4802 [hep-ph].

[89] P. Nason, "A new method for combining NLO QCD with shower Monte Carlo algorithms," *JHEP*, vol. 11, p. 040, 2004. DOI: 10.1088/1126-6708/2004/11/040. arXiv: hep-ph/0409146 [hep-ph].

[90] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with parton shower simulations: The POWHEG method," *JHEP*, vol. 11, p. 070, 2007. DOI: 10.1088/1126-6708/2007/11/070. arXiv: 0709.2092 [hep-ph].

[91] S. Alioli, S. Ferrario Ravasio, J. M. Lindert, and R. Röntsch, "Four-lepton production in gluon fusion at NLO matched to parton showers," *Eur. Phys. J. C*, vol. 81, p. 687, 2021. DOI: 10.1140/epjc/s10052-021-09470-5. arXiv: 2102.07783 [hep-ph].

[92] L. Evans and P. Bryant, "LHC machine," *JINST*, vol. 3, S08001, 2008. DOI: 10.1088/1748-0221/3/08/S08001.

[93] O. Brüning, H. Burkhardt, and S. Myers, "The large hadron collider," *Progress in Particle and Nuclear Physics*, vol. 67, pp. 705–734, 2012. DOI: https://doi.org/10.1016/j.ppnp.2012.03.001.

[94] A. Lopes and M. L. Perrey, "FAQ-LHC The guide," 2022. https://cds.cern.ch/record/2809109.

[95] E. Mobs, "The CERN accelerator complex in 2019," 2019, General Photo. https://cds.cern.ch/record/2684277.

[96] H. Bartosik *et al.*, "Injectors Beam Performance Evolution during Run 2," 2019. https://cds.cern.ch/record/2750276.

[97] R. Steerenberg *et al.*, "Operation and performance of the CERN Large Hadron Collider during proton Run 2," in *10th International Particle Accelerator Conference*, 2019, MOPMP031. DOI: 10.18429/JACoW-IPAC2019-MOPMP031.

[98] W. Herr and B. Muratori, "Concept of luminosity," 2006. DOI: 10.5170/CERN-2006-002.361. https://cds.cern.ch/record/941318.

[99] CMS Collaboration, "The CMS experiment at the CERN LHC," *JINST*, vol. 3, S08004, 2008. DOI: 10.1088/1748-0221/3/08/S08004.

[100] CMS Collaboration, "Development of the CMS detector for the CERN LHC Run 3," *JINST*, vol. 19, P05064, 2024. DOI: 10.1088/1748-0221/19/05/P05064. arXiv: 2309.05466 [physics.ins-det].

[101] CMS Collaboration, "Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$," *JINST*, vol. 15, P10017, 2020. DOI: 10.1088/1748-0221/15/10/P10017. arXiv: 2006.10165 [hep-ex].

[102] CMS Collaboration, "The CMS trigger system," *JINST*, vol. 12, P01020, 2017. DOI: 10.1088/1748-0221/12/01/P01020. arXiv: 1609.02366 [physics.ins-det].

[103] CMS Collaboration, "Performance of the CMS high-level trigger during LHC Run 2," *JINST*, vol. 19, P11021, 2024. DOI: 10.1088/1748-0221/19/11/P11021. arXiv: 2410.17038 [physics.ins-det].

[104] CMS Collaboration, "Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC," *JINST*, vol. 16, P05014, 2021. DOI: 10.1088/1748-0221/16/05/P05014. arXiv: 2012.06888 [hep-ex].

[105] CMS Collaboration, "Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$," *JINST*, vol. 13, P06015, 2018. DOI: 10.1088/1748-0221/13/06/P06015. arXiv: 1804.04528 [physics.ins-det].

[106] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker," *JINST*, vol. 9, P10009, 2014. DOI: 10.1088/1748-0221/9/10/P10009. arXiv: 1405.6569 [physics.ins-det].

[107] G. Viehauser and T. Weidberg, *Detectors in Particle Physics: A Modern Introduction*. CRC Press, Taylor & Francis Group, 2024, ISBN: 9781032263182.

[108] T. Sakuma and T. McCauley, "Detector and Event Visualization with SketchUp at the CMS Experiment," *J. Phys. Conf. Ser.*, vol. 513, D. L. Groep and D. Bonacorsi, Eds., p. 022 032, 2014. DOI: 10.1088/1742-6596/513/2/022032. arXiv: 1311.4942 [physics.ins-det].

[109] CMS Tracker Group Collaboration, "The CMS Phase-1 pixel detector upgrade," *JINST*, vol. 16, P02027, 2021. DOI: 10.1088/1748-0221/16/02/P02027. arXiv: 2012.14304 [physics.ins-det].

[110] CMS Collaboration, "Track impact parameter resolution for the full pseudo rapidity coverage in the 2017 dataset with the CMS Phase-1 pixel detector," CMS Detector Performance Summary CMS-DP-2020-049, 2020. https://cds.cern.ch/record/2743740.

[111] CMS Collaboration, "2017 tracking performance plots," CMS Detector Performance Summary CMS-DP-2017-015, 2017. https://cds.cern.ch/record/2290524.

[112] CMS Collaboration, "Operation and performance of the CMS silicon strip tracker with proton-proton collisions at the CERN LHC," *JINST*, vol. 20, no. 08, P08027, 2025. DOI: 10.1088/1748-0221/20/08/P08027. arXiv: 2506.17195 [physics.ins-det].

[113] CMS Collaboration, "The Phase-2 Upgrade of the CMS Tracker," CERN, Geneva, Tech. Rep., 2017. DOI: 10.17181/CERN.QZ28.FLHW. https://cds.cern.ch/record/2272264.

[114] CMS Collaboration, "ECAL 2016 refined calibration and Run2 summary plots," CMS Detector Performance Summary CMS-DP-2020-021, 2020. https://cds.cern.ch/record/2717925.

[115] CMS Collaboration, "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV," *JINST*, vol. 12, P02014, 2017. DOI: 10.1088/1748-0221/12/02/P02014. arXiv: 1607.03663 [hep-ex].

[116] M. Abbas *et al.*, "Detector Control System for the GE1/1 slice test," *JINST*, vol. 15, no. 05, P05023, 2020. DOI: 10.1088/1748-0221/15/05/P05023.

[117] CMS Collaboration, "The GEM GE1/1 station of the CMS muon detector: status, commissioning and operation in magnetic field," CERN, Geneva, Tech. Rep., 2022. DOI: 10.22323/1.414.0632. https://cds.cern.ch/record/2842396.

[118] CMS Collaboration, "The Performance of the CMS Muon Detector in Proton-Proton Collisions at $\sqrt{s} = 7$ TeV at the LHC," *JINST*, vol. 8, P11002, 2013. DOI: 10.1088/1748-0221/8/11/P11002. arXiv: 1306.6905 [physics.ins-det].

[119] CMS Muon Collaboration, "Measurement of the background in the CMS muon detector in pp-collisions at $\sqrt{s} = 13$ TeV," *Eur. Phys. J. C*, vol. 84, no. 9, p. 955, 2024. DOI: 10.1140/epjc/s10052-024-13077-x.

[120] CMS Collaboration, "Triple GEM detectors for the Phase-2 upgrade of the CMS experiment at the LHC," *JINST*, vol. 20, no. 02, p. C02003, 2025. DOI: 10.1088/1748-0221/20/02/C02003. https://cds.cern.ch/record/2925538.

[121] CMS Collaboration, "CMS Technical Design Report for the Level-1 Trigger Upgrade," Tech. Rep., 2013. https://cds.cern.ch/record/1556311.

[122] J. Hegeman *et al.*, "The CMS Timing and Control Distribution System," in *2015 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, Oct. 2015, pp. 1–3. DOI: 10.1109/NSSMIC.2015.7581984.

[123] CMS Collaboration, "Precision luminosity measurement in proton-proton collisions at $\sqrt{s} = 13$ TeV in 2015 and 2016 at CMS," *Eur. Phys. J. C*, vol. 81, p. 800, 2021. DOI: 10.1140/epjc/s10052-021-09538-2. arXiv: 2104.01927 [hep-ex].

[124] CMS Collaboration, "CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13$ TeV," CMS Physics Analysis Summary CMS-PAS-LUM-18-002, 2019. https://cds.cern.ch/record/2676164/.

[125] P. Grafström and W. Kozanecki, "Luminosity determination at proton colliders," *Progress in Particle and Nuclear Physics*, vol. 81, pp. 97–148, 2015, ISSN: 0146-6410. DOI: 10.1016/j.ppnp.2014.11.002. https://www.sciencedirect.com/science/article/pii/S0146641014000878.

[126] "Addendum to the report on the physics at the HL-LHC, and perspectives for the HE-LHC: Collection of notes from ATLAS and CMS," *CERN Yellow Rep. Monogr.*, vol. 7, Addendum, 2019. DOI: 10.23731/CYRM-2019-007.Addendum. arXiv: 1902.10229 [hep-ex].

[127] CMS Collaboration, "Performance of the Pixel Luminosity Telescope for Luminosity Measurement at CMS during Run2," *PoS*, vol. EPS-HEP2017, P. Checchia *et al.*, Eds., p. 504, 2017. DOI: 10.22323/1.314.0504.

[128] M. Hempel, "Development of a Novel Diamond Based Detector for Machine Induced Background and Luminosity Measurements," Ph.D. dissertation, Brandenburg Tech. U., Hamburg, 2017. DOI: 10.3204/PUBDB-2017-06875.

[129] G. S. Millan, D. Perrin, and L. Scibile, "RAMSES: the LHC radiation monitoring system for the environment and safety," in *Proceedings of the 10th International Conference on Accelerator and Large Experimental Physics Control Systems (ICALEPCS 2005)*, 2005.

[130] A. Ledeul, A. Savulescu, G. Segura, B. Styczen, and D. Vazquez Rivera, "CERN Supervision, Control and Data Acquisition System for Radiation and Environmental Protection," in *12th International Workshop on Personal Computers and Particle Accelerator Controls*, 2019, FRCC3. DOI: 10.18429/JACoW-PCaPAC2018-FRCC3.

[131] S. van der Meer, "Calibration of the effective beam height in the ISR," CERN, Geneva, Tech. Rep., 1968. https://cds.cern.ch/record/296752.

[132] C. Rubbia, "Measurement of the luminosity of $p\bar{p}$ collider with a (generalized) van der meer method," CERN-$p\bar{p}$-Note-38, Tech. Rep., 1977.

[133] LHCb Collaboration, "Precision luminosity measurements at LHCb," *JINST*, vol. 9, no. 12, P12005, 2014. DOI: 10.1088/1748-0221/9/12/P12005. arXiv: 1410.0149 [hep-ex].

[134] ATLAS Collaboration, "Luminosity determination in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector at the LHC," *Eur. Phys. J. C*, vol. 76, no. 12, p. 653, 2016. DOI: 10.1140/epjc/s10052-016-4466-1. arXiv: 1608.03953 [hep-ex].

[135] CMS Collaboration, "CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13$ TeV," CMS Physics Analysis Summary CMS-PAS-LUM-17-004, 2018. https://cds.cern.ch/record/2621960/.

[136] Carlo Oleari, *MATCHING NLO CALCULATIONS WITH PARTON SHOWER: THE POSITIVE-WEIGHT HARDEST EMISSION GENERATOR*, URL: https://indico.in2p3.fr/event/6053/, Nov. 2011.

[137] Gionata Luisoni, *An introduction to POWHEG*, Presented at Dartmouth-UW Experimental/Theory Discussion.URL: https://indico.global/event/732/, Mar. 2017.

[138] S. Hoeche, F. Krauss, M. Schonherr, and F. Siegert, "A critical appraisal of NLO+PS matching methods," *JHEP*, vol. 09, p. 049, 2012. DOI: 10.1007/JHEP09(2012)049. arXiv: 1111.1220 [hep-ph].

[139] M. Bengtsson and T. Sjöstrand, "Coherent parton showers versus matrix elements - implications of PETRA/PEP data," *Physics Letters B*, vol. 185, p. 435, 1987. DOI: 10.1016/0370-2693(87)91031-8.

[140] A. Buckley *et al.*, "General-purpose event generators for LHC physics," *Phys. Rept.*, vol. 504, p. 145, 2011. DOI: 10.1016/j.physrep.2011.03.005. arXiv: 1101.2599 [hep-ph].

[141] S. Frixione and B. R. Webber, "Matching NLO QCD computations and parton shower simulations," *JHEP*, vol. 06, p. 029, 2002. DOI: 10.1088/1126-6708/2002/06/029. arXiv: hep-ph/0204244.

[142] S. Catani, F. Krauss, R. Kuhn, and B. R. Webber, "QCD matrix elements + parton showers," *JHEP*, vol. 11, p. 063, 2001. DOI: 10.1088/1126-6708/2001/11/063. arXiv: hep-ph/0109231.

[143] J. Alwall *et al.*, "Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions," *Eur. Phys. J. C*, vol. 53, pp. 473–500, 2008. DOI: 10.1140/epjc/s10052-007-0490-5. arXiv: 0706.2569 [hep-ph].

[144] H. Brooks and C. T. Preuss, "Efficient multi-jet merging with the VINCIA sector shower," *Comput. Phys. Commun.*, vol. 264, p. 107 985, 2021. DOI: 10.1016/j.cpc.2021.107985. arXiv: 2008.09468 [hep-ph].

[145] M. L. Mangano, M. Moretti, and R. Pittau, "Multijet matrix elements and shower evolution in hadronic collisions: $Wb\bar{b}$ + n jets as a case study," *Nucl. Phys. B*, vol. 632, p. 343, 2002. DOI: 10.1016/S0550-3213(02)00249-3. arXiv: hep-ph/0108069.

[146] R. Frederix and S. Frixione, "Merging meets matching in MC@NLO," *JHEP*, vol. 12, p. 061, 2012. DOI: 10.1007/JHEP12(2012)061. arXiv: 1209.6215 [hep-ph].

[147] L. Lönnblad and S. Prestel, "Merging Multi-leg NLO Matrix Elements with Parton Showers," *JHEP*, vol. 03, p. 166, 2013. DOI: 10.1007/JHEP03(2013)166. arXiv: 1211.7278 [hep-ph].

[148] K. Hamilton, P. Nason, and G. Zanderighi, "MINLO: Multi-Scale Improved NLO," *JHEP*, vol. 10, p. 155, 2012. DOI: 10.1007/JHEP10(2012)155. arXiv: 1206.3572 [hep-ph].

[149] S. Hoeche, F. Krauss, M. Schonherr, and F. Siegert, "QCD matrix elements + parton showers: The NLO case," *JHEP*, vol. 04, p. 027, 2013. DOI: 10.1007/JHEP04(2013)027. arXiv: 1207.5030 [hep-ph].

[150] P. F. Monni, P. Nason, E. Re, M. Wiesemann, and G. Zanderighi, "MiNNLO$_{PS}$: a new method to match NNLO QCD to parton showers," *JHEP*, vol. 05, p. 143, 2020, [Erratum: DOI 10.1007/JHEP02(2022)031]. DOI: 10.1007/JHEP05(2020)143. arXiv: 1908.06987 [hep-ph].

[151] K. Hamilton, P. Nason, C. Oleari, and G. Zanderighi, "Merging H/W/Z + 0 and 1 jet at NLO with no merging scale: a path to parton shower + NNLO matching," *JHEP*, vol. 05, p. 082, 2013. DOI: 10.1007/JHEP05(2013)082. arXiv: 1212.4504 [hep-ph].

[152] S. Höche, Y. Li, and S. Prestel, "Drell-Yan lepton pair production at NNLO QCD with parton showers," *Phys. Rev. D*, vol. 91, p. 074 015, 2015. DOI: 10.1103/PhysRevD.91.074015. arXiv: 1405.3607 [hep-ph].

[153] S. Alioli, C. W. Bauer, C. Berggren, F. J. Tackmann, J. R. Walsh, and S. Zuberi, "Matching Fully Differential NNLO Calculations and Parton Showers," *JHEP*, vol. 06, p. 089, 2014. DOI: 10.1007/JHEP06(2014)089. arXiv: 1311.0286 [hep-ph].

[154] P. F. Monni, E. Re, and M. Wiesemann, "MiNNLO$_{\rm PS}$: optimizing $2 \to 1$ hadronic processes," *Eur. Phys. J. C*, vol. 80, p. 1075, 2020. DOI: 10.1140/epjc/s10052-020-08658-5. arXiv: 2006.04133 [hep-ph].

[155] D. Lombardi, M. Wiesemann, and G. Zanderighi, "Advancing MiNNLO$_{PS}$ to diboson processes: Zγ production at NNLO+PS," *JHEP*, vol. 06, p. 095, 2021. DOI: 10.1007/JHEP06(2021)095. arXiv: 2010.10478 [hep-ph].

[156] D. Lombardi, M. Wiesemann, and G. Zanderighi, "W^+W^- production at NNLO+PS with MINNLO$_{PS}$," *JHEP*, vol. 11, p. 230, 2021. DOI: 10.1007/JHEP11(2021)230. arXiv: 2103.12077 [hep-ph].

[157] J. Mazzitelli, P. F. Monni, P. Nason, E. Re, M. Wiesemann, and G. Zanderighi, "Next-to-Next-to-Leading Order Event Generation for Top-Quark Pair Production," *Phys. Rev. Lett.*, vol. 127, no. 6, p. 062 001, 2021. DOI: 10.1103/PhysRevLett.127.062001. arXiv: 2012.14267 [hep-ph].

[158] C. Biello, A. Sankar, M. Wiesemann, and G. Zanderighi, "NNLO+PS predictions for Higgs production through bottom-quark annihilation with MiNNLO$_{PS}$," *Eur. Phys. J. C*, vol. 84, p. 479, 2024. DOI: 10.1140/epjc/s10052-024-12845-z. arXiv: 2402.04025 [hep-ph].

[159] GEANT4 Collaboration, "GEANT4—a simulation toolkit," *Nucl. Instrum. Meth. A*, vol. 506, p. 250, 2003. DOI: 10.1016/S0168-9002(03)01368-8.

[160] J. M. Campbell and R. K. Ellis, "MCFM for the Tevatron and the LHC," *Nucl. Phys. B Proc. Suppl.*, vol. 10, p. 205, 2010. DOI: 10.1016/j.nuclphysbps.2010.08.011. arXiv: 1007.3492 [hep-ph].

[161] F. Caola, K. Melnikov, R. Röntsch, and L. Tancredi, "QCD corrections to ZZ production in gluon fusion at the LHC," *Phys. Rev. D*, vol. 92, p. 094 028, 2015. DOI: 10.1103/PhysRevD.92.094028. arXiv: 1509.06734 [hep-ph].

[162] NNPDF Collaboration, "Parton distributions for the LHC run II," *JHEP*, vol. 04, p. 040, 2015. DOI: 10.1007/JHEP04(2015)040. arXiv: 1410.8849 [hep-ph].

[163] CMS Collaboration, "Particle-flow reconstruction and global event description with the CMS detector," *JINST*, vol. 12, P10003, 2017. DOI: 10.1088/1748-0221/12/10/P10003. arXiv: 1706.04965 [physics.ins-det].

[164] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm," *JHEP*, vol. 04, p. 063, 2008. DOI: 10.1088/1126-6708/2008/04/063. arXiv: 0802.1189 [hep-ex].

[165] M. Cacciari, G. P. Salam, and G. Soyez, "FastJet user manual," *Eur. Phys. J. C*, vol. 72, p. 1896, 2012. DOI: 10.1140/epjc/s10052-012-1896-2. arXiv: 1111.6097 [hep-ph].

[166] CMS Collaboration, "Technical proposal for the Phase-II upgrade of the Compact Muon Solenoid," CMS Technical Proposal CERN-LHCC-2015-010, CMS-TDR-15-02, 2015. http://cds.cern.ch/record/2020886.

[167] CMS Collaboration, "Measurements of production cross sections of the Higgs boson in the four-lepton final state in proton-proton collisions at $\sqrt{s} = 13$ TeV," *Eur. Phys. J. C*, vol. 81, p. 488, 2021. DOI: 10.1140/epjc/s10052-021-09200-x. arXiv: 2103.04956 [hep-ex].

[168] CMS Collaboration, "Muon identification using multivariate techniques in the CMS experiment in proton-proton collisions at $\sqrt{s} = 13$ TeV," *JINST*, vol. 19, no. 02, P02031, 2024. DOI: 10.1088/1748-0221/19/02/P02031. arXiv: 2310.03844 [hep-ex].

[169] CMS Collaboration, "Pileup mitigation at CMS in 13 TeV data," *JINST*, vol. 15, P09018, 2020. DOI: 10.1088/1748-0221/15/09/P09018. arXiv: 2003.00503 [hep-ex].

[170] M. Cacciari and G. P. Salam, "Pileup subtraction using jet areas," *Phys. Lett. B*, vol. 659, p. 119, 2008. DOI: 10.1016/j.physletb.2007.09.077. arXiv: 0707.1378 [hep-ph].

[171] CMS Collaboration, "Measurement of the inclusive W and Z production cross sections in pp collisions at $\sqrt{s} = 7$ TeV with the CMS experiment," *JHEP*, vol. 10, p. 132, 2011. DOI: 10.1007/JHEP10(2011)132. arXiv: 1107.4789 [hep-ex].

[172] CMS Collaboration, "Jet algorithms performance in 13 TeV data," CMS Physics Analysis Summary CMS-PAS-JME-16-003, 2017. https://cds.cern.ch/record/2256875.

[173] CMS Collaboration, "Determination of jet energy calibration and transverse momentum resolution in CMS," *JINST*, vol. 6, P11002, 2011. DOI: 10.1088/1748-0221/6/11/P11002. arXiv: 1107.4277.

[174] Particle Data Group, R. L. Workman, *et al.*, "Review of particle physics," *Prog. Theor. Exp. Phys.*, vol. 2022, p. 083C01, 2022. DOI: 10.1093/ptep/ptac097.

[175] CMS Collaboration, "Measurement of the properties of a Higgs boson in the four-lepton final state," *Phys. Rev. D*, vol. 89, p. 092 007, 2014. DOI: 10.1103/PhysRevD.89.092007. arXiv: 1312.5353 [hep-ex].

[176] G. D'Agostini, "A multidimensional unfolding method based on Bayes' theorem," *Nucl. Instrum. Meth. A*, vol. 362, p. 487, 1995. DOI: 10.1016/0168-9002(95)00274-X.

[177] T. Adye, "Unfolding algorithms and tests using RooUnfold," in *PHYSTAT 2011*, Geneva: CERN, 2011, p. 313. DOI: 10.5170/CERN-2011-006.313. arXiv: 1105.1160 [physics.data-an].

[178] CMS Collaboration, "Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 13$ TeV," *JHEP*, vol. 07, p. 161, 2018. DOI: 10.1007/JHEP07(2018)161. arXiv: 1802.02613 [hep-ex].

[179] J. Butterworth *et al.*, "PDF4LHC recommendations for LHC Run II," *J. Phys. G*, vol. 43, p. 023 001, 2016. DOI: 10.1088/0954-3899/43/2/023001. arXiv: 1510.03865 [hep-ph].

[180] F. Chollet, *Deep Learning with Python*. Manning Publications, 2018.

[181] F. Chollet *et al.*, *Keras*, https://keras.io, 2015.

[182] Martín Abadi *et al.*, *TensorFlow: Large-scale machine learning on heterogeneous systems*, Software available from tensorflow.org, 2015. https://www.tensorflow.org/.

[183] CMS Collaboration, "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*, vol. 774, p. 682, 2017. DOI: 10.1016/j.physletb.2017.10.020. arXiv: 1708.02812 [hep-ex].

[184] CMS Collaboration, "The CMS statistical analysis and combination tool: COMBINE," *Comput. Softw. Big Sci.*, vol. 8, p. 19, 2024. DOI: 10.1007/s41781-024-00121-4. arXiv: 2404.06614 [physics.data-an].

[185] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics," *Eur. Phys. J. C*, vol. 71, p. 1554, 2011, [Erratum: DOI 10.1140/epjc/s10052-013-2501-z]. DOI: 10.1140/epjc/s10052-011-1554-0. arXiv: 1007.1727 [physics.data-an].

[186] *HEPData record for the ZZ+jets analysis*, 2023. DOI: 10.17182/hepdata.145862.

[187] ATLAS Collaboration, "Snowmass White Paper Contribution: Physics with the Phase-2 ATLAS and CMS Detectors," CERN, Geneva, Tech. Rep., 2022. https://cds.cern.ch/record/2805993.

[188] CMS Collaboration, "Vector Boson Scattering prospective studies in the ZZ fully leptonic decay channel for the High-Luminosity and High-Energy LHC upgrades," CERN, Geneva, Tech. Rep., 2018. https://cds.cern.ch/record/2650915.

[189] C. Amendola and on behalf of the CMS Collaboration, "An overview of the CMS High-Granularity Calorimeter," *Journal of Instrumentation*, vol. 20, no. 01, p. C01031, Jan. 2025. DOI: 10.1088/1748-0221/20/01/C01031.