Jets produced in association with W-bosons in CMS at the LHC

by

Kira S. Grogg

A dissertation submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

(Physics)

at the

University of Wisconsin – Madison

2011

Jets produced in association with W-bosons in CMS at the LHC

Kira S. Grogg

Under the supervision of Prof. Wesley H. Smith

At the University of Wisconsin–Madison

Presented here is a study of jets produced in association with W bosons production in pp collisions at center-of-mass energy of 7 TeV using the full CMS 2010 data set, corresponding to a luminosity of 36 pb^{-1}. The transverse energy distribution of the reconstructed leading jets is measured and compared to theoretical expectations. The measured jet multiplicity distributions are efficiency corrected and unfolded to enable direct comparison to theoretical models. The ratios of multiplicities, $\sigma(W+n)/\sigma(W+(n-1))$ and $\sigma(W+n)/\sigma(W)$ where n stands for the number of jets produced, are also presented.

Abstract

Presented here is a study of jets produced in association with W bosons production in pp collisions at center-of-mass energy of 7 TeV using the full CMS 2010 data set, corresponding to a luminosity of 36 pb^{-1}. The transverse energy distribution of the reconstructed leading jets is measured and compared to theoretical expectations. The measured jet multiplicity distributions are efficiency corrected and unfolded to enable direct comparison to theoretical models. The ratios of multiplicities, $\sigma(W+n)/\sigma(W+(n-1))$ and $\sigma(W+n)/\sigma(W)$ where n stands for the number of jets produced, are also presented.

To my advisors for their knowledge, to my parents for their wisdom, to my fellow students for their ideas, and to my friends for their support.

"Let's think the unthinkable, let's do the undoable. Let us prepare to grapple with the ineffable itself, and see if we may not eff it after all."

— Douglas Adams

Acknowledgements

I want to thank the many people who have guided me to this point in my study of physics. I would first like to express my gratitude to my professors at Carleton College for nurturing my love of physics and providing me a solid foundation in the field. My introduction into high energy experimental physics at UW-Madison was led by Professors Wesley Smith and Sridhara Dasu. Professor Smith's enthusiasm for particle physics inspired me to purse this field of study. Professor Dasu first showed me the details involved in conducting a particle physics experiment and provided me with valuable instruction throughout my graduate school experience. While working at CERN, scientist Monika Grothe encouraged me to try new analysis strategies and to become more involved with the CMS collaboration. Scientist Pam Klabbers inspired me by the extent of her knowledge and how well she performed her job. Post-doc Jonathan Efron helped me develop my analysis. Scientists Vitaliano Ciulli and Ilaria Segoni, along with the EWK and V+Jets groups, provided me an opportunity to work on one of the early physics results with CMS. In addition I want to thank the UW computing staff for helping solve a variety of computing problems related to my work, and also thank the UW and CMS support staff for help with the logistics of relocating to CERN. Fellow UW students Marc Weinberg, Ian Ross, Christos Lazaridis, Michalis Bachtis, and Mike Anderson, thank you for hashing out ideas and strategies with me and providing much needed commiseration. I would also like to thank students Jessica Leonard, Josh Swanson, and Lindsey Gray for their friendship and help, and, in particular, thank Jeff Klukas for providing an excellent and easy-to-use tool for making beautiful plots. Finally, I would like to thank my parents, Tom and Jeri, for their enthusiastic support in my six-year quest for a Ph.D.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

> *In the beginning the Universe was created. This has made a lot of people very angry and has been widely regarded as a bad move.*
>
> — Douglas Adams

1.1 Standard Model

The current theory of elementary particles, known as the Standard Model (SM) [1, 2, 3], is a powerful predictor of the behavior of the universe at the quantum level. Ever since the discovery of subatomic particles in the early 20th century, scientists have been refining the models used to describe and explain particle behavior. While the Standard Model is by no means complete, or entirely correct, it is the best tool presently available for describing particle interactions.

The Standard Model describes two elementary matter particle types—leptons and quarks. These particles interact with each other through the electromagnetic, strong, and weak forces, manifesting themselves as "force carrier" particles. Gravity, the only other known force, is not part of the SM because it is too weak to easily study at the particle level.

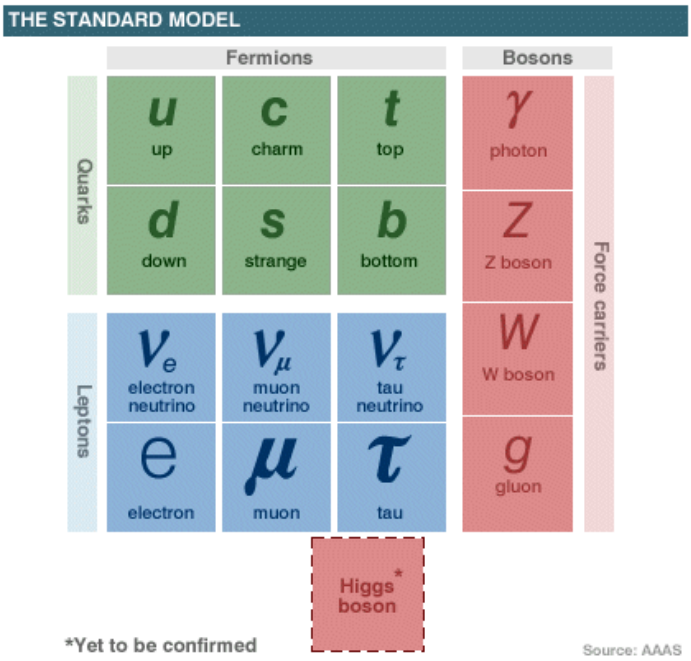


Figure 1.1: Schematic of the particles in the Standard Model and their associated symbols and groupings.

1.1.1 Particles and Interactions

Figure 1.1 lists the known and postulated fundamental particles of the SM. Leptons come in three types or "flavors"—electron, muon, and tau—that, except for mass, share all other properties (such as a charge of -1). For each of the three flavors there is a corresponding neutrino (e.g., electron neutrino, ν_e) that is nearly massless. Table 1.1 on the left lists the properties of each. The anti-leptons have the same properties as leptons except with an opposite charge.

Quarks come in six flavors ranging in mass from 0.003 GeV to 172 GeV[1]. The six flavors are paired into three generations, each having one quark with a +2/3 electric

[1]All masses, momentums, and energies of relativistic particles will henceforth be given in units of energy (eV) since they can all be related through the speed of light using the famous equation $E = mc^2$ and the approximation $pc \approx E$.

Leptons		
Fermion	mass	charge
electron	115 keV	-1
ν_e	< 2 eV	0
muon	106 MeV	-1
ν_μ	< 0.170 MeV	0
tau	1.78 GeV	-1
ν_τ	< 15.5 MeV	0

Quarks		
Fermion	mass	charge
up (u)	3 MeV	+2/3
down (d)	6 MeV	-1/3
charm (c)	1.3 GeV	+2/3
strange (s)	100 MeV	-1/3
top (t)	172 GeV	+2/3
bottom (b)	4.2 GeV	-1/3

Table 1.1: Tables of the fermions and their properties [4], including charge and mass (approximate for quarks, as they cannot be isolated). Antimatter has opposite charges.

charge and one with a -1/3 electric charge, relative to the charge of the electron. The two lightest quarks, up and down, make up protons and neutrons. See Table 1.1 on the right for a list of quark properties. Free quarks have not been observed; they instead come in groupings of two or three. Quarks have one of three "color" charges—red, green and blue—such that the groupings result in colorless objects called hadrons. Hadrons with a quark (e.g., red) and an anti-quark (e.g., anti-red) are designated mesons. Three quarks (anti-quarks) with different colors—red, green, and blue (anti-red, anti-green, anti-blue)—combine into baryons. Collectively, quarks and leptons are known as fermions, sharing the property of half-integer spin. Spin is quantized angular momentum, fundamental to the particle type and used for classification and interaction characteristics.

Interactions between matter particles occur through the exchange of bosons, which are thus known as force carriers. See Table 1.2 for a list of known and postulated bosons and their properties. The photon is massless, has no charge, and mediates electromagnetic interactions known at the particle level as Quantum Electrodynamics

(QED) (see for example [5, 6]). Particles interacting through their charge, such as an electron orbiting the nucleus of an atom, exchange photons. Two massive bosons, the charged $W^{\pm}$ and the neutral Z, carry the weak force. The weak force has a short range and permits fermions to decay and change flavor. The W and Z can interact with all quarks and leptons.

The electromagnetic interactions described by the exchange of photons (QED) and the weak interactions described by the massive W and Z bosons are unified in the Standard Model into a single "electroweak" (EWK) theory [7, 8, 9]. The unification requires the addition of another massive particle, the Higgs boson [10, 11, 12]. The interaction of the SM particles with the Higgs boson, including its self-interaction, results in their masses, giving the Higgs boson special status within the SM. However, the existence of the Higgs boson has not yet been verified experimentally.

The gluon mediates the strong force and, like the photon, is massless and electrically neutral. Gluons carry a color and an anti-color charge and are exchanged by quarks. The strong force binds quarks into mesons ($q\bar{q}$) and baryons (qqq) through gluon exchange and is what holds together the nucleus of an atom. The theory of Quantum Chromodynamics (QCD) describes the strong interactions of quark and gluons, collectively known as partons.

Note that all fermions interact weakly (i.e., through the weak force), whereas only quarks interact strongly. Only the electrically charged particles, quarks, electrons, muons, and taus, interact electromagnetically. Finally, neutrinos, having neither color nor electric charge, only interact weakly.

Boson	force	spin	mass (GeV)	charge	force strength
photon	electromagnetic	1	0.0	0	10^{36}
Z^0	weak	1	91.2	0	10^{25}
$W^{\pm}$	weak	1	80.4	± 1	10^{25}
gluon	strong	1	0.0	0	10^{38}
Higgs	–	0	> 114	0	—

Table 1.2: Table of boson properties [4], including the force each carries, the mass, and the charge. The strength of the forces are given relative to gravity.

1.2 Electroweak Physics

The electroweak force is a unification of the weak and electromagnetic forces, the idea being that above a unification energy of several hundred GeV, the two forces are essentially the same. Electroweak interactions involve the photon, Z boson, and $W^{\pm}$ boson force carries. The weak force as a function of distance (spatial derivative of the potential) is given by the equation

$$F_{weak} \approx \frac{d}{dr}[\frac{1}{r}e^{-m\,r}] \tag{1.1}$$

where m is the mass of the W or Z boson, and r is the distance. By using the mass of the photon (zero) instead, one recovers the familiar inverse square law for the electromagnetic force. Because of the exponential part of the function and the high masses of the W and Z bosons (80.4 and 91.2, respectively), the weak force has a very short range.

These bosons are formed from, or decay into, pairs of leptons and quarks. Depictions of the interaction vertices between W/Z bosons and fermions are given in Figure 1.2 and the boson self interaction[2] is given in Figure 1.3. Each fermions-to-

[2]It is also possible for massive bosons to self-interact, with either a 3 or 4-boson vertex, because they carry weak charge.

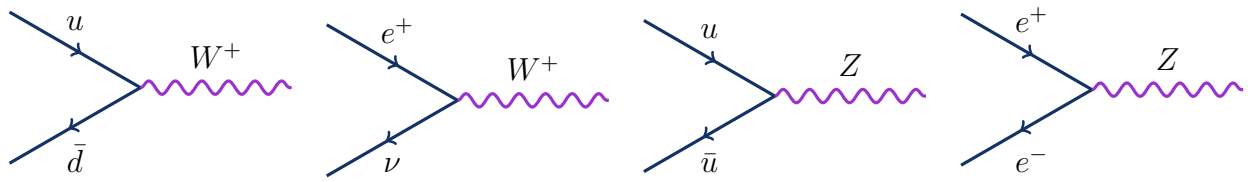


Figure 1.2: Sample interactions vertices between quarks and leptons and W or Z bosons.

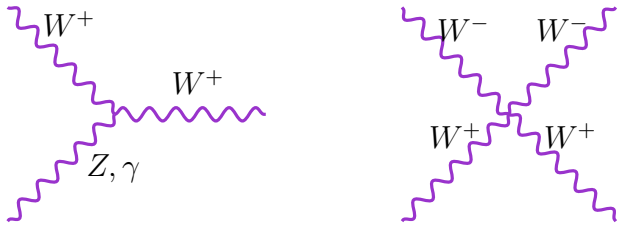


Figure 1.3: Sample self-interaction vertices for W and Z bosons.

W/Z interaction vertex adds another factor of the EWK coupling strength, α_W, to calculations of the likelihood of the interaction. In this paper, the focus will be on the formation of $W^{\pm}$ bosons from partons (in protons), such as $u\bar{d} \rightarrow W^+$, and on the decay of the W into an electron and a neutrino, $W^{\pm} \rightarrow e^{\pm}\nu$.

1.3 Quantum Chromodynamics (QCD)

Quantum Chromodynamics (QCD) governs the interactions of partons through the strong force. Unlike the electrically neutral force carrier photon in QED, the force carrier for QCD, the gluon, carries a color charge itself and comes in eight color variations [13]. Therefore, QCD gluons can self interact, unlike the photons of QED, leading to significantly different behavior for the strong force, such that the strength increases with distance (although it acts only over small distances). Figure 1.4 shows possible interactions between quarks and gluons, as well as gluon self-interaction.

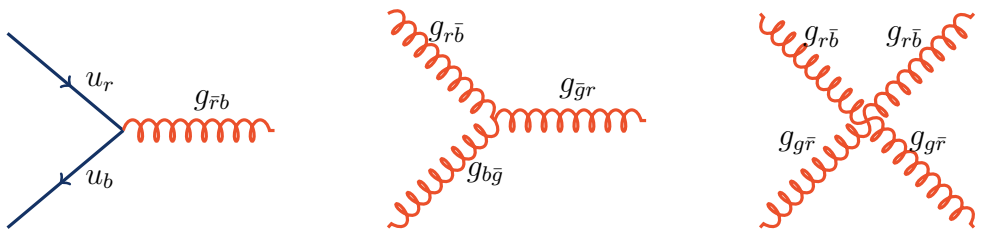


Figure 1.4: Sample interactions vertices between quarks and gluons, or gluon self-interaction, with the red (r), green (g), blue (b) color flow indicated.

Because the strong coupling strength is small at high energies or short distances, QCD can be treated with perturbation theory[3]. However, the coupling for the strong force, α_s, varies with the energies being studied. As the energy decreases, the coupling increases, such that at low enough energies the theory can no longer be approximated with small perturbations. Further discussion of QCD is provided in Chapter 2.

Quarks are "color confined" [14] and always form into colorless hadrons. This confinement occurs because the strong force increases with distance up to a point after which it levels off, about 1 fm. As partons leave the interaction point, the strong force essentially begins to decelerate them. The deceleration causes the partons to radiate more partons in the form of hadrons (much like bremsstrahlung for accelerating electrons) [15]. In terms of energy, it is easier to create new $q\bar{q}$ pairs than to let the existing quark move farther away at greater energies. This process continues with the new hadrons as the individual partons move apart. The two families of hadrons, mesons and baryons form a wide range of particles, depending on the flavor and orientation of the quarks. Many hadrons are unstable and decay quickly, while others

[3]Starting from a known mathematical solution, A_0, small adjustments, $\epsilon^n A_n$, are made to fit the more complicated physical situation, A, that itself cannot be solved fully: $A = A_0 + \epsilon A_1 + \epsilon^2 A_2 + \cdots$. The "order" of the perturbation corresponds to the power of ϵ.

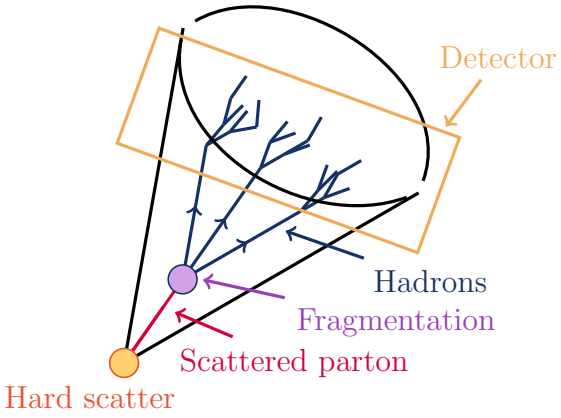


Figure 1.5: Illustration of the evolution from the hard scattering parton to the jet in the detector.

like charged pions, kaons, protons, and neutrons are able to travel a detectable distance before decaying into lighter hadrons or leptons and photons. The particles that are formed from the original parton during the hard interaction leave the interaction point as a collimated spray of particles. The spray or "jet" is what the experiment eventually detects. A cartoon depicting the hadronization process is given in Figure 1.5.

1.4 Proton Collisions

Because protons are composite particles, collisions between protons do not involve the entire object. Instead, one of the partons from a proton will interact with one from the other proton. Protons are made up of three valence quarks, uud, but the valence quarks exchange and radiate gluons that produce virtual "sea" quarks [16] within the proton after splitting into $q\bar{q}$ pairs. Any of these "sea" quarks may interact, making $q\bar{q}$ interactions possible, but less likely, between two protons. More likely are quark-gluon interactions, especially at higher energies. In proton-proton collisions, in-

Figure 1.6: Feynman diagrams for sample QCD (left) and EWK/QED (right) processes possible from pp collisions.

teractions occur that involve QCD, QED, and EWK. An example of each is shown in Figure 1.6 as a Feynman diagram illustrating the $2 \rightarrow 2$ process, using the interaction vertices diagrammed earlier.

The "cross section" associated with these processes are calculated using Quantum Field Theory (QFT) [17] and measured in particle colliders. Cross section, given in units of area[4], essentially refers to the likelihood of a particular interaction occurring (such as those in Figure 1.6). The more likely an interaction is to occur, the larger the effective "area", and the more frequent the occurrence. The rate of events produced for a given process depends on the luminosity, $\mathcal{L}$, i.e., the total number of pp interactions per unit time, and the cross section for the process:

$$rate = dN/dt = \sigma\mathcal{L} \tag{1.2}$$

The size of the cross section depends on the strength of the interactions involved. In general, the QCD cross section is the largest, being governed by the strong(est) force. The next largest cross section is QED followed by EWK processes involving the weakest force.

The Standard Model offers predictions of cross sections which need to be verified at colliders with higher energy levels. Figure 1.7 shows the cross sections of several

[4]The preferred unit of area for particle cross sections is the barn = 10^{-24} cm^2.

QCD and EWK processes expected from proton-(anti)proton collisions, including that for W production, σ_W. The cross sections of most processes are dependent on the center of mass energy of the collision, $\sqrt{s} = \sqrt{P_1^2 + P_2^2}$. More details on cross section calculations are presented in Chapter 2.

1.5 W+jets cross section

The W and Z bosons were first discovered at CERN[5] using a $p\bar{p}$ collider which achieved the high energy collisions needed to produce such massive particles [19, 20]. Because of conservation of energy (and $E = mc^2$), real particles of mass M (GeV/c^2) cannot be created unless the interacting particles have a center-of-mass energy of at least M (GeV). The W^+ bosons were produced primarily through the interaction of $u\bar{d}$ quarks in the protons/anti-protons (or $\bar{u}d \rightarrow W^-$). The heavy bosons had already been postulated, along with their masses, as a necessary element of the Standard Model. Thus, their discovery was an achievement in the predictive power of the SM.

The desired measurement in this paper is the production cross section of W+$n$jets from proton-proton collisions. The interaction producing W+jets (partons) can be represented by a variety of Feynman diagrams, two examples of which are shown in Figure 1.8. The cross section for each process are calculated from these diagrams using the Feynman rules for the interaction vertices and force propagators. The sum of all possible diagrams for W+$\geq n$ jets gives the inclusive cross section for n jets.

[5]The European Organization for Nuclear Research, an international collaboration that, among other projects, operates particle colliders.

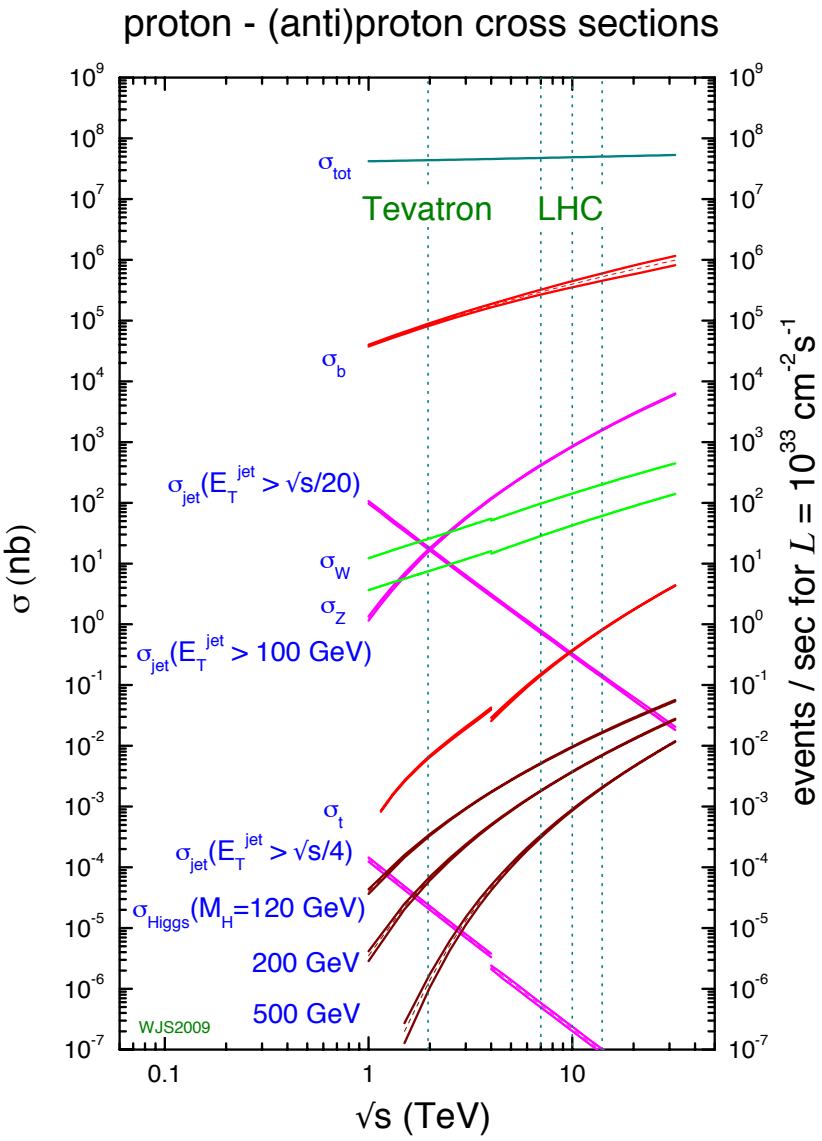


Figure 1.7: Production cross section as a function of center of mass energy for several processes at proton-(anti) proton colliders, with σ_W around 30 nb for the LHC [18].

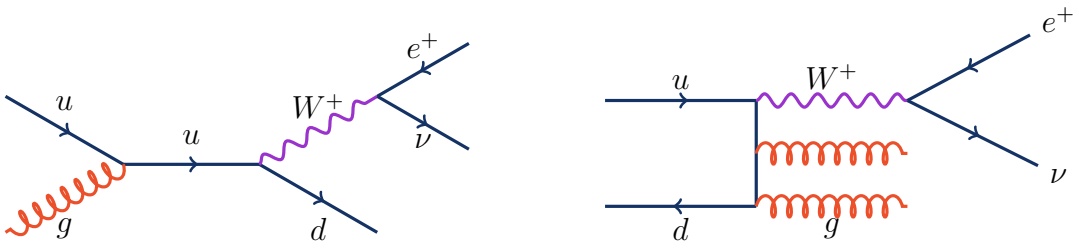


Figure 1.8: Sample Feynman diagrams for W+jets production. On the left, starting from an up quark and gluon and resulting in electron, neutrino, and down quark (becomes a jet). On the right, starting from an up quark and down anti-quark and resulting in electron, neutrino, and two gluon jets.

1.6 Need for Higgs and/or Physics Beyond Standard Model

Thus far, the Standard Model has served physicists well, but does not provide a complete description of particle interactions. There are a number of areas where it breaks down, such as in very high energy regions not yet explored through experiment, that need to be explained. In addition, there are 25 parameters that cannot be calculated directly from the current theory which must be determined from experiment only. Among them are the masses of the fermions, the strength of the forces, and those parameters describing the coupling of the Higgs boson (its mass still being unknown).

The Higgs boson, which is integral to the SM, is a particle postulated to account for spontaneous symmetry breaking [21] and the presence of mass in the universe. The Higgs mechanism is a necessary element to give the W and Z bosons mass; otherwise, the bosons in the SM are massless. Since bosons are known to have considerable mass, finding evidence of the Higgs boson is one of the primary goals of modern particle physics. Because Higgs boson production can have final states similar to W+jets, such as when a Higgs is formed from and decays into two W bosons, W+jets is a major "background" process to Higgs boson production. A detailed understanding

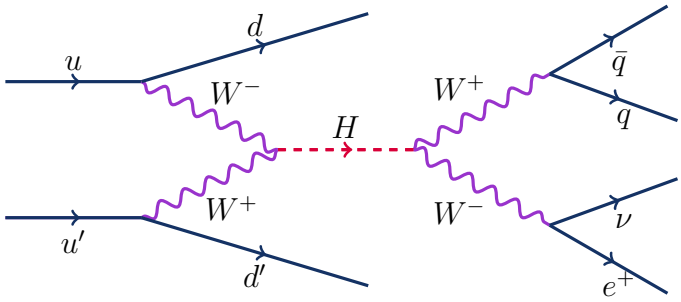


Figure 1.9: Feynman diagram for vector boson fusion Higgs production resulting in an electron, neutrino, and two quark jets.

of SM processes, including their expected cross sections and kinematic distributions, is necessary to isolate potential new physics signals. Figure 1.9 presents a Feynman diagram showing theoretically how Higgs boson production might occur through W bosons with a final state including an electron, a neutrino, and two quarks (jets) [22, 23]. W+jets is also potentially a background process to the production of other new physics particles beyond the Standard Model, especially ones that decay to W bosons.

Chapter 2

Theory of W+Jets Production

Parton production in association with W bosons ("W+jets") is a relatively clean signal with a high cross section. The theoretical understanding of the electroweak portion of the W boson production is well developed, thus enabling a thorough study of the associated jet production mechanism described by perturbative QCD (pQCD) [24]. Section 2.1 will provide an overview of the quantum field theory behind the production of W+jets (partons), while Section 2.2 will cover the specifics of the W boson decay and detection. Section 2.3 will focus on the final state partons and pQCD aspects of W+jets.

2.1 W Boson + Jets Production

The full cross section of hadronic scattering, illustrated in Figure 2.1, is represented by the equation

$$\sigma = \frac{1}{3}\sum_{q,q'}\int dx_1 dx_2 f_1(x_1,Q^2) f_2(x_2,Q^2)\ \hat{\sigma}(\hat{s}). \tag{2.1}$$

where x_1 and x_2 are the proton momenta fractions that are carried by the interacting partons and $\hat{\sigma}(\hat{s})$ is the cross section at the parton level [25]. The hard scattering cross section at the parton level is calculated using EWK theory. For example, the cross

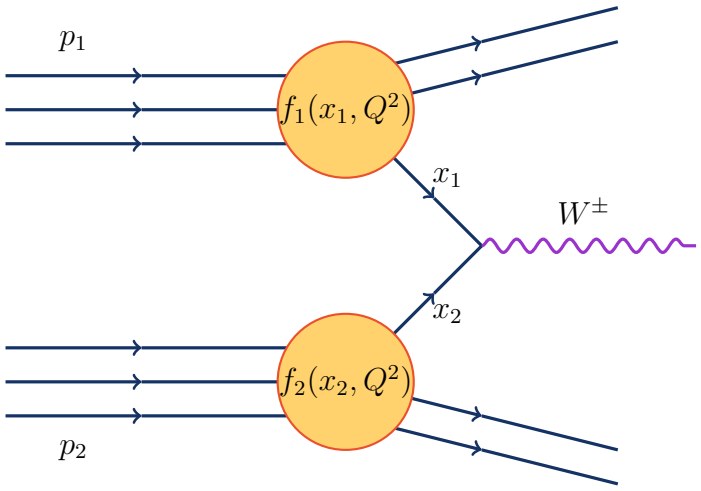


Figure 2.1: Formation of a W boson through deep inelastic scattering of two protons, p_1 and p_2, shown as an interaction between two of the constituent particles with momentum fractions x_1 and x_2.

section for $u\bar{d} \to e^+\nu$ as a function of the parton center-of-mass energy $\hat{s}$ is calculated to be

$$\hat{\sigma}(u\bar{d} \to e^+\nu) = \frac{|V_{ud}|^2}{3\pi}\left(\frac{G_F M_W^2}{\sqrt{2}}\right)^2 \frac{\hat{s}}{(\hat{s} - M_W)^2 + \Gamma_W M_W)^2}. \tag{2.2}$$

The up (u) and anti-down ($\bar{d}$) quarks may be exchanged for other quark types such as $u\bar{s}$ or $c\bar{s}$. The parameter V_{ud} indicates the strength of the flavor changing decays, with ud and cs being the most likely.

In Equation 2.1, the summation is over the possible quarks and anti-quarks, and the factor of 1/3 is derived from the three possible quark color charges. Parton distribution functions (PDFs) [26, 27], $f_i(x_1, Q^2)$ and $f_i(x_2, Q^2)$, describe the composition of the protons in terms of the constituent partons. The parameter Q is the characteristic momentum scale of the scattering, in this case using $Q = M_W$. The cross section in Equation 2.1 corresponds to the W cross section shown in the middle part of Figure 1.7.

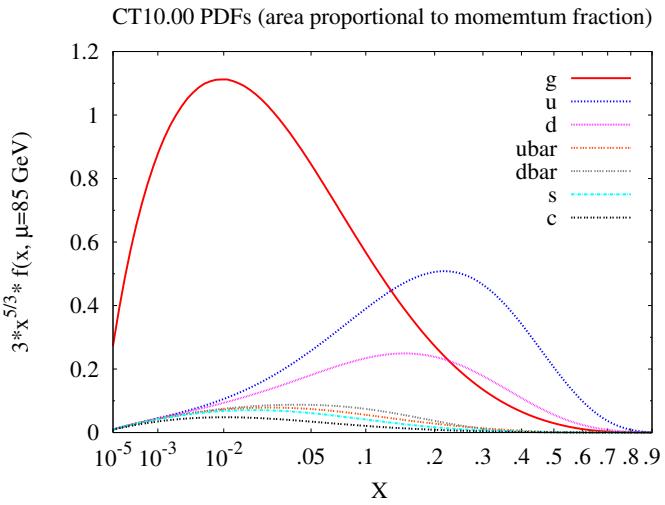


Figure 2.2: Example parton distribution function from the CT10 set of PDFs, for $Q = 85$ GeV [28].

Figure 2.2 shows a sample composition of the particles in a proton for two different Q^2 values. The parton distribution functions describe the number of particles of type i in the proton with a momentum fraction between x and dx. Because neither of the two protons has valence anti-quarks, the most likely interaction for pp colliders is a gluon from one proton interacting with a quark from the other, shown in Figure 1.8 on the left. The PDFs have been previously determined experimentally [28] at the Tevatron collider [29].

The process described above covers only the $q\bar{q}$ production and $e\nu$ decay of W bosons and does not take into account our interest in studying the partons that are produced along with the W bosons. The probability of producing an additional parton is roughly dependent on the strong coupling constant, α_s (i.e., each quark-quark-gluon interaction vertex adds another factor of α_s), although higher order corrections from internal loops in the diagrams can change this dependency by about 10% [30]. The

coupling constant α_s governs the strength of the strong interaction and depends on the momentum transfer scale, Q, of the interaction: $\alpha_s(Q^2) \propto 1/\ln(Q^2/\Lambda^2_{\rm QCD})$ where $\Lambda_{\rm QCD}$ is the momentum scale for QCD above which pQCD becomes valid.

2.2 W Boson Decay

W bosons couple to both quarks and leptons roughly equally and can decay to any of six quark combinations (excluding the top, which is more massive than the W [31, 32]), or one of the three lepton flavors. Each of the six quark combinations is multiplied by the three possible quark colors. However, the likelihood of the W decaying to a given quark combination is less than one, as described by the CKM matrix $V_{qq'}$ [33, 34]. The three combinations for the up and charm generations each add to about one. The likelihood, or branching ratio (BR), of the W decaying to quarks is thus 2/3 and to leptons is 1/3. Each of the three leptonic decays—electron, muon, or tau—is roughly equally likely to occur and thus the electron-neutrino branching fraction is about 1/9 of all W bosons decays. When measuring only the decay of W bosons to an electron and a neutrino, one may start from the total theoretical cross section and multiply by the branching ratio:

$$\sigma(W \to e\nu + n\ partons) = \sigma(W + n\ partons) \cdot BR\,(W \to e\nu). \tag{2.3}$$

The measured branching ratio for $W \to e\nu$ is 10.8% [4]. Because the W boson is massive (80.4 GeV) and has a very short lifetime ($10^{-}25$s, particle detectors can only measure the decay products, e.g., the electron and neutrino. Electrons from the W boson usually carry away about half the W mass (the other half going to the neutrino) and thus the expected energy of the relatively light electrons is about

40 GeV, a readily detectable amount. Complicating the measurement is the fact that neutrinos are nearly massless and only interact via the weak force, leaving no trace in the detector. The only evidence that neutrinos are present is an imbalance in the transverse momentum of all detected particles. In a two beam collider, the total initial transverse momentum (i.e., the momentum in the plane perpendicular to the beam path) is zero. During and after the collision, momentum should be conserved. Any significant transverse momentum that does not cancel is attributed to an undetected neutrino and is called "missing transverse energy", $\not\!\!E_T$. It is calculated from the x and y components of the energy of all of the particles i in an interaction event:

$$\not\!\!E_T = -\sum_i \left(E_x^i \, \hat{\mathbf{x}} + E_y^i \, \hat{\mathbf{y}} \right). \tag{2.4}$$

(It is impossible to calculate the total longitudinal momentum since much of it continues along the beam pipe and is not detected.) The W boson may be partially reconstructed from the electron and the $\not\!\!E_T$. Because only the transverse missing energy can be used, the transverse mass of the W, M_{T}, rather than the full mass, is calculated:

$$M_{\mathrm{T}} = \sqrt{2 \times \not\!\!E_T \times P_T(e)(1 - cos(\theta))} \tag{2.5}$$

where $cos(\theta)$ is the angle between the electron and the $\not\!\!E_T$ in the transverse plane. The maximum value of the M_{T} is the full W mass, 80.4 GeV (plus the width and any smearing effects from the detector).

2.3 Associated Jets

The production cross section of W bosons in conjunction with other outgoing partons was outlined briefly in Section 1.5. At higher values of Q^2, such as in W

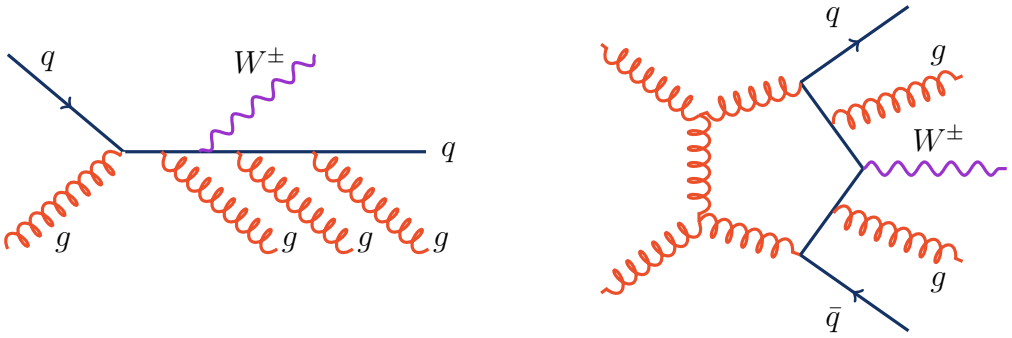


Figure 2.3: Feynman diagrams showing two examples of W + 4-jet events.

boson production, the coupling constant α_s becomes smaller and perturbative QCD is applicable. Calculations done at the tree-level diagram, with just incoming and outgoing particles (no internal loops), are fairly straightforward and give leading order (LO) results. Calculations including virtual contributions (internal gluons and particle loops) and additional partons are considered to be the next-to-leading order (NLO) in perturbative QCD, or even next-to-next-to-leading order (NNLO) and are more complex. With more partons being produced along with the boson, higher orders (NNLO) of pQCD can be tested.

The calculation of cross sections and simulations becomes much more complicated with additional partons, since there are many subprocesses contributing to a single outcome. For instance, there are 498 possible subprocesses for W + 4-parton production in *pp* collisions [35], two examples of which are shown in Figure 2.3.

In addition, gluons emitted from quarks can form gluon pairs. This additional coupling between gluons because of their color charge causes an inverted screening effect, such that the closer the probe, the smaller the effective color charge. These

internal loops of gluons added to tree-level diagrams create ultraviolet divergences[1] in pQCD. The divergences are removed by introducing renormalization, a mathematical strategy to deal with infinities in cross section integrations. The renormalization scale, μ_R, is not fixed, but rather depends upon the momenta exchanged in an interaction, often set to the same value as the factorization scale μ_f, which deals with low energy, large distance, divergences [15]. The energy scale chosen for calculations cannot affect the final observables.

2.4 Previous Studies of W+Jets

The most recent precision measurements of the W boson come from the two major experiments at the Tevatron collider, D0 [36] and CDF [37]. The dominant production of W bosons at the $p\bar{p}$ Tevatron collider is due to the reaction $q\bar{q'} \rightarrow W$. The center of mass energy, $\sqrt{s}$, at the Tevatron during its Run II was 1.96 TeV. For comparison, the Large Hadron Collider (LHC) [38] is currently running at $\sqrt{s} = 7$ TeV and will eventually increase to $\sqrt{s} = 14$ TeV. Both of the multipurpose experiments at the LHC, the Compact Muon Solenoid (CMS) [39] and ATLAS [40], are currently conducting studies of W+jets production.

2.4.1 Collider Detector at Fermilab (CDF)

The CDF analysis [41] is based on an integrated luminosity of 320 $\pm$ 18 pb^{-1}, where integrated luminosity is a measure of the total number of $p\bar{p}$ collisions. The electrons were required to have an $E_{\mathrm{T}} > 20$ GeV and missing transverse energy, $\not\!\!E_T$, greater than 30 GeV, since the W decay products each share part of M_T^W. The

[1]Ultraviolet (UV) divergences are unphysical effects in calculations using Feynman diagrams when infinitely high energies or small distances are present.

transverse mass was required to be greater than 20 GeV. These requirements are necessary to remove background events that are unlikely to be true W+jets events, but result instead from QCD or $t\bar{t}$ processes. Only the responses from the matter in a particle detector give indications of type and characteristics of the particles coming from an interaction. Requirements must be put on the detectable decay products, electrons, $\not\!\!E_T$, and jets in order to select events from the desired interactions. Events are grouped according to the number of jets with transverse energy $E_{\mathrm{T}} > 15$ GeV. All of the jet counting is done inclusively such that events with 2 jets, for example, are included in both the 1-jet and 2-jet bins. The $\not\!\!E_T$ distribution, distinct for events with W bosons, is used to extract the signal from the remaining background events. Figure 2.4 shows the results of a fit to the $\not\!\!E_T$ with the signal and two types of background events. The background fraction varies from 10% to 40% between 1-jet and 4-jet events, and up to 90% for the high jet E_{T} range.

Figure 2.5 on the left shows the ratio of data to theory for the cross section as a function of jet multiplicity n and the ratio of the cross section for $\geq n$ jets over $\geq n-1$ jets, σ_n/σ_{n-1}. On the right is a plot of the differential cross section in data compared to three different theoretical models as a function of jet E_{T} for events with ≥ 1 jets. Close agreement between the data and NLO theory predictions is seen in both cases.

2.4.2 D-Zero (D0)

The D0 analysis [42] is based on an integrated luminosity of 4.2 fb^{-1}. Events are selected with electron $p_{\mathrm{T}} > 15$ GeV, $\not\!\!E_T > 20$ GeV, and $M_{\mathrm{T}} > 40$ GeV. The jet counting is done for jet $p_{\mathrm{T}} > 20$ GeV. Differential and total cross sections are normalized to the measured inclusive W boson cross section to cancel some of the

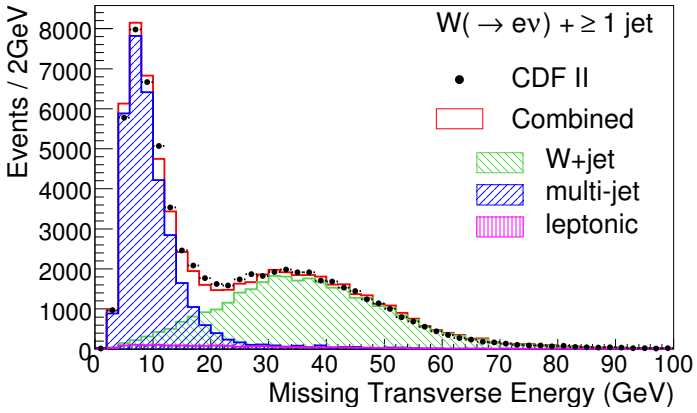
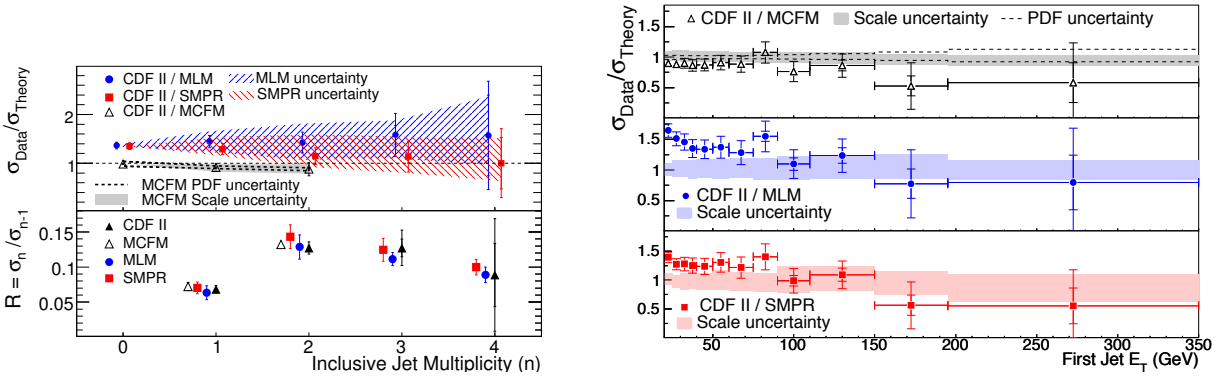
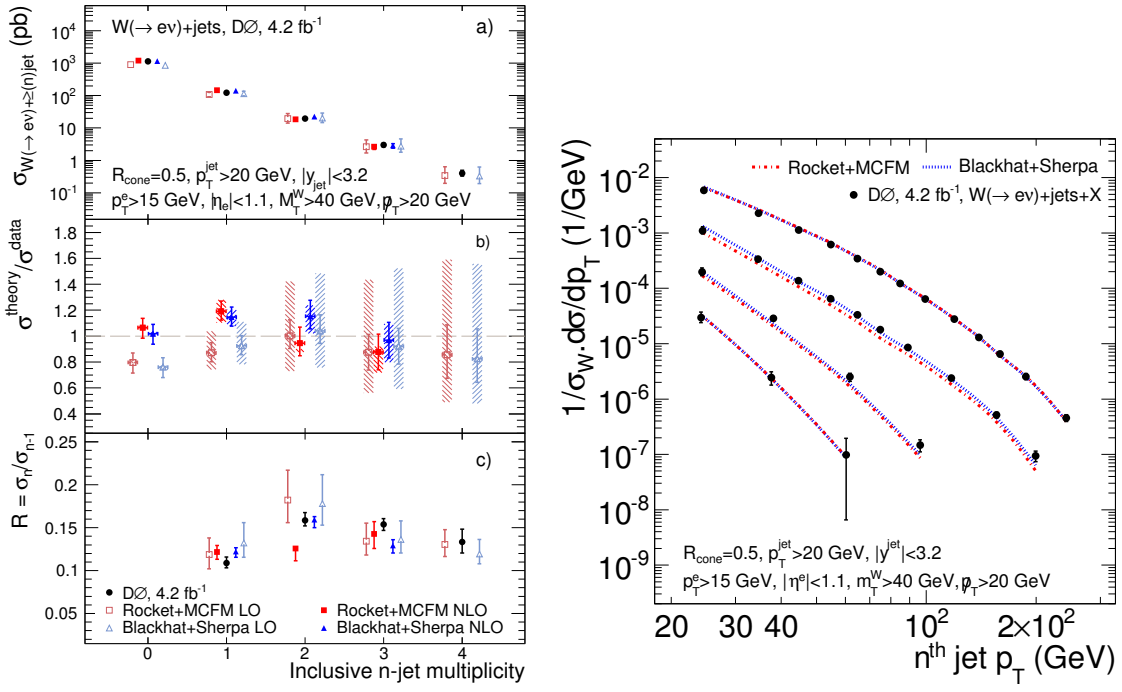


Figure 2.4: CDF results for fitting template distributions of the $\not\!\!E_T$ to data for the W + $\geq$ 1 jet sample [41].



(a) (b)

Figure 2.5: CDF results for the cross section of W + $\geq$ n jets. (a) The ratio of measured to theoretical cross section as a function of inclusive jet multiplicity, (b) the ratio of measured to theoretical cross section as a function of jet E_{T} [41].



(a) (b)

Figure 2.6: D0 cross section results. (a) The total cross section as a function of the inclusive jet multiplicity, the ratio of theory to data measurements, and the ratio of σ_n/σ_{n-1} for data and several theoretical models. (b) The differential cross section as a function of the n^{th} jet p_{T}, normalized to the inclusive $W \to e\nu$ cross section and compared to NLO predictions. W+1 jet is the top curve, W+4 jets is the bottom curve [42].

systematic uncertainties. Significant agreement is shown between the theory and data, although NLO theory predictions for the cross section are slightly higher for the 1-jet bin than that found in data. Figure 2.6(a) shows the inclusive W + n jets cross section for each jet multiplicity, compared to two theoretical calculations. Figure 2.6(b) shows the differential cross section by jet p_{T} normalized to the inclusive W cross section. The agreement between theory and data is good, with some deviation at the highest and lowest jet p_{T}.

Chapter 3

Experimental Setup

With the Large Hadron Collider (LHC) operational, physicists are now able to conduct research at far higher energy levels (7 TeV) than those provided by previous accelerators. Measurements of the Standard Model processes are being conducted along with searches for new physics processes. To make these measurements, the Compact Muon Solenoid (CMS) detector was constructed to detect the particles resulting from the interactions created in the LHC.

3.1 Large Hadron Collider

As part of physicists' endless endeavor to probe higher energies and smaller distances, the Large Hadron Collider (LHC) [38] was constructed at CERN. The LHC uses an underground ring 27 km in circumference that originally housed an electron-positron collider (LEP). The energies intended for the LHC made it impossible to continue using electrons, as they are light-weight and radiate too much energy when accelerated around a ring. Instead, two beams of protons are collided, each at an energy of 3.5 TeV, producing a wide range of physics events to study at a center of mass energy of $\sqrt{s} = 7$ TeV. Eventually the LHC will move to the design energy of

$\sqrt{s} = 14$ TeV. The maximum center of mass energy is constrained by the size of the LHC ring and the limits of magnet technology.

Luminosity is important because the number of events that are produced for a particular process is directly proportional to the time-integrated luminosity and the cross section of the event, σ_{event}:

$$N_{event} = L\sigma_{event} \tag{3.1}$$

A high luminosity is necessary to search for very rare processes. To achieve a high luminosity, the LHC uses proton-proton collisions, which are easier to produce than the proton-anti-proton collisions used at the Tevatron. The cross-section dependence on center of mass energy was shown in Figure 1.7.

Protons begin their journey in the Linac2, a linear accelerator that uses radio frequency (RF) cavities acting on the charged protons to boost them to 50 MeV. They are then injected into the Proton Synchrotron Booster (PSB) which ramps up the energy to 1.4 GeV, also by using RF cavities. At this stage magnets begin squeezing the beam down in size, because smaller beam size implies a higher luminosty. The PSB starts with six bunches of protons that then get split into three. The next stage, the Proton Synchrotron (PS), continues increasing the beam energy, this time up to 24 GeV. Also at this point there is more bunching of the protons, so that each bunch gets split into two, twice. The bunches are shortened using an 80 MHz RF system so that they can fit into the 200 MHz brackets of the next stage. The final stage before injection into the main LHC ring is the Super Proton Synchrotron (SPS). The SPS increases the energy to 450 GeV. At this point there are ideally 2,808 bunches, each with about 10^{11} protons. The RF fields oscillate at the frequency of 40MHz so that the

bunches are spaced 25 ns apart. When the LHC was first being commissioned, only one bunch was placed in each beam. As running continued smoothly, more and more bunches, and more protons per bunch, were added until each beam had 368 bunches for the final 2010 runs. Once the protons are in the LHC ring, they are accelerated to the target energy of 3.5 GeV for each beam [38]. A diagram of the LHC accelerator complex, from Linac2 to the detectors, can be seen in Figure 3.1.

Given that both beams have the same electric charge and need to accelerate in opposite directions, they cannot share a common beam pipe and magnets. Instead, special dipole magnets are used to direct the beams in opposite directions. The magnets have a compact, twin-bore design to fit into the existing 3.7 m tunnel width. They have two sets of coils and two beam channels in the same structure and cooling system, and thus are both mechanically and magnetically coupled. The beams share a common beam pipe only at the interaction regions, for a length of about 130 m. The magnets are made of NbTi and cooled using liquid helium to a superconductive state at a temperature of 1.9K. In all, 1232 dipole magnets, each with a field of 8.3 Tesla, are used to bend the proton beams around the ring. Quadrupole magnets are used around the tunnel for focusing the beam.

3.1.1 Luminosity

Luminosity is a function of the number of protons in each bunch (N_b), the number of bunches in the beams (n_b), the revolution frequency of the beams (f), and the size of the beams (effective collision area, A_{eff}). It is roughly calculated as

$$\mathcal{L} = \frac{N_b^2 \times n_b \times f}{A_{\text{eff}}} \tag{3.2}$$

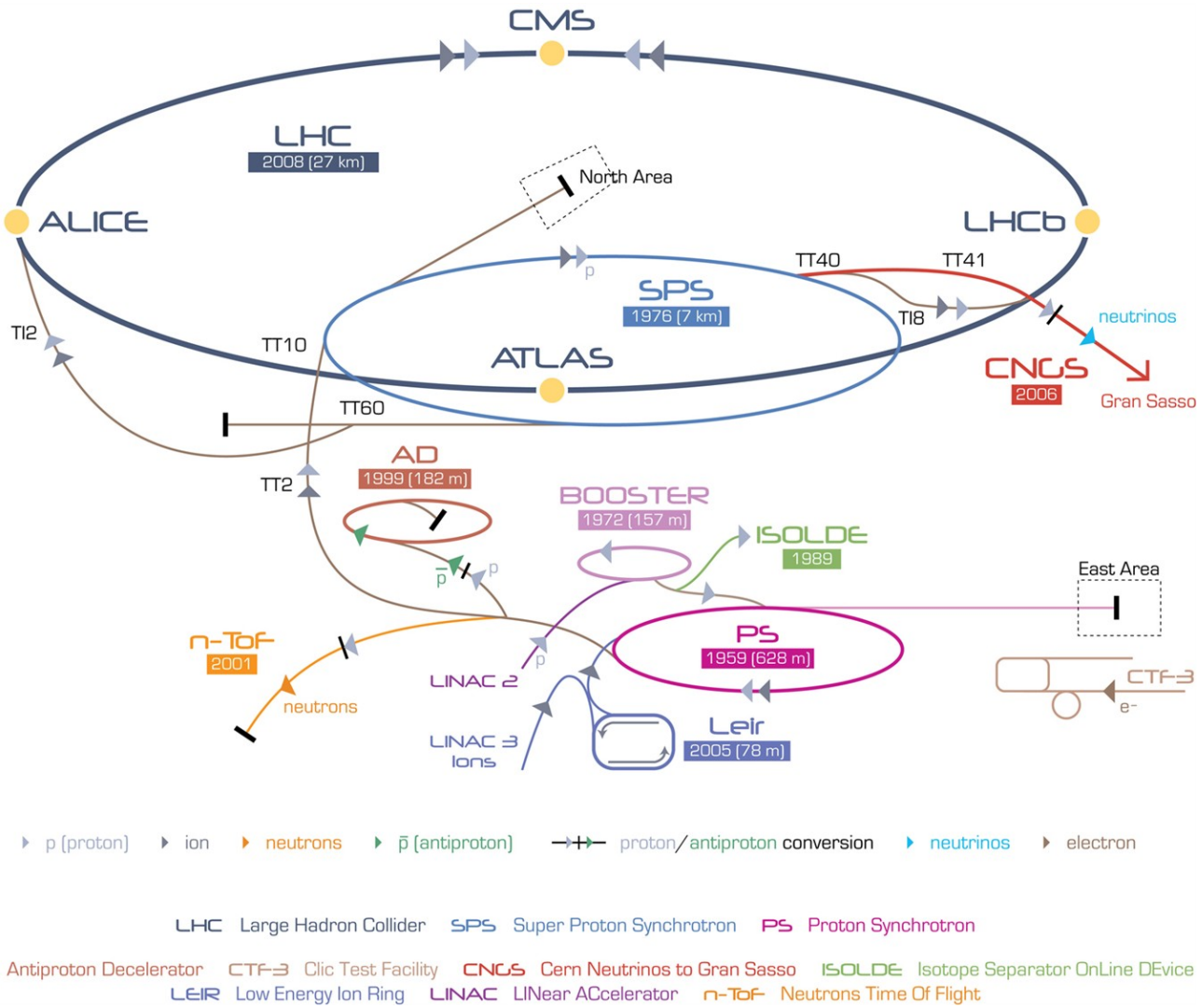


Figure 3.1: Diagram of the CERN accelerator complex showing the initial synchrotrons, the LHC ring, and the locations of the four major experiments, CMS, Atlas, LHCb, and ALICE [43].

for two colliding beams with the same parameters. "Integrated luminosity", $L = \int \mathcal{L}\, dt$, corresponds directly to the number of collisions collected in a given time frame.

Approximately 400 quadripole magnets are used to focus the beams to 16.7 μm. The smaller beam size means a smaller effective collision area, A_{eff}, and a higher luminosity. The frequency of the revolution is 11.246 kHz. The number of particles, number of bunches, and the frequency of revolution are well known quantities. The effective area is calculated from the transverse widths of the beams, σ_x and σ_y, assuming gaussian profiles:

$$A_{\text{eff}} = 4\pi \times \sigma_x \times \sigma_y. \tag{3.3}$$

During the first year of running, the instantaneous luminosity rose to 2.07×10^{32} cm^{-2}s^{-1} (design luminosity is $L = 10^{34}$cm^{-2}s^{-1}). The total integrated luminosity for the 2010 proton-proton physics runs is 36.1 pb^{-1}. The luminosity uncertainty is roughly 4% and is dominated by uncertainty in the size of the beams (A_{eff}). The luminosity is related to the cross section of a physics process through

$$\sigma = \frac{N_{\text{obs}}}{(A \times \epsilon \times L)} \tag{3.4}$$

where σ is the cross section, N_{obs} is the number of observed events, A is the acceptance (fraction of the events that can be found by the detector), ϵ is the efficiency of the selection, and L is the integrated luminosity [38].

3.2 Compact Muon Solenoid

On the French side of the LHC is the 12,500-ton Compact Muon Solenoid (CMS) detector (Figure 3.2) [39, 44]. It is located at one of four interaction points, where the two proton beams are diverted into collisions. Along with ATLAS, it is one of two

general purpose detectors searching for the Higgs boson and new physics signatures. CMS is sectioned into several sub-detectors, each with specific functions. The innermost part of the detector is the tracker, followed by the electromagnetic calorimeter and the hadronic calorimeter. These three sub-detectors are within a solenoid magnet, which is surrounded by muon chambers.

The location of the calorimeters inside the solenoid creates a much more compact design (hence the name), and better detection of particles in the calorimeters since they do not first traverse the material of the magnet. The CMS magnet is a NbTi solenoid that generates a large magnetic field of 3.8 T. The field bends charged particles in the tracker to enable measurement of their momentum. An iron yoke interspersed with the muon chambers around the solenoid carries a return field of 2 T, making further momentum measurements of muons possible. The magnet length is 12.5 m, while the inner diameter is 6.3 m, allowing a large radius for tracking (i.e., better resolution). The high length to radius ratio of the solenoid means a very uniform magnetic field. Figure 3.3 shows a slice of the detector with the main sub-components labeled along with which particles are detected in each area.

3.2.1 CMS Geometry

CMS has a cylindrical shape with a diameter of 14.6 m and length of 21.6 m. It is divided into three major sections: the barrel, two endcaps, and two forward regions. CMS uses a right handed Cartesian coordinate system with the origin at the interaction point. The x-direction points towards the center of the LHC, the y-axis points up, and the z-axis points west along the beam direction. The radial distance in the $x - y$ (transverse) plane is r. The angle around the detector in the transverse

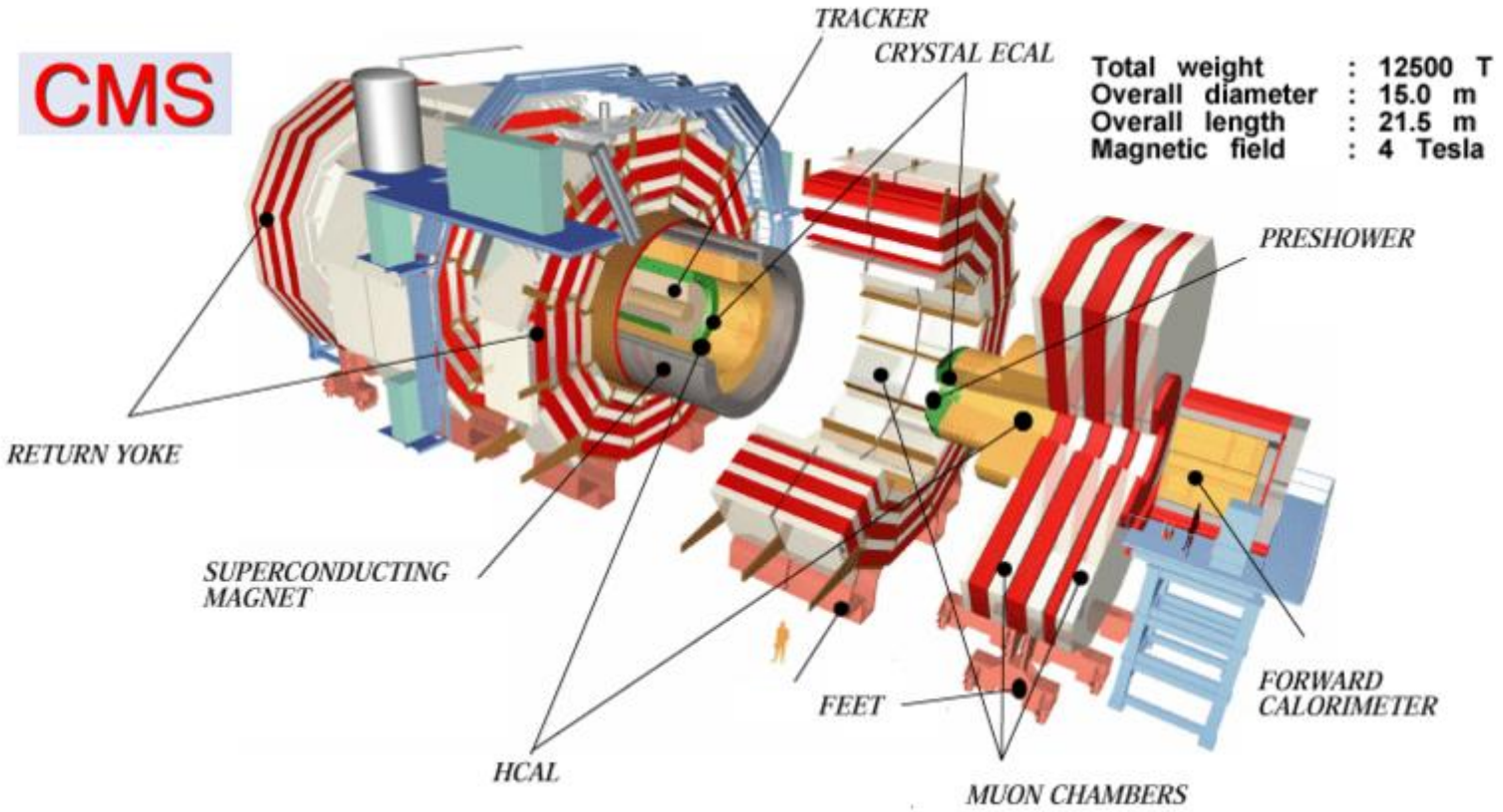


Figure 3.2: 3D Drawing of the CMS detector using with sub-detectors labeled. A human figure is included for scale [45].

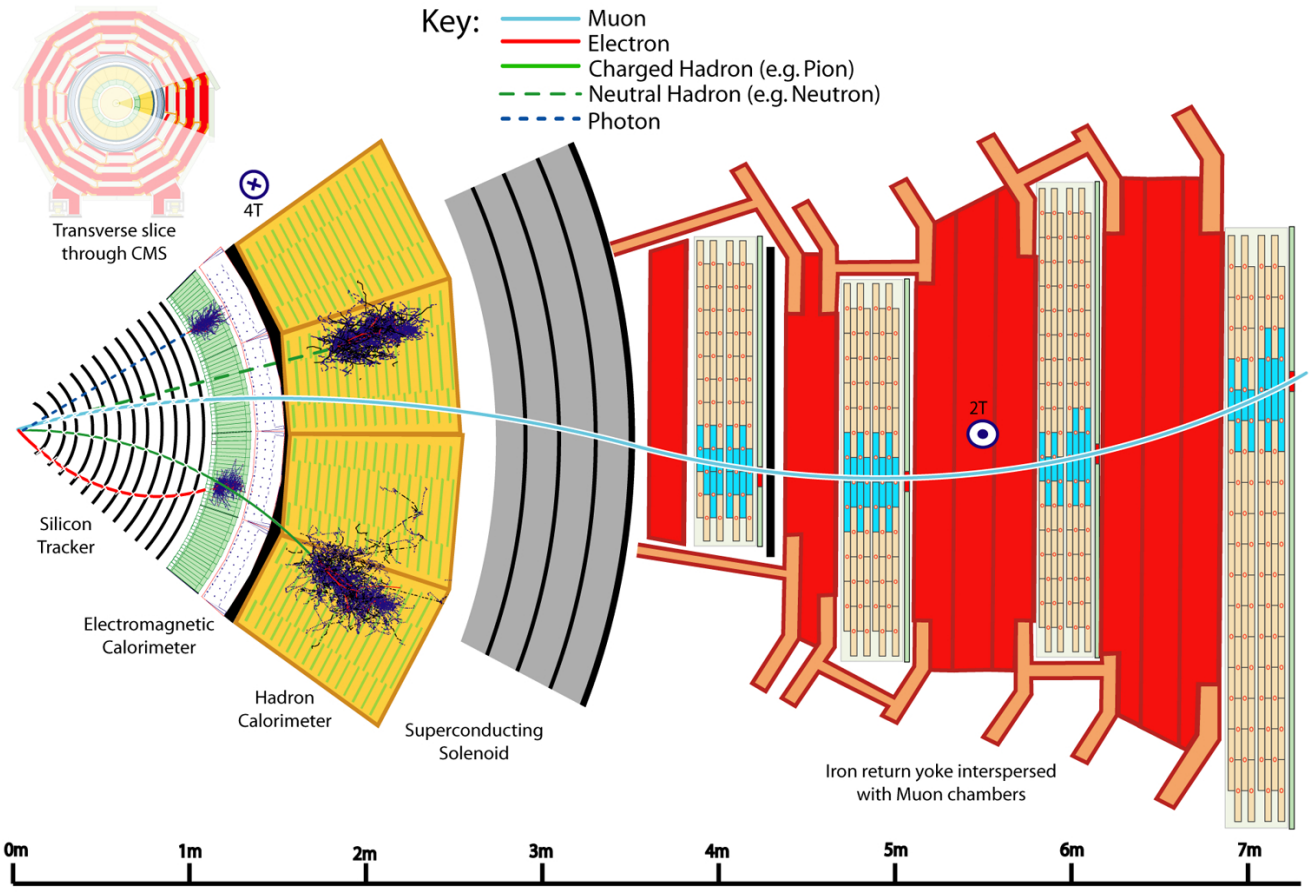


Figure 3.3: A slice of the CMS detector and the path that various particles travel through it. Electrons (red solid lines) leave a hits in the trackers and deposit all of their energy in the Electromagnetic Calorimeter. Charged and neutral hadrons, components of jets, leave the bulk of their energy in the Hadron Calorimeter, while muons continue to chambers outside the solenoid [46].

plane is ϕ. The detector and interactions are roughly uniform in phi, since there is no significantly preferred direction. The angle above the beam path (along the detector) is θ, but almost always "pseudorapidity" (η) is used, since the distribution of particles in η is roughly constant and is defined such that it ranges from 0 (vertical) to infinity (along the beam path):

$$\eta = -ln[tan(\theta/2)] \tag{3.5}$$

For the calorimeters, "barrel" refers to the region of $|\eta| < 1.4442$, "endcap" to the region $|\eta| > 1.566$ and $|\eta| < 3.0$. The gap between the barrel and endcap is used for cables and has fewer detecting components. The hadron forward region covers $3.0 < |\eta| < 5$. The tracker extends to $|\eta| < 2.5$.

3.2.2 Tracker

The tracker subdetector [47, 48] is the first layer that particles from the proton-proton interaction traverse. It uses millions of silicon sensors to measure the trajectory of charged particles that pass through it. Nearest the interaction point are pixels for fine granularity at the region of highest particle flux, while further away, where the particle flux is lower, larger strips are used. The sensors are reverse biased p-n junction diodes. When charged particles pass through the depletion region of the diode, they cause ionization currents that are read out through the electronics. Using silicon sensors makes it possible to have thin layers, and thus short response times and excellent position resolution. Figure 3.4 shows the location and layout of the tracker. Nearest to the interaction region is the pixel detector [49] with 65.9 million pixels covering an area of about 1 m^2. Each pixel has dimensions of about $100\mu m \times 150\mu m \times 250\mu m$. The resolution in the r-ϕ plane is $\sim 10\mu m$, and in the r-z plane is

$\sim$ 20 μm. The pixels are situated in three concentric layers in the barrel with two endcap discs on each side. They range from -2.5 to 2.5 in pseudorapidity. The pixels are kept at -10^{0}C to reduce radiation damage, to keep noise levels low, and to maintain stability. The pixels help establish the vertex of the collision event and initial track hits. They are also important in distinguishing particles with a displaced secondary vertex, such as those from B-hadrons which can travel a detectable distance before decaying. The secondary vertex information is used in this analysis to distinguish $t\bar{t} \rightarrow WbWb$ from W+light flavored jets events.

Surrounding the pixels are 11.4 million silicon strips. Ten layers of these strips in the barrel cover a radius of 20 cm to 110 cm from the beam axis. The barrel is divided into the Tracker Inner Barrel (TIB) for $|r| < 65$ cm and Tracker Outer Barrel (TOB) for $65 < |r| < 110$ cm. The TIB and TOB have a resolution of 230 μm and 530 μm respectively and are oriented along the beam axis. Filling in the gap between the barrel and endcap are Tracker Inner Disks (TID) next to the TIB. The Tracker EndCap (TEC) has 9 layers from 124 cm $< |z| <$ 282 cm, covering up to $|\eta| < 2.5$. Both the TID and TEC are oriented radially to the beam.

The signal to noise ratio is better than 25:1, and the number of dead or noisy strips is less than 3/1000. All together, there is an impressive 210 m^2 of silicon. With tracker information, one can measure the curvature of a charged particle track, and from that, its momentum. Because the tracker is within the solenoid, any charged particle will bend in the magnetic field:

$$p_{\mathrm{T}} = qRB \tag{3.6}$$

where p_{T} is the transverse momentum, q is the charge, R is the radius of the track, and

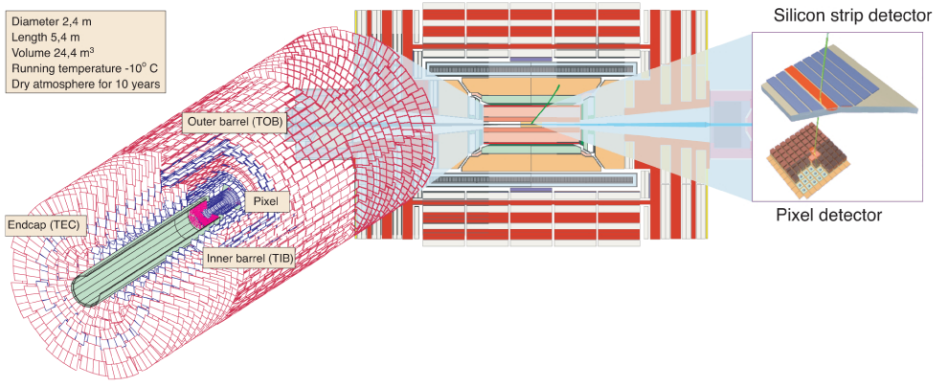


Figure 3.4: Drawing of the CMS tracker subcomponents and detecting elements, pixels and strips [50].

B is the strength of the magnetic field. The tracker has a high resolution because of the large radius and large magnetic field, as well as the fine granularity of the tracker elements:

$$\frac{\sigma_{p_{\mathrm{T}}}}{p_{\mathrm{T}}} \approx (15 p_T \oplus 0.5)\%\ (TeV), |\eta| < 1.6 \tag{3.7a}$$

$$\frac{\sigma_{p_{\mathrm{T}}}}{p_{\mathrm{T}}} \approx (60 p_T \oplus 0.5)\%\ (TeV), |\eta| = 2.5 \tag{3.7b}$$

Because a higher p_{T} results in less curvature, the resolution decreases for higher p_{T} objects. An object with $p_{\mathrm{T}} = 40$ GeV has a tracker resolution of 0.6%, while a 100 GeV object has a resolution of 1.5%. The tracker helps identify electrons from the W boson decay, as well as measure their transverse momentum, p_{T}, and is used to eliminate photons, which being neutral, leave no tracks.

3.2.3 Electromagnetic Calorimeter

After the tracker, particles encounter the electromagnetic calorimeter (ECAL) [51, 52]. Electrons and photons deposit almost all of their energy in the ECAL. It is used, along with the tracker information, to determine the energy and location of

the electrons from W boson decays. The ECAL is composed of 76,832 lead tungstate ($PbWO_4$) crystals which produce electromagnetic showers and transmit their light (energy). As electrons pass through the crystal they emit bremsstrahlung photons that convert to electron pairs, resulting in a shower of light collected by vacuum photodiodes (barrel) or vacuum phototriodes (endcap).

The ECAL is separated into the barrel (EB), the endcap (EE), and a preshower (PS) detector in the endcaps. A drawing of the ECAL and subcomponents is shown in Figure 3.5. The barrel, $|\eta| < 1.4442$, has 61,200 crystals, and each of the endcaps, ranging from $1.56 < |\eta| < 3.0$, has 7234 crystals. The pre-shower, made of Pb-Si, is used to reject photons from π^0 decays. The calorimeter crystals are grouped into supermodules consisting of 1700 crystals, 20 in ϕ and 85 in η. One half-barrel cylinder has eighteen supermodules. The endcaps have 15,632 tapered crystals, with a larger rear face of 30x30 mm^2. The lead tungstate crystals have a density of 8.3g/cm^3, a Molière radius of 22 mm, a radiation length of 0.89 cm, and dimensions of 22 mm $\times$ 22 mm $\times$ 230 mm. The Molière radius, R_M, corresponds to 90% of the lateral spread of an electromagnetic shower. The crystals are each 1 R_M, offering a highly granular ECAL with good position resolution. One of the advantages of using lead tungstate crystals is that they are very dense but optically clear. Particles are likely to interact with the crystal atoms, and the light from the interaction is collectable with the crystal itself. They are radiation tolerant and produce a quick shower: 80% of the light is produced in 25 ns. The energy of an electron traveling a distance x is given by

$$E(x) = E_0 e^{-x/X_0} \tag{3.8}$$

where X_0 is the radiation length of the material. The crystals in the barrel (endcap)

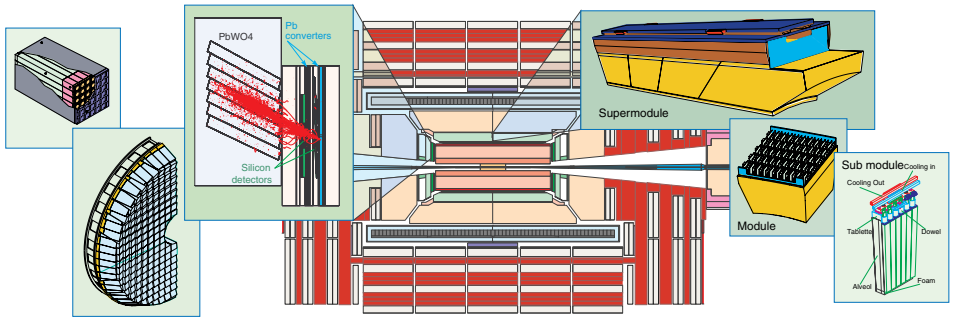


Figure 3.5: Drawing of the CMS ECAL subcomponents, illustrating the organization of the crystals and showing an example EM shower [50].

are 25.8 (24.7) X_0 long and are able to capture nearly all of the electromagnetic (EM) energy of electrons and photons.

The energy resolution of the ECAL is excellent and is given as

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{2.7\%}{\sqrt{E}}\right)^2 + \left(\frac{0.15\ GeV}{E}\right)^2 + (0.5\%)^2 \tag{3.9}$$

The first term is the "stochastic" term covering statistical fluctuations and intrinsic shower fluctuations. The second term is for electronic noise and pile-up energy, while the final term is a constant to cover detector non-uniformity and calibration uncertainty. The resolution is optimal for energies between 1 GeV and 1 TeV. A 40 GeV deposit of energy has a resolution of less than 1%. After being exposed to radiation, the crystals began to lose some of their transparency and were re-calibrated using a laser system.

3.2.4 Hadronic Calorimeter

The hadronic calorimeter (HCAL) [53] measures the energy and location of hadronic showers (jets), in this case those produced in association with a W boson. It also plays an important role in the measurement of missing transverse energy, in

this case energy from the neutrino in the W boson decay. The HCAL is a sampling calorimeter, where the barrel/endcap (HB/HE) region is made up of brass and scintillator layers and the forward region (HF) of steel plates and quartz fibers. The HB extends to $|\eta| < 1.305$, the HE from $1.305 < |\eta| < 3.0$, and the HF from $3.0 < |\eta| < 5.0$. A drawing of the HCAL and subcomponents is shown in Figure 3.6. Because hadronic activity is associated with the strong force, the particles in jets interact with nuclei. The HCAL covers 6-8 interaction lengths, where the energy of a particle drops by a factor of $1/e$ for each interaction length. The particles passing through the HB/HE (HF) interact with the brass (steel) material and are detected in the plastic scintillator (quartz fibers) layers. Brass was chosen for being non-magnetic and having a short interaction length. Steel and quartz fiber were chosen for the endcaps because they are better able to withstand the high rate of particles passing through and the larger amount of radiation in the forward region.

Light emitted from the brass tiles as particle pass through is absorbed in the scintillator fibers. The wavelength shifting fibers act as waveguides to hybrid photodiodes that collect the energy information. Cherenkov light is produced in the HF when charged particles pass through the quartz fibers.

The resolution of the HCAL is given by the equation:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{90\%}{\sqrt{E}}\right)^2 + (4.5\%)^2 \tag{3.10a}$$

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{172\%}{\sqrt{E}}\right)^2 + (9.0\%)^2 \tag{3.10b}$$

where 3.10a is for the barrel/endcap region, and 3.10b is for the HF. The first term is the "stochastic" term covering statistical fluctuations and intrinsic shower fluctuations. The second term is a constant term to cover detector non-uniformity and calibration

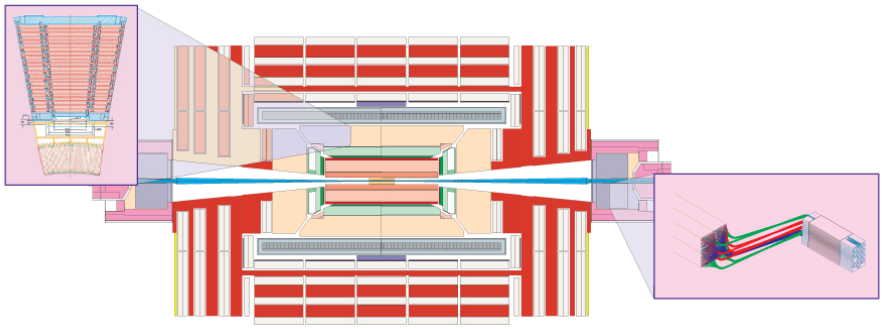


Figure 3.6: Drawing of the CMS HCAL subcomponents. On the left, the brass and scintillator of the HB/HE are detailed. To the right the steel and quartz of the HF are highlighted [50].

uncertainty. The resolution of the HCAL is inferior to the ECAL partly because it is a sampling calorimeter, but also because hadronic showers have fewer particles, thereby increasing the statistical fluctuations.

3.2.5 Muon chambers

Muons are not integral for an analysis of W bosons decaying to an electron and a neutrino (aside from requiring no muons), but are important for numerous other studies. Being relatively massive, muons have little loss of energy when passing through the calorimeters, but being charged, they do leave information in the tracker. For a more accurate measurement of their momentum, a second sub-detector, outside of the solenoid, is needed [54]. Because muons are minimum ionizing particles, it is not possible to collect all of their energy, as can be done with photons, electrons, and hadrons. Instead, the transverse momentum is measured as they pass through first the tracker and then the detecting elements in the magnetic field outside the solenoid.

To measure muons there are three types of detecting elements: cathode strip chambers (CSC), drift tubes (DT), and resistive plate chambers (RPC). The DTs

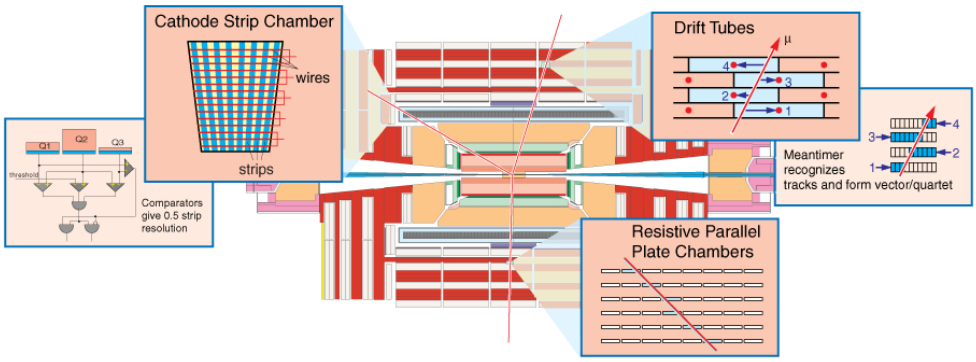


Figure 3.7: Drawing of the CMS Muon subcomponents including cathode strip chambers (CSCs), drift tubes (DTs), and resistive place chambers (RPCs) [50].

are used in the barrel region, the CSCs in the endcaps, and the RPCs in both. The magnetic field outside the solenoid is enhanced by an iron yoke, interspersed with the detecting elements. The field is about 2 T and bends the muons enough so that they leave a curved track from which the transverse momentum may be calculated. The DTs in the barrel have four concentric layers, each segmented into three for a total of 12 sectors. The spatial resolution is 100 μm in the r-ϕ plane, and 150 μm in the z direction. They have a drift time of 380 ns. CSCs are used in the endcaps because they can withstand the higher flux of particles and a less uniform magnetic field. They are in concentric rings of 18 or 36 chambers. While the CSCs and DTs both contribute to the triggering of muons, the RPCs in both the barrel and endcaps are ideal for triggering. They have a very short time resolution of about 1.5 ns and a spacial resolution better than 1 cm. Figure 3.7 gives a schematic view of the muon subsystems.

3.2.6 Trigger

As the detector collects information about the particles in the event, this information moves on to the Level 1 trigger (L1) which uses specialized hardware to decide whether to keep or throw away the event. The L1 triggered events are then sent to a software-based high level trigger (HLT), which reduces the rate to a manageable 300 Hz that can be written to disk for later analysis. Triggering allows only the most promising events to continue on for full reconstruction and analysis.

3.2.6.1 Level-1 Trigger

Given the interaction frequency of 0.5 GHz—there are about 2.2 interactions (for the 2010 run) every 25 ns, and roughly 20 interactions per bunchs—not all of the events (each about 0.2 MB) can be stored. The L1 trigger is the first step to reducing the number of events by selecting only 50-100 kHz of the most interesting events, using hardware algorithms that can make decisions in less than 3.2 μs [55].

Events are selected for electrons/photons, jets, $\not\!\!E_T$, taus, muons, and combinations of these objects. Information from the calorimeters and muon systems is sent to local hardware. In the case of calorimeter objects, the data from ECAL and HCAL are sent to the Regional Calorimeter Trigger (RCT) which identifies jet candidates and the most energetic electrons/photons (with no distinction made between them at this stage). The results of the RCT algorithm are sent to the Global Calorimeter Trigger (GCT) where the candidates are counted and sorted. The calorimeter trigger results are then combined with the muon trigger results in the Global Trigger (GT). At this point, only 100 kHz of events pass the selection algorithms and get sent to the high level trigger. A schematic view of the trigger flow is shown in Figure 3.8.

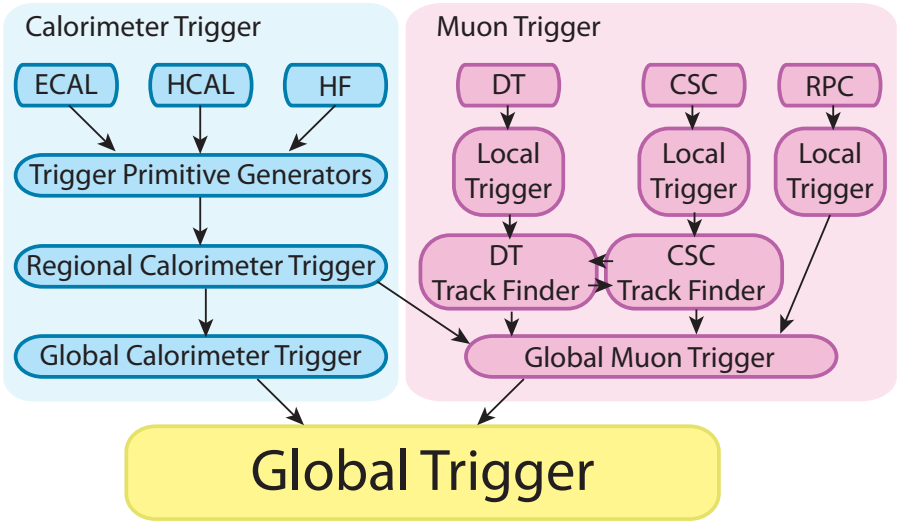


Figure 3.8: Schematic of how information passes through the various systems within the Level-1 Trigger [55].

This analysis uses exclusively single electron triggers. At L1, the electron/photon trigger uses only calorimeter information to make a decision about keeping the event. Groups of 5×5 calorimeter crystals (in the barrel) correspond to single trigger towers. The Trigger Primitive Generator (TPG) sums the E_{T} of the 25 crystals to get the E_{T} for the trigger tower, and associates it to the correct bunch crossing. The ECAL uses selective readout (SR), such that towers (and their eight surrounding towers) are only read out if they register E_{T} above a 1 GeV threshold.

The TPG sends 8 bits for E_{T} and a fine grain bit (see Section 3.2.6.2) from each tower to the RCT. There are 4176 trigger towers binned as $72\phi \times 56\eta$ in HB/HE, and $18\phi \times 4\eta \times 2$ in HF. Each EB trigger tower has a size of $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$. The EE towers have roughly a similar size up to $\eta < 2$, at which point the width in η increases with η. The HCAL follows the same dimensions as the ECAL up to $\eta < 1.74$, at which point the towers become twice as large in ϕ. The HF, not used

in electron/photon triggering at the L1, has larger ϕ binning and is used for jet and $\not\!\!E_T$ triggering. The TPG information from each tower is sent to the RCT where the electron/photon is classified as isolated or non-isolated. The eight most energetic electron/photon candidates, four isolated and four non-isolated, for $|\eta| < 2.5$, are sent to the GCT. The RCT also calculates the four most energetic jets, for $|\eta| < 3$, and four for $3 < |\eta| < 5$. See Section 3.2.6.2 for details on the candidate creation/selection algorithm used in the RCT. The GCT then forms E_x and E_y using lookup tables, sorts the electrons, jets, and taus, counts the number of jets, and sends the top four calorimeter candidates, along with the $\not\!\!E_T$, to the global trigger. The GT combines the muon and calorimeter information. Different L1 paths for electrons are used, depending on the desired E_{T} threshold and multiplicity [55].

3.2.6.2 Regional Calorimeter Trigger

The Regional Calorimeter Trigger (RCT) [56] is responsible for the initial interpretation of data coming from the calorimeters. It is the first stage for constructing electron/photon and jet trigger objects. The RCT is divided into eighteen crates, covering the region $|\eta| < 5$. Each crate has seven cards, with two regions per card, for a total of 14 regions per crate. Each region is a 4x4 group of trigger towers. The RCT receives energy information for ECAL, HCAL, and HF from the TPG for each trigger tower. A "fine grain" (FG) bit is sent with the E_{T} from the ECAL that gives information on the lateral spread of the EM shower. It is active when the highest energy strip pair of crystals in the 5x5 tower has too small of a fraction of the tower energy. The fraction used during the 2010 running is 90%. Electrons are expected to be very narrow in η, so if the FG bit is set, the triggering object is less likely to be

a true (isolated) electron. A second test is done to check for hadronic energy in the HCAL tower corresponding to an ECAL tower. If the ratio of the HCAL tower E_{T} to the ECAL tower E_{T} (H/E) is greater than a given threshold (currently set to 5%), then the H/E veto is set to true.

The RCT algorithm, diagrammed in Figure 3.9, shows the strips of crystals going into a tower and the 3x3 window searching for the highest ranked towers. Added to the highest ranked tower is the rank of the highest adjacent tower. Each candidate is assessed for isolation or non-isolation. If the highest ranked tower or any of its eight neighbors has the FG bit or the H/E veto set, the candidate is non-isolated. In addition, the candidate is non-isolated if none of the four corner "L"s around the main tower is quiet ($E_{\mathrm{T}} < 1.5$ GeV).

For each 4x4 region, the highest ranked isolated and non-isolated candidates are chosen. Then, within each of the 18 crates, the highest four isolated and highest four non-isolated electron/photon candidates of the 14 regions are sent to the GCT. The jet part of the RCT algorithm sums the total energy in each region and sends the E_{T} of the highest four regions to the GCT.

3.2.6.3 High Level Trigger

The high level trigger (HLT) [57, 58] uses software algorithms to further filter out unwanted events (mostly QCD), and to pass along potentially interesting interactions for archiving and subsequent offline processing. It reduces the overall rate to about 300 Hz, a rate that is possible to store. The HLT aims to select events as quickly and efficiently as possible. Unlike in many experiments, there is no distinction between Level-2 and Level-3 triggers, since everything is processed with a single processor farm.

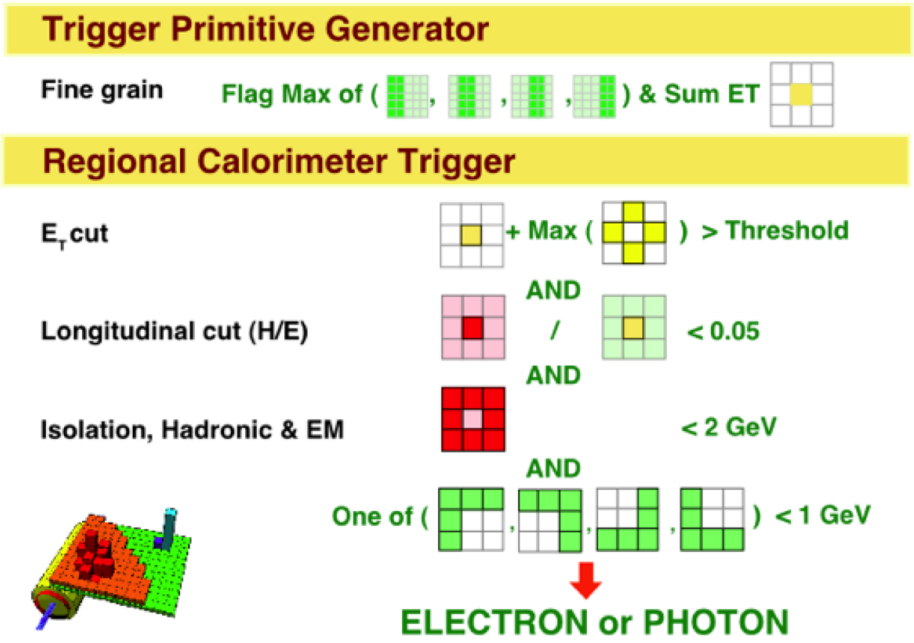


Figure 3.9: Diagram of the Regional Calorimeter Trigger algorithm decisions for isolated vs. non-isolated electron/photons.

However, events are still rejected in several steps to reduce the overall CPU usage. The Level-1 trigger results are used to decide what needs to be reconstructed. For instance, if the L1 indicates only a possible electron/photon, then the HLT does not bother checking the muon system.

For the electron HLT paths, an energy deposit is identified in the ECAL first, and only if it passes the desired energy threshold requirement is a track reconstructed from the tracker information. The ECAL deposit and track are then matched, and if they pass a loose selection, the event is saved. More detailed information on reconstruction of the HLT objects is given in Chapter 5. Because the luminosity changed quickly in the first months of data taking, the HLT requirements also changed quickly: the more luminosity, the more events that need to be filtered out, the tighter the selection needs

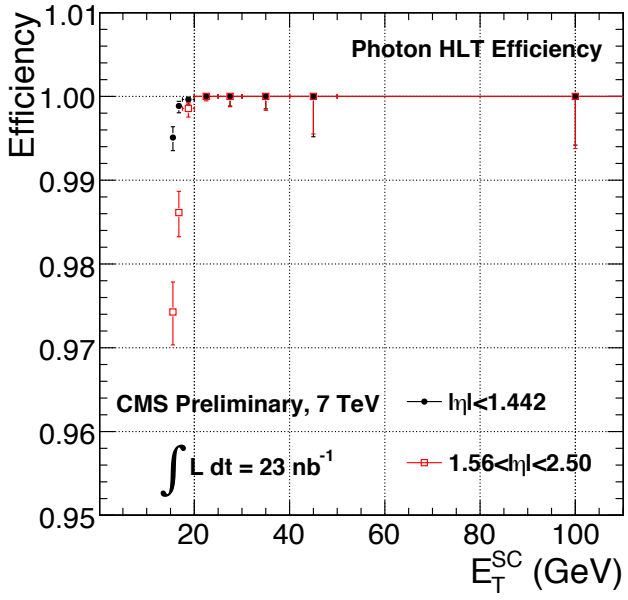


Figure 3.10: The efficiency of the HLT_Photon15 path for a reconstructed electron matched to an L1_SingleEG5 candidate as a function of E_T [59].

to be in order to keep the rate reasonable. Table 6.1 in Chapter 6 lists the runs, the triggers used, and the specific selection criteria for each run. The electron efficiency turn-on curve for the HLT in early runs is given in Figure 3.10, showing nearly 100% efficiency starting at 20 GeV.

Chapter 4

Event Simulation

To model high energy processes of interest accurately, physicists rely heavily on computer simulations. Simulations allow physicists to perform non-analytical calculations by weighting random numbers based on the theoretical predictions behind the model being reproduced. The method is known as Monte Carlo (MC) [60], after the city known for gambling, because of the randomness used when producing the millions of events needed to study a single process. MC simulations are useful for comparing data to the current theory, to derive calibration corrections, and to understand systematic uncertainties in measurements derived from data.

4.1 Monte Carlo Event Generation

Event generation through simulation is done in several steps, occasionally using different computer programs for various steps [61]. The basic flow occurs as follows. Input information, such as colliding particle types, in this case proton-proton and center of mass energy, $\sqrt{s} = 7$ TeV, are chosen to match the operating parameters of the collider. Parton distributions based on Standard Model calculations are included for initial particles, modeling the substructure in terms of flavors and energy sharing

using the parameters measured by previous experiments. One parton from each proton interacts with another parton from the oncoming proton in a "hard interaction" where new particles, such as a W boson and multiple partons, are produced with substantial $p_{\mathrm{T}} \gg \Lambda_{\mathrm{QCD}}$. The colored remnants of the initial-state protons undergo "soft interactions" and multiply into a shower of additional "soft" partons with low transverse momentum $p_{\mathrm{T}} \approx \Lambda_{\mathrm{QCD}}$. The W boson in this analysis is selected to decay to an electron and neutrino. The outgoing partons shower into more partons. The individual quarks and gluons, behaving according to the QCD confinement mechanism, then "fragment" into colorless hadrons. The unstable hadrons decay in a cascade until only stable particles remain. The $\sqrt{s} = 7$ TeV collisions typically result in ∼500 particles with p_{T} greater than 1 GeV traversing the detector.

4.1.1 PDF and Hard Scatter

The processes of interest, and the theory behind them, have been discussed in Chapter 2. Parton distribution functions (PDFs) are used to model the composition of the protons in a collision. PDFs are obtained from earlier experiments and are extrapolated for use at the LHC. See Section 2.1 and reference [28]. They give the probability of finding a parton with a momentum fraction x in a proton. The interaction between two partons, one from each proton, produces the hard scattering of interest. The matrix element calculations corresponding to the Feynman diagrams shown in Figure 1.8, are used to generate events within the phase space. Some of the cross sections used to scale the number of simulation events are taken separately from the MC generator that produced the events, in order to include higher order effects.

4.1.2 Parton Showers and Hadronization

The basic hard scattering must be expanded upon as it does not account for other physical effects before and after the interaction. There are bremsstrahlung-type initial and final state radiations, both of which are added into the Monte Carlo simulation. These do not depend heavily on the particular process being simulated, and more on the momentum transfer scale of the process. As the p_{T} of the showering partons gets smaller, the perturbative QCD calculations become unreliable. Therefore, approximate models of parton showers, e.g., the event generators SHERPA[62], HERWIG[63], PYTHIA [64], are used to form parton showers. Algorithms such as CKKW [65] and MLM [66] are used to match the matrix element to the parton shower to avoid over or undercounting jets relative to the original number of partons.

The hadronization is done using phenomenological models as it cannot be described from first principles. For example, PYTHIA uses the "Lund string model" [67, 68] to simulate the fragmentation of hadrons. Quarks are located at the end of strings that are collections of field lines from the gluons which are much narrower than electromagnetic fields because the gluon couples to itself (whereas the photon does not). Strings break for the production of new quark/anti-quark pairs, and the new quarks can form into mesons with quarks from adjacent strings.

In every event there are other partons from the colliding protons that do not participate in the hard interaction but cannot be ignored. These are considered to be the "underlying event". In MC generators the extra hadron multiplicities and the soft interactions are simulated based on phenomenological models.

In addition to extra partons within the interacting proton, there are other pro-

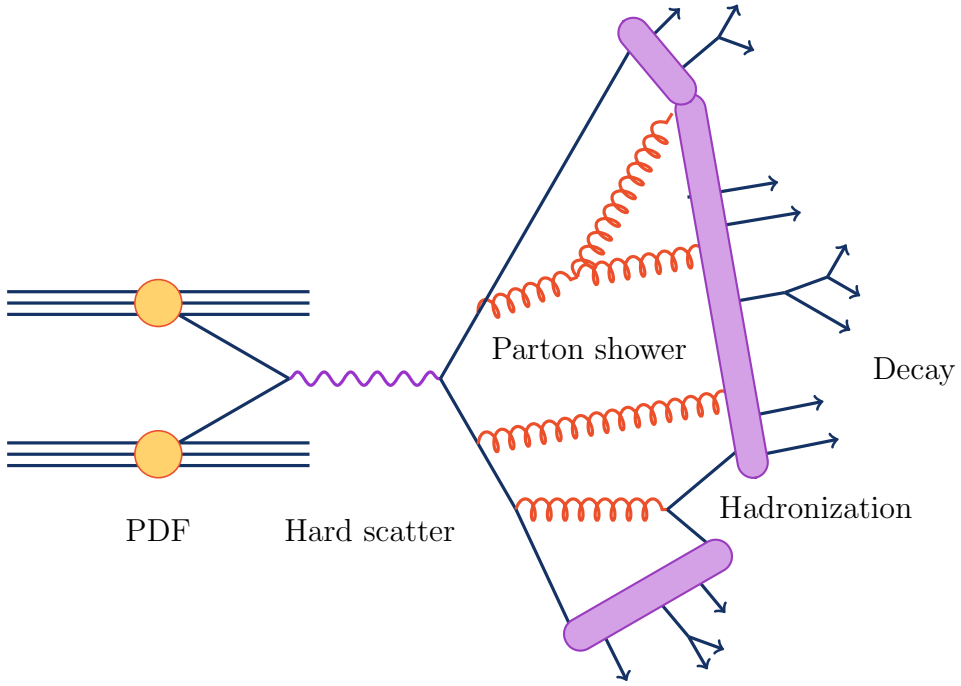


Figure 4.1: A pictorial representation of a collision with the hard interaction and the resulting fragmentation, hadronization, and decay.

tons in a bunch crossing that may also interact, known as "pile-up". In an effort to simulate pile-up in the generated events, minimum-bias events are superimposed on the samples roughly according to the distribution of multiple proton-proton collisions observed during the 2010 data taking. For the majority of data taken there were about 2.2 collisions for each triggered bunch crossing.

4.1.3 Monte Carlo Software

Monte Carlo programs can be divided into two types. Some, such as PYTHIA [64], HERWIG[63] and SHERPA[62], include all the steps in event generation from the PDF to the final state particles. The hard interaction portion of the cross section calculation for 2→2 processes is based on perturbative QCD to leading order (LO).

However, these are not sufficient for a study of W bosons produced with jets. Next-to-leading order (NLO) calculations and those including multiple final state partons are necessary for an accurate modeling of the W+jets data. See Figure 1.8, The $2 \rightarrow n$ processes are calculated under some simplifying assumptions. To model the W+Jets signal, the second type of MC generator is used. This type of generator, such as MadGraph [35] or AlpGen [69], focuses on the more complex hard matrix elements and can produce multiple parton final states. These generators are more reliable for producing events with multiple jets with the correct kinematics and cross sections. They usually do not include the hadronization step and therefore are often interfaced with the other, more complete generators to get a full picture of the event.

The simulated signal events in this analysis were generated with MadGraph [35] interfaced with pythia [64]. MadGraph produces the parton level event—PDF and hard scattering from a matrix-element calculation—for a W boson and up to four jets. pythia handles the parton showering and hadronization. For the backgrounds, MadGraph + pythia is used for Z+jets and top quark samples, while only pythia is used for the γ+jets and QCD samples. The pythia simulation uses parameters from the "Z2" tune [70], a modification of the "Z1" tune [71], to better describe the soft particles from the underlying event and pileup. Simulations of the signal using only pythia or MadGraph + pythia with the "D6T" tune [72] are employed for additional comparisons and systematic studies. The parton distribution function model used for MadGraph is CTEQ5L [73]. The factorization and renormalization scales are $\mu_F = \mu_R = M_W$.

When possible, the normalization of the MC samples comes from next-to-leading

order (NLO) or next-to-next-to-leading order (NNLO) [74] calculations using the Fully Exclusive W and Z Production (FEWZ) program [75] or the Monte Carlo for FeMtobarn processes (MCFM) program [76]. MCFM is a specially designed code for calculating cross sections in hadron-hadron collisions at the femtobarn level. The full set of Monte Carlo samples, the kinematic cuts, and their cross sections (or MCFM predicted (N)NLO cross-sections), are given in Table 4.1. All of the samples have pile-up simulations included which correspond to the latest data taking of the 2010 run [77, 78].

4.2 Detector Simulation

After the basic scattering has been modeled with MC simulations, the resulting particles are put through a simulation of the detector. The characterization of the CMS detector is done with GEANT4, a toolkit [79] for the simulation of the interactions of particles in detector matter. It accounts for the behavior of various types of particles passing through the different types of matter from which the detector is made. The simulation models the materials, the magnetic fields, and the specific geometry of the CMS detector. The information of the path and energy of the simulated particles is output in a form similar to the data acquired from the real detector so that it can be reconstructed in the CMS software. The Monte Carlo has the original particle information from the hard scattering stored so that it can be used to calibrate measured quantities to the generated values [80].

Generator	Process	Kinematic cuts	σ (pb)	
MadGraph (Z2)	$W \to \ell\nu$ + jets	None	31314	NNLO
MadGraph	$Z \to \ell\ell$ + jets	$M_{ll} > 50$ GeV	3048	NNLO
MadGraph	$t\bar{t}$ + jets (tauola)	None	157	NLO
MadGraph	t(s - chan)	None	1.4	NLO
MadGraph	t(t - chan)	None	20.9	NLO
MadGraph	tW	None	10.6	NLO
pythia	EM enriched QCD	$20 < \hat{p}_T < 30$ GeV	1719150	LO
pythia	EM enriched QCD	$30 < \hat{p}_T < 80$ GeV	3498700	LO
pythia	EM enriched QCD	$80 < \hat{p}_T < 170$ GeV	134088	LO
pythia	BCtoE QCD	$20 < \hat{p}_T < 30$ GeV	108330	LO
pythia	BCtoE QCD	$30 < \hat{p}_T < 80$ GeV	138762	LO
pythia	BCtoE QCD	$80 < \hat{p}_T < 170$ GeV	9422	LO
pythia	γ + jets	$15 < \hat{p}_T < 30$ GeV	171700	LO
pythia	γ + jets	$30 < \hat{p}_T < 50$ GeV	16690	LO
pythia	γ + jets	$50 < \hat{p}_T < 80$ GeV	2722	LO
Additional W+Jets MCs used for testing unfolding				
pythia	$W \to \ell\nu$	$\lvert\eta_e\rvert < 2.7$	8159	LO
MadGraph (D6T)	$W \to \ell\nu$ + jets	None	31314	NNLO

Table 4.1: List of MC signal and background samples

Chapter 5

Event Reconstruction

After collecting data from the various subsystems of the CMS detector (or simulating events using Monte Carlo generators and detector simulation), the information must be synthesized into usable objects. In particular, for W+jets we are interested in reconstructing the electron and measuring the $\not\!\!E_T$ due to the neutrino from W decay and also in reconstructing jets from the associated partons (as described in Chapter 2).

5.1 HLT objects

The high level trigger [55] uses a specialized version of the CMS reconstruction software to select interesting events during collisions and reduces data to a manageable 300 Hz. It has time to process more information than the Level 1, including "hits" from the tracker. Events are selected using several different HLT "paths", seeded by a L1 single electron/photon trigger with an E_{T} threshold of either 5 or 8 GeV (L1_SingleEG5 or L1_SingleEG8) [59]. Starting from the region specified by the L1 seed, electrons are reconstructed from ECAL energy deposits, including energy dispersed from bremsstrahlung in the nearby crystals. To distinguish electrons from photons, two or three pixel hits must be found within $\Delta\phi < 40$ mrad when

extrapolating from the cluster of ECAL hits to the nominal vertex. Finally, full track reconstruction using the Kalman Filter formalism [81] is performed starting from the pixel hits. Requirements are put on the ratio of energy to momentum: E/p < 1.5 in the barrel and E/p < 2.45 in the endcaps [82] since the energy from the calorimeter and the momentum from the tracker are more likely to be similar for true electrons than for fake electrons or those from jets. It also removes calorimeter noise spikes. Additionally, a requirement is made that the fraction of energy in the HCAL relative to the ECAL be less than 15% in order to reject electrons from jets that are more likely to leave energy deposits in the HCAL. The E_{T} of the final reconstructed electron must then be above the threshold designated in the HLT path name, again in order to remove fake electrons. For the earliest runs, when the electron HLT paths were still being commissioned, the photon path without track information was used [83].

5.2 Particle Flow Overview

Jets, missing transverse energy (for the neutrino), and electrons are all reconstructed using the "Particle Flow" (PFlow) algorithm [84]. The algorithm tries to create a complete event description by collecting information from all of the sub-detectors (tracker, calorimeters, muon chambers). Clusters of related information are formed in each of the sub-detectors before being linked to other sub-detectors and becoming reconstructed particles. Every indication of activity in the detector that is above a minimum noise threshold is assigned to one of the particles. Particles are initially formed into the categories of muons, electrons, photons, charged hadrons, and neutral hadrons. From these elements composite objects, such as jets and missing

energy (neutrino), are also reconstructed.

5.2.1 Track and Vertex

The reconstruction of both electrons and jets are improved over calorimeter-only objects through the use of the tracker sub-detector (Section 3.2.2). The high magnetic field and large volume of silicon in the tracker allows for separation of photons, which leave no tracker information, from charged particle deposits, which ionize in the tracker. Given its excellent resolution, tracking is a key component of the PFlow algorithm. The tracker adds valuable location, direction, and momentum information.

Tracks are initially seeded with pairs or triplets of track hits in the pixel and strips. The hits are propagated outward using pattern recognition to find new hits. All of these hits are then combined to form a full track and are fitted to obtain the track parameters, such as η and ϕ locations and p_{T}. Iterations of the algorithm are performed, after removing hits already assigned to a track, to gather information for all of the tracks in the event. Tight seeding criteria, such as distance to the beam spot and Δz to the HLT primary vertex, are initially used to ensure a low fake-rate. The cut values are dependent on p_{T} and are relaxed as the number of hits increases. The seeding criteria are loosened in a second iteration to increase the efficiency to $> 90\%$ for charged hadrons. Further iterations reduce constraints on the vertex to include particles from photon conversions and interactions in the material of the tracker [81].

5.2.2 Calorimeters

Combining the hadronic information from the HCAL with the higher resolution deposit in the ECAL allows for a hadron energy resolution of $\sim 10\%$ at 100 GeV.

As described in Sections 3.2.3 and 3.2.4, electron and photon energy is deposited in the electromagnetic calorimeter (ECAL), while hadronic energy is deposited in both the ECAL and the hadronic calorimeter (HCAL). Calorimeters are useful for measuring the energy and direction of both neutral and charged particles. They can also be used with the tracker to separate the neutral from the charged hadrons and are excellent at identifying electrons. PFlow specific cluster algorithms are used separately in each sub-detector of the calorimeters (ECAL barrel, ECAL endcap, HCAL barrel, HCAL endcap, first and second pre-shower (PS) layers). The algorithm begins with "cluster seeds", local calorimeter-cell energy maxima above a minimum threshold to avoid noise. "Topological clusters" are formed by adding more cells, again above a threshold, near the initial cluster. Cell thresholds are two standard deviations from the electronics noise level. Each seed eventually becomes a "particle-flow cluster" [85].

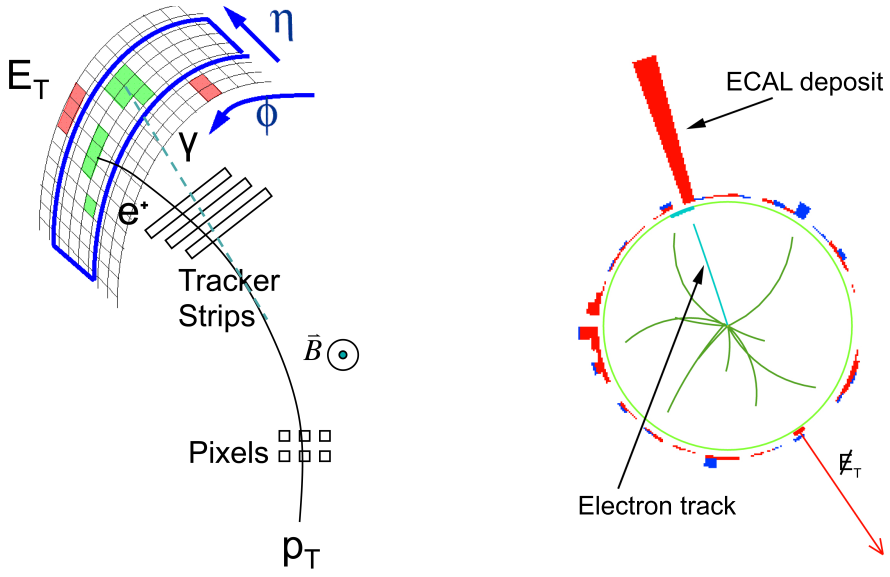
5.2.3 Object reconstruction

Particles passing through the CMS detector can leave information in multiple sub-detectors, and the strength of the PF algorithm is that it combines together all of the elements into an accurate description of each particle. The "link algorithm" begins constructing particles from the detector elements while also avoiding possible double counting of tracker or calorimeter deposits. Distances between all pairs of elements are quantified for their quality and then the elements are formed into "blocks". To link a track to a calorimeter cluster, possibly as part of electron or jet reconstruction, several steps are used. The last hit in the tracker is extrapolated first to the pre-shower, then to the ECAL up to the typical electron shower depth, and finally to the typical hadron shower depth in the HCAL ($\sim$ one interaction length). The track is

linked to clusters within the extrapolated position. The link distance is then defined as the distance between the extrapolated track and the cluster. Tangents to the track are also extrapolated into the ECAL to account for bremsstrahlung from electrons. Clusters that fall in this tangent are also linked to the track as possible radiated photons. Links are made between ECAL and HCAL clusters and between the PS and ECAL when a cluster in the more granular calorimeter falls within a cluster of the less granular calorimeter. Muons are also included in the linking algorithm, combining tracker tracks and muon tracks to form a "global muon".

5.3 Electron Reconstruction

For high energy electrons the particle flow algorithm uses essentially the standard electron reconstruction algorithm [86]. This is because the algorithm is fully commissioned and has associated identification variables with good performance for selecting electrons from W bosons decays (discussed in Chapter 6). The ECAL-driven algorithm used for high energy electrons starts from an ECAL cluster with $E_{\mathrm{T}} > 4$ GeV and uses the property that the center-of-mass of the deposits does not depend on the bremsstrahlung. The hits in the pixels and strips are inferred inward and tracks can be built from the hits [85]. The electron-like tracks are fit using a Gaussian-Sum Filter [87] to follow them into the ECAL and which also accounts for kinks in the trajectory after emitted photons. Because electrons radiate photons as they curve in the tracker, the associated ECAL cluster needs to be wider in ϕ to include the emitted photon energy as part of the original electron energy. All of the energy in a strip in ϕ surrounding the local maximum energy is gathered from adjacent clusters into a "supercluster" [44]. The path in the tracker and the hits in the ECAL are



(a) (b)

Figure 5.1: (a) Drawing of an electron passing through the tracker and creating hits in the ECAL crystals, including a bremsstrahlung photon, γ. (b) Event display in the x-y plane of an electron with a high momentum track and a large ECAL deposit [88].

shown in Figure 5.1.

There is also a requirement that the ratio between the HCAL and ECAL energy deposits is smaller than 0.15 to reduce the rate of fake seeds. Loose requirements of $|\Delta\phi| < 0.15$ and $|\Delta\eta| < 0.02$, the distance between the supercluster and track in the ϕ and η directions, are applied as a preselection, with tighter cuts applied during the full selection.

The electron momentum is computed as the weighted mean of the supercluster energy and the track momentum when $|E/p - 1| < 2.5\ \sigma_{E/p}$, where σ is the error on the track and ECAL measurements. Otherwise, the ECAL information is used. The track momentum is generally more favored at lower p_{T}, while ECAL E_{T} is preferred

at higher energies [89].

5.4 Hadron and Photon Reconstruction

With electrons accounted for, jet reconstruction can begin. Tracks for particles found in jets have tighter quality criteria applied. In particular, the relative uncertainty on the p_{T} must be smaller than the relative calorimetric energy resolution for charged hadrons. Photons and neutral hadrons are identified by comparing the track momentum to the energy detected in the calorimeters. (The energy in the calorimeters is first calibrated for a reliable comparison.) If several tracks are linked to an HCAL cluster, their momentum are summed and then compared to the calorimeter energy. If one track is linked to several HCAL clusters, only the closest cluster is used for the comparison. Links of tracks to the closest ECAL clusters are also kept, and the other ECAL clusters linked to a given track are ordered by distance to nearest track. The link is kept only if the total calorimetric energy is smaller than the total charged particle momentum. This process ignores links from overlapping photons but keeps clusters that are from hadronic shower fluctuations. If the calorimeter cluster energy linked to a track(s) is much larger than the track momentum and the excess is larger than the calorimeter energy resolution, then a photon and/or neutral hadron is created. The excess is identified as a photon and neutral hadron if not all of the excess is in the ECAL. If the excess is covered by the ECAL energy alone, the excess only gives rise to a photon (which makes up a much larger percentage of jet energy). Any calorimeter clusters not linked to a track, or for which the link was not kept, are used as photons and neutral hadrons, respectively. Remaining tracks in a block are used for charged hadrons with the momentum and energy taken directly from the track

momenta. The momenta are redefined using a fit to the corresponding calorimeter energy. This process is most useful when track information has a lower resolution, such as at very high energies or large η (maximum η for the tracker is 2.5).

5.5 Jet Reconstruction

Partons from a hard interaction split into many other particles as they leave the interaction point and pass through the detector, as explained in Section 1.3. Constructing the many individual particles, presumably originating from a single parton, into entities known as "jets" allows for easier analysis and comparison to theory. Jets used here are reconstructed from particle flow objects, described above, rather than from detector elements (as in most other algorithms). By using pre-identified detector particles, reconstructed jets more closely resemble jets reconstructed from MC-truth particles. Roughly 90% of the jet energy can be reconstructed with precision while 10% is affected by the poor HCAL resolution and calibration corrections. Only jets within the acceptance of the tracker $|\eta| < 2.4$ are kept in order to make full use of the PFlow algorithm. Jets are created by means of a clustering algorithm known as "anti-k_T" [90] which is infrared and collinear (IRC) safe. QCD calculations at the parton level have divergences, such as when a parton emits a soft gluon, or a parton divides into two collinear partons, or a parton is emitted such that is carries away only longitudinal momentum. A jet algorithm is infrared safe if the jet reconstruction does not change with the presence of infinitely soft particles. The algorithm is collinear safe if it automatically recombines partons that divide into collinear partons [91]. The IRC safe property indicates that the jets can be compared to theoretical computations of any perturbative order.

The anti-k_T algorithm clusters particles into jets according to proximity in both momentum space and physical distance. It starts with the hardest momentum particle and first clusters it with "nearby" particles that have smaller relative momentum. Distances d_{ij} between two entities (either particles or pseudojets), i and j, are calculated, along with the distance d_{iB} between a particle i and the beam, B. The two distances are compared and if $d_{i,B}$ is smaller, the entity i is declared a jet and is removed from the list of entities. If d_{ij} is smaller, i and j are combined into a pseudojet, weighting the combined position by the momenta. The procedure is repeated until all entities have been removed. The distances are defined

$$d_{ij} = min(k_{t,i}^{2p}, k_{t,j}^{2p}) \frac{\Delta_{ij}^2}{R^2} \tag{5.1a}$$

$$d_{iB} = k_{t,i}^{2p} \tag{5.1b}$$

where $\Delta_{ij} = (\Delta y_{ij}^2 + \Delta\phi_{ij}^2)$ and $k_{t,i}^{2p}$, y_i, and ϕ_i are the transverse momentum, rapidity and azimuth of particle i. R is the radius of the cone and p is a parameter added concerning the relative power of the energy vs spatial scales. A p value of 1 gives the inclusive "k_T" algorithm and $p = 0$ gives the "Cambridge/Aachen" algorithm [90]. In the anti-k_T algorithm p is set to -1, meaning that it starts with higher p_{T} particles and adds to it the softer particles.

The effect of this algorithm is that softer momentum particles tend to cluster with harder ones before clustering with other soft particles. A hard particle with no nearby neighbors will have all the soft particles within R added to it resulting in a conical jet. If another hard particle is between $R < \Delta_{ij} < 2R$, the two hard particles will split the surrounding softer particles, weighted according to their relative transverse momenta, resulting in only partly conical jets. Two hard particles within $\Delta_{ij} < R$

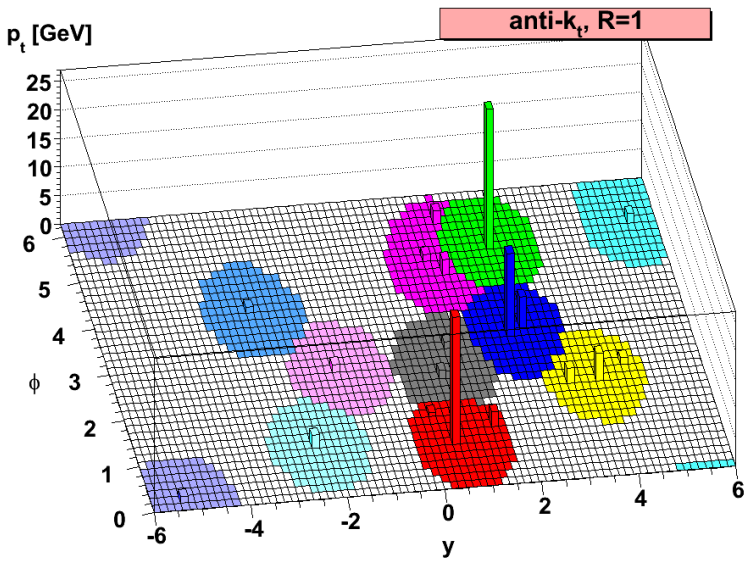


Figure 5.2: The results of applying the anti-k_T algorithm to a generated parton-level event with additional soft radiation. The colored areas indicate the areas defined by each jet. Harder jets take up a larger portion of the surrounding area than nearby softer jets [90].

will combine into one jet, the center weighted by their relative momenta to the harder particle and the shape formed from the union of the two cones around the particles plus the cone around the final jet. One strength of this reconstruction method is that softer particles do not modify the shape of the jet, but it is still flexible for harder radiation. An example of the anti-k_T algorithm applied to a parton-level event with added soft radiation is shown in Figure 5.2. The jets used in this analysis have a cone size R of 0.5.

The anti-k_T algorithm is less susceptible to underlying event (UE) and pile-up events than other clustering algorithms. The UE and pile-up can add to the energy of each jet as well as change the distribution of particles among the jets. The suppression of additional energy on the anti-k_T algorithm is proportional to the

jet p_{T} and has a only small effect on reasonably high energy jets. This suppression leads to less smearing of the jet momenta. Smearing eventually needs to be corrected before comparing reconstructed jets to generator level jets. The anti-k_T algorithm takes noticeably more time to compute than similar algorithms, $N^{2/3}$ vs N ln N, where N is the number of iterations. Because computing resources are finite, special FastJet [92, 93] calculations are used to maintain a clustering that is as rapid as other algorithms such as k_T and iterative cone, and can remove pile-up energy efficiently.

Energy scale corrections are necessary to adjust for the non-uniform and non-linear response of the calorimeters, electronics noise, and pile-up interactions. The corrections attempt to adjust the energy scale of the measured and reconstructed jets closer to the original particle jet energies. The corrections are done in steps. The offset correction (level 1, L1) attempts to reduce extra energy from noise and pile-up events. The relative corrections (level 2, L2) rectify the variations in η relative to a central reference region. The absolute correction (level 3, L3) removes variations in p_{T}. The full sequence is

$$E_{Corrected} = (E_{Uncorrected} - E_{offset}) \times C_{Rel}(\eta, p''_{\mathrm{T}}) \times C_{Abs}(p'_{\mathrm{T}}) \tag{5.2}$$

where p''_{T} is the transverse momentum of the jet corrected for offset and $p'_{\mathrm{T}} = p''_{\mathrm{T}} \times C_{Rel}(\eta, p''_{\mathrm{T}})$ is the transverse momentum of the jet corrected for offset and pseudorapidity dependence [94].

The correction factors are obtained using MC truth information or using data driven techniques for in-situ calibration from known processes. MC corrections are useful until a sufficiently large data sample is available for in-situ jet calibration. Energy scale corrections on the order of 5% in the barrel and 10% in the endcaps were

applied for PFlow jets [95].

Electrons included in the list of jets are "cleaned" of the electrons from the W decay as the electrons are considered separately. The cleaning is done with a simple ΔR comparison of the electron position with respect to each jet. If $\Delta R < 0.5$, then the "jet" is really an electron and not included in the jet counting or kinematics.

5.5.1 B jets tagging

Identification of jets as coming from b-quarks is important when dealing with the irreducible top background. In $t\bar{t}$ production one top may decay to $W \rightarrow e\nu + b$ jet while the other top decays to $W \rightarrow jj + b$ jet, leading to 4 jets, two of which are from b-quarks. Hadrons composed of b-quarks travel a measurable distance before decaying into lighter particles and this algorithm exploits their long lifetime. A jet is identified as a b-jet if at least N of its tracks are within a significance, S, of the impact parameter exceeding $S = \mathrm{IP}/\sigma_{IP}$. The impact parameter (IP) is the distance between the track and the vertex at the point of closest approach. Because the uncertainty, σ_{IP}, for this quantity varies with the number of measurements along the track, the significance is the preferred discriminator [96]. The tagging algorithm returns the numerical discriminator S for the Nth track on which a cut may be applied. For high efficiency we have chosen to use N=2 tracks so that the discriminator S is known as "track counting high efficiency" (TCHE). Performance of the algorithm for various cuts on S has been studied and a mis-tag rate and tag-rate are established for a given S for jets above a p_{T} threshold [97]. These rates are included in the probability distribution function used to distinguish W+jets from top events in Section 7.2.1.

5.6 Missing Transverse Energy and M_{T} Construction

Neutrinos—colorless, neutral, and nearly massless—interact so seldom that they pass through the detector without registering their presence in any of the subsystems. The only way to detect their contribution to an interaction is to add up the energy in the plane transverse to the beam direction. Any energy that does not add up to zero (the initial transverse energy of the p-p collision) is attributed to the neutrino. Because only the transverse energy can be accounted for, longitudinal information is lost for the neutrino. In the PFlow algorithm [84] used here, the missing energy is designated as the opposite of the sum of the transverse momentum of all of the reconstructed particles. The missing transverse energy, or $\not\!\!E_T$, is sensitive to the underlying event and calorimeter noise. The resolution of the $\not\!\!E_T$ is lower than that of the electron, but the use of calorimeter noise cleaning and the PFlow algorithm improves the resolution as much as possible. The PFlow based $\not\!\!E_T$ reconstructs about twice as much of the true $\not\!\!E_T$ as the calorimeter-based $\not\!\!E_T$ calculations [98].

Because some information is lost from the neutrino, it is not possible to fully reconstruct the mass of the W boson. Instead we use the transverse mass defined in Equation 5.3, which only requires knowledge of the transverse information from the two decay particles.

$$M_{\mathrm{T}} = \sqrt{2 \times \not\!\!E_T \times P_T(e)(1 - cos(\theta))} \tag{5.3}$$

where $cos(\theta)$ is the angle between the electron and the missing transverse energy. M_{T} is used, along with $n_{jet}^{b-tagged}$, in the fitting method to determine the number of signal events, described in Section 7.2.1. It does an excellent job at discriminating between signal, with large $\not\!\!E_T$ from the neutrino, and QCD, with very little natural $\not\!\!E_T$.

Chapter 6

Event Selection

After the reconstruction of the physics objects, it is necessary to apply a selection in both online and offline settings. By selecting certain criteria, events are reduced to those potentially interesting for a W+jets study, while preserving control samples for estimating backgrounds. The background processes that are considered, and that the selection attempts to reduce, are outlined below.

- QCD multi-jet: The largest background, but also the most susceptible to selection cuts, as the reconstructed "electrons" in this sample tend to have low energy or are fakes. Monte Carlo simulations are not reliable for QCD and thus the shape and yield are estimated using data driven techniques.

- Photon+jets: Some background comes from $pp \rightarrow \gamma + jets$ processes when photons convert to electrons in the detector material.

- Z+jets: If one of the two electrons from the Z decay is not reconstructed or outside the fiducial region, then Z+jets events can resemble W+jets. This background is small, especially after rejecting di-electrons, because of the smaller cross section.

- $W \rightarrow \tau\nu$+jets: Small background, mostly from the τ decaying to an electron and neutrino. The electrons from τs tend to be less energetic than electrons directly from the W boson.

- Top quark: Includes $t\bar{t}$ and single top processes. Only becomes significant at 2 or more jets.

The data sample is selected from events collected during the 2010 LHC running period and corresponds to a luminosity of 36.1 $\pm$ 1.4 pb^{-1} [99]. Only runs with good beam and detector conditions were considered. All subsystems, including the trigger, need to be operational during the run and performing according to specifications. The detector control system (DCS) must display a state ready for running and the data quality monitoring (DQM) must indicate that online checks match expectations.

Events must first be selected during data taking using a trigger in hardware and then in speed-optimized software, where processing can keep up with the rate of collisions. Only about 0.00025% of collisions can be written to disk with the trigger filtering out all but the most promising. At CMS, the online selection is done in two stages, first with the Level-1 Trigger (L1, hardware) and then with the High Level Trigger (HLT, software), both described in Section 3.2.6 and Chapter 5.

For this analysis, the L1 trigger is set to select on a single electromagnetic object with p_{T} of at least 8 GeV. This corresponds to the L1_SingleEG8 path, although the L1_SingleEG5 path ($p_{\mathrm{T}} > 5$GeV) was used in the earliest runs (run number ≤ 137028). The L1 simply selects on the presence of at least one electromagnetic object, based on ECAL trigger towers, above the chosen E_T threshold.

As luminosity increased, the HLT selection needed to be tightened to keep the data taking rate at a manageable level. Initially the p_{T} threshold increased from 10 GeV to 15 GeV, and then, for the majority of the 2010 run, to 17 GeV. The full list of HLT paths and run ranges are given in Table 6.1. Note that starting from run 141956 some identification and isolation selection is included in the HLT to keep the rate at a reasonable level. Table 6.2 gives the identification and/or isolation selection for the relevant paths. Section 6.1 explains these selection variables. They are all equally or less restrictive as the offline selection. The efficiency of the HLT relative to offline reconstructed electrons above the p_{T} threshold is nearly 100%.

Run Range	Trigger Name
136033 - 137028	HLT_Photon10_L1R
138564 - 140401	HLT_Photon15_Cleaned_L1R
141956 - 144114	HLT_Ele15_SW_CaloEleId_L1R
146428 - 147116	HLT_Ele17_SW_CaloEleId_L1R
147196 - 148058	HLT_Ele17_SW_TightEleId_L1R_v1
148822 - 149063	HLT_Ele17_SW_TighterEleIdIsol_L1R_v2
149181 - 149442	HLT_Ele17_SW_TighterEleIdIsol_L1R_v3

Table 6.1: HLT paths used by run range

HLT type	H/E	$\delta\eta_{in}$	$\delta\phi_{in}$	$\sigma_{i\eta i\eta}$	Iso_{Ecal}/p_T	Iso_{Hcal}/p_T	Iso_{Track}/p_T
CaloEleId	0.15	-	-	0.014 (0.035)	-	-	-
TightEleId	0.15	0.01 (0.01)	0.08 (0.08)	0.012 (0.032)	-	-	-
TigherEleIdIsol	0.05	0.008 (0.007)	0.1 (0.1)	0.011 (0.031)	0.125 (0.075)	0.05	0.15 (0.1)

Table 6.2: HLT selection by path type. Selection for barrel (endcap, if different).

After events are selected online using the trigger, the raw data is written to permanent storage and then fully reconstructed, as is described in Chapter 5. At this

point, more detailed selection may be applied to obtain a cleaner sample dominated by events of interest.

6.1 Electron Selection

Electrons from W boson decay have a signature distinct from most other common processes, especially from the largest background, QCD. The following criteria were applied offline to reduce backgrounds, principally targeting fake electrons and electrons inside jets. The variables below were chosen for their power in distinguishing signal from background processes. This analysis uses a simple cut-based approach to event selection similar to that used for the inclusive W/Z analysis in CMS [100]. The cut-based approach is transparent, quick, and allows the inversion of cuts to obtain a background rich sample (needed for fitting parameters).

Starting from 15,041,836 events passing the HLT, the first requirement for the electron is to match it to the HLT triggering object within $\Delta R < 0.1$. This requirement makes it possible to correct for the HLT efficiency using the tag and probe method (Section 6.4). Requiring a reconstructed electron that matches to an HLT object leaves 10,321,965 data events. The next requirement is that the electrons are within the acceptance of the detector, leaving 6,823,434 events. This basic requirement ensures that only electrons detectable by CMS, in terms of momentum and position, are considered in both data and Monte Carlo. Acceptance is defined as follows.

- Electrons must be in the fiducial volume of the detector: the pseudorapidity, η, must be $|\eta| < 2.5$, while excluding the gap between the barrel and endcap, $1.444 < |\eta| < 1.56$.

- The highest transverse momentum electron (i.e., leading electron) must have $p_{\mathrm{T}} \geq 20$ GeV. The W boson produces an energetic electron while the electrons from most background processes are much less energetic. A cut at 20 GeV is high enough to avoid any inefficiencies in the HLT, and low enough to maintain a high signal efficiency.

The electron selection is divided into three steps, identification (ID), isolation, and conversion rejection. The cut values were chosen to maximize background rejection while maintaining a predefined acceptable efficiency loss of about 80% for the inclusive W signal. An iterative method was used; the algorithm starts from loose cuts and chooses a target signal to background ratio (S/B). All of the cuts (ID, conversion, and isolation) are examined to determine what value for each cut achieves the target S/B with the least efficiency loss. The cut list is updated with the new number, and the process is repeated until the full set of cuts is determined. The algorithm provides a set of points in S/B versus efficiency space such that a point can be chosen with the desired efficiency (e.g., 80%). One can then apply the corresponding set of cuts. Because of the differences in the behavior of electrons in the ECAL barrel and endcap regions, a different set of cuts is used for each.

There are several electron ID variables that have been found to be most powerful in distinguishing between true electrons and those which are fake or part of a jet [59].

- The shape variable, $\sigma_{i\eta i\eta}$, is a measure of the width of the electromagnetic cluster in the η direction. It is measured in units of crystals in a 5x5 block centered on the super-cluster seed crystal. This quantity should be small because electromagnetic showers from a single electron have little spread in the η direction.

(Spread in ϕ is expected from bremsstrahlung.) We require $\sigma_{i\eta i\eta} < 0.01$ (0.03) in the barrel (endcap).

- The track to cluster matching variables, $\Delta\phi_{in}$ and $\Delta\eta_{in}$, compare electron location according to GSF track information at the vertex to the location of the associated supercluster. $\Delta\phi_{in}$ ($\Delta\eta_{in}$) measures the spread in the ϕ (η) direction between the two. For true electrons, this number is small, although some width is expected for electrons that radiate in the tracker. We require $\Delta\phi_{in} < 0.06$ (0.03) in the barrel (endcap) and $\Delta\eta_{in} < 0.004$ (0.007) in the barrel (endcap).

- The measure of hadronic activity, H/E, determines how much energy of a reconstructed particle reaches the hadronic calorimeter. The ECAL crystals are designed such that true electrons deposit almost all of their energy before reaching the HCAL, so the ratio H/E is small for true electrons. H is the energy in the HCAL towers in a cone of $\Delta R < 0.15$ around the electron position, and E is the energy of the electron super-cluster. We require $H/E < 0.04$ (0.025) in the barrel (endcap).

These electron identification variables are found in Figures 6.1 and 6.2. They are shown with all selection cuts applied except for the cut on the variable plotted. There is excellent agreement between the Monte Carlo and the data, although the QCD MC cross section is consistently underestimated. Some differences in the HLT paths that are available for data and MC, with data HLT paths being tighter, lead to divergences at the tails of the $\sigma_{i\eta i\eta}$ and H/E distributions. Figures 6.3 (barrel) and 6.4 (endcap), with the cut locations indicated, show that the cuts are placed to balance efficiency and significance, $S/\sqrt{S+B}$. After the ID cuts there are 1,205,840 remaining data

events.

One of the most critical requirements for electrons from W decays is that they be isolated from other activity in the event. Events where the leading electrons are found in jets are removed at this stage, such that there is a reduction in the number of QCD events by a factor of three, even after ID cuts have already been applied. Three types of isolation—track, ECAL, and HCAL—are applied and outlined below. All are taken relative to the electron p_{T} for each event [59].

- Track Isolation is the sum of the p_{T} of tracks with $p_{\mathrm{T}} > 0.7$ GeV, each within an annulus of $0.04 < \Delta R < 0.3$ centered around the track direction at the vertex. We require $I_{trk}/E_{\mathrm{T}} < 0.09$ (0.04) in the barrel (endcap).

- ECAL Isolation is the sum of the ECAL reconstructed hits (RecHits) within a cone of $\Delta R = 0.3$ centered around the ECAL supercluster position. The electron energy itself is excluded from the sum by an inner veto cone of three crystals and a strip of three crystals width in η. Only RecHits above a minimum threshold of energy > 0.08 GeV in the barrel, or $E_{\mathrm{T}} > 0.1$ GeV in the endcap, are included in the sum. We require $I_{ECAL}/E_{\mathrm{T}} < 0.07$ (0.05) in the barrel (endcap).

- HCAL Isolation is the sum of the HCAL calo towers within an annulus of $0.15 < \Delta R < 0.3$ centered around the HCAL supercluster position. We require $I_{HCAL}/E_{\mathrm{T}} < 0.10$ (0.025) in the barrel (endcap).

Plots of the individual isolation variables, relative to p_{T}, are shown in Figure 6.5. They are shown in log scale with all selection cuts applied except for the cut on the variable plotted. There is excellent agreement between the Monte Carlo and the data,

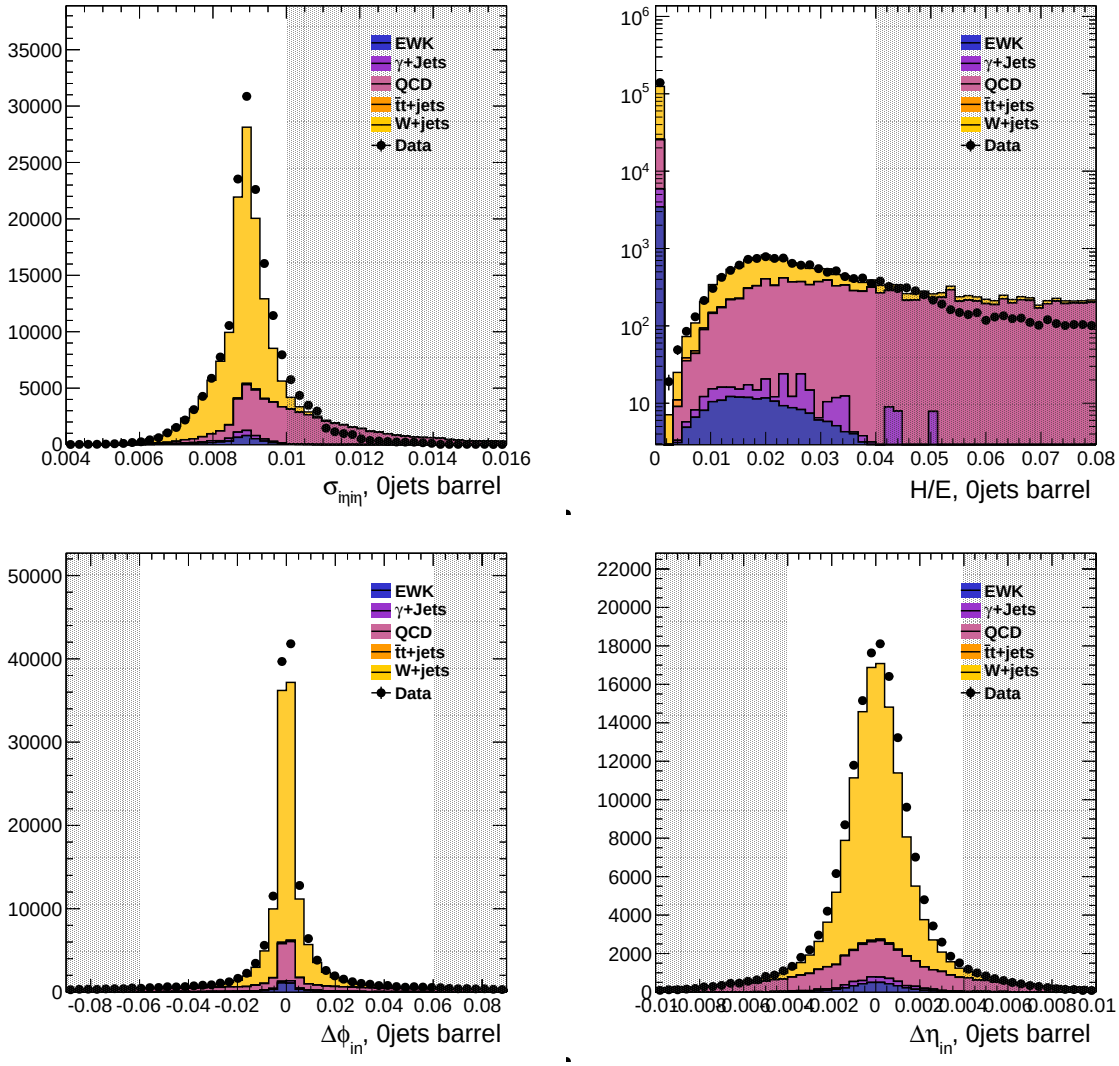


Figure 6.1: Luminosity normalized ID variables in MadGraph and pythia MC along with data for barrel electrons. Shaded area is the rejected events.

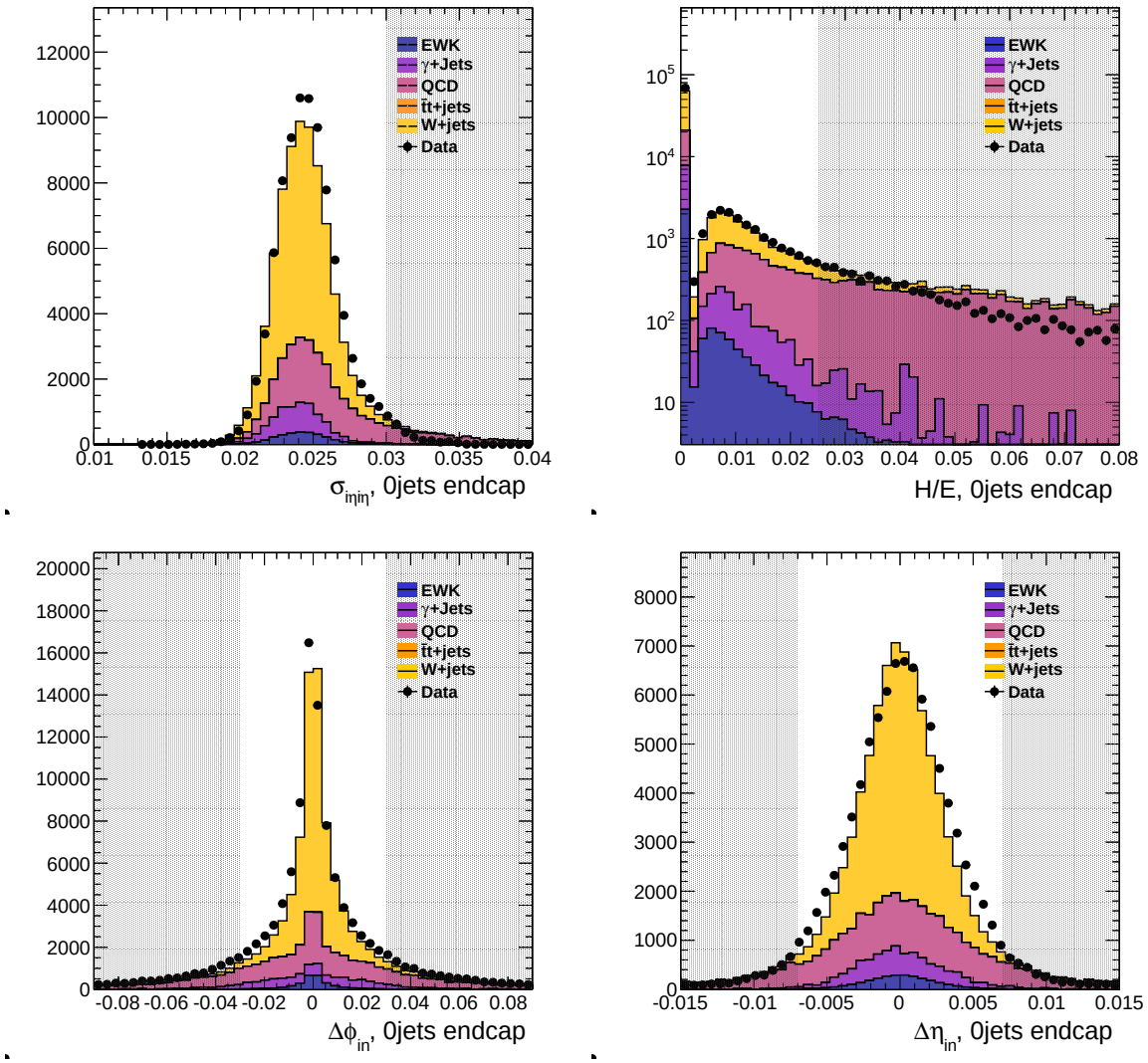


Figure 6.2: Luminosity normalized ID variables in MadGraph and pythia MC compared to data for endcap electrons. Shaded area is the rejected events.

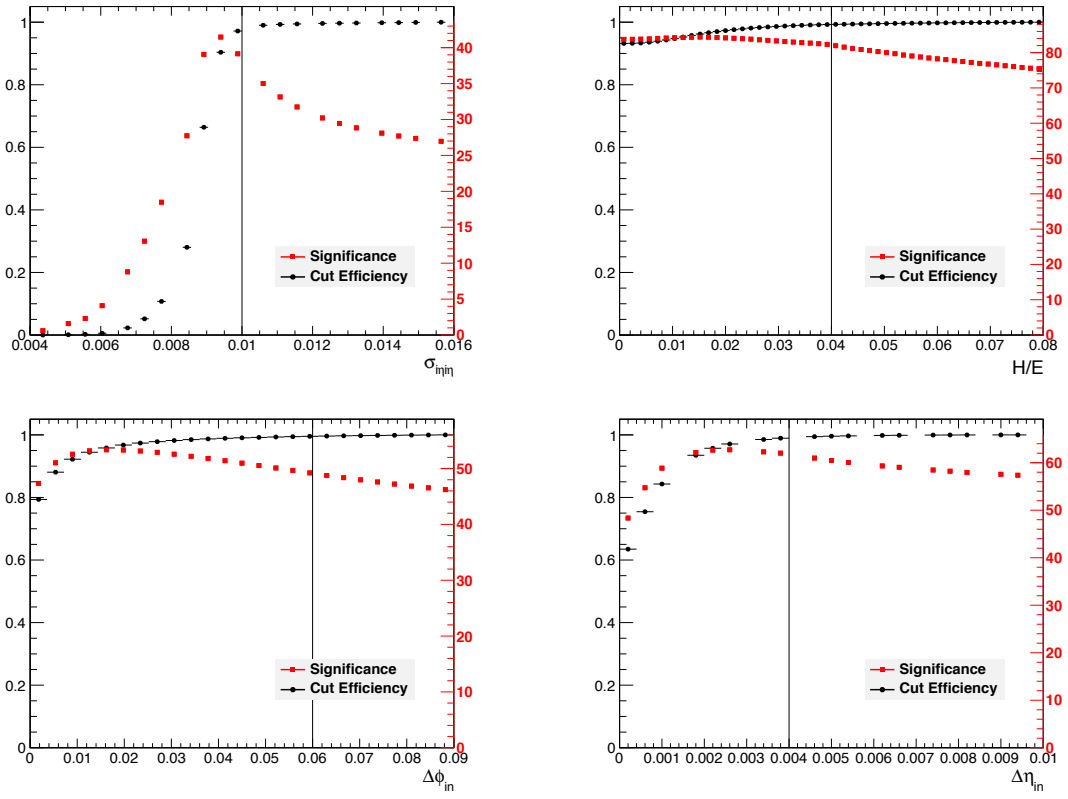


Figure 6.3: ID selection efficiency and $S/\sqrt{S+B}$ for barrel electrons according to MadGraph and pythia MC.

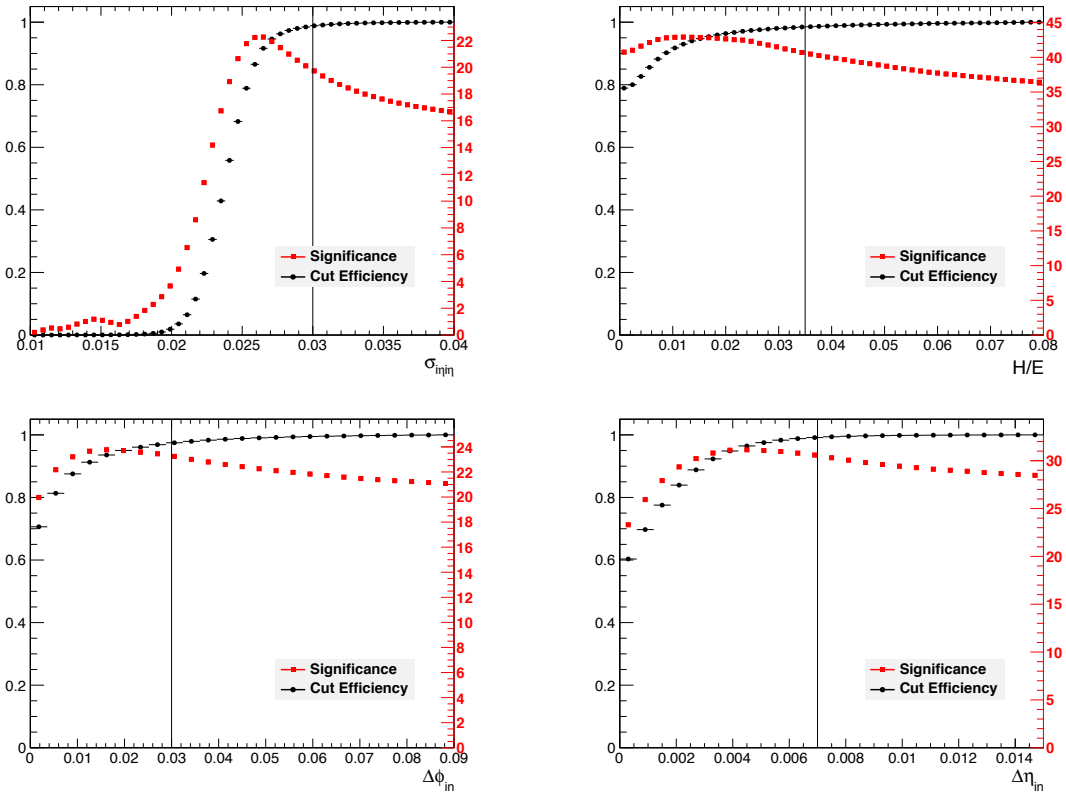


Figure 6.4: ID selection efficiency and $S/\sqrt{S+B}$ for endcap electrons according to MadGraph and pythia MC.

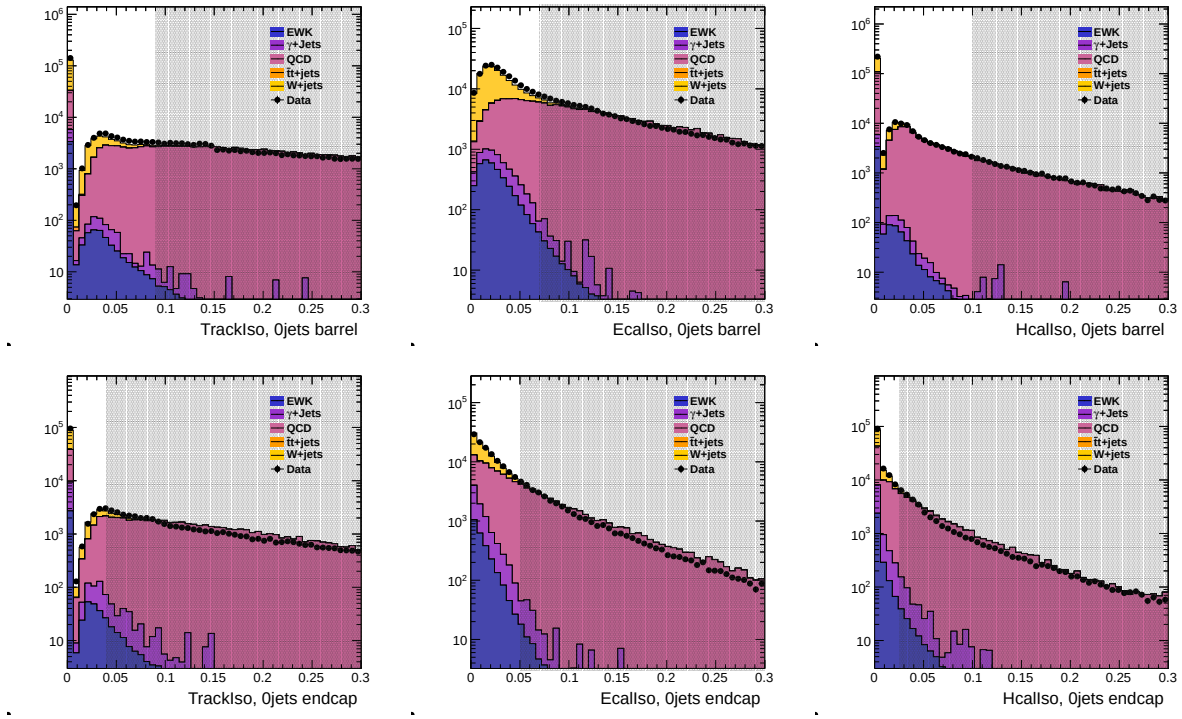


Figure 6.5: Luminosity normalized isolation variables in MadGraph and pythia MC along with data for electrons, all relative to p_{T}. Top plots for electrons in the barrel, bottom plots for electrons in the endcap. Shaded area is the rejected events.

except at the higher values where the effect of the different HLT paths in data and MC is evident. The efficiencies and significances, Figure 6.6, show that the cuts are placed to maintain efficiency without sacrificing too much significance. After the isolation cuts there are 514,511 remaining data events.

Photons converting to electrons are rejected through two requirements, the number of missing tracker hits, and $\Delta cot(\theta)$ and Dist. We require no missing track hits between the vertex and the first measured hit of the reconstructed electron track. Track hits from electrons start at the very first tracker layers, while photons will not register hits until they have converted to electrons, often passing through several tracker layers first. Electrons from conversion will have a "partner" track of opposite

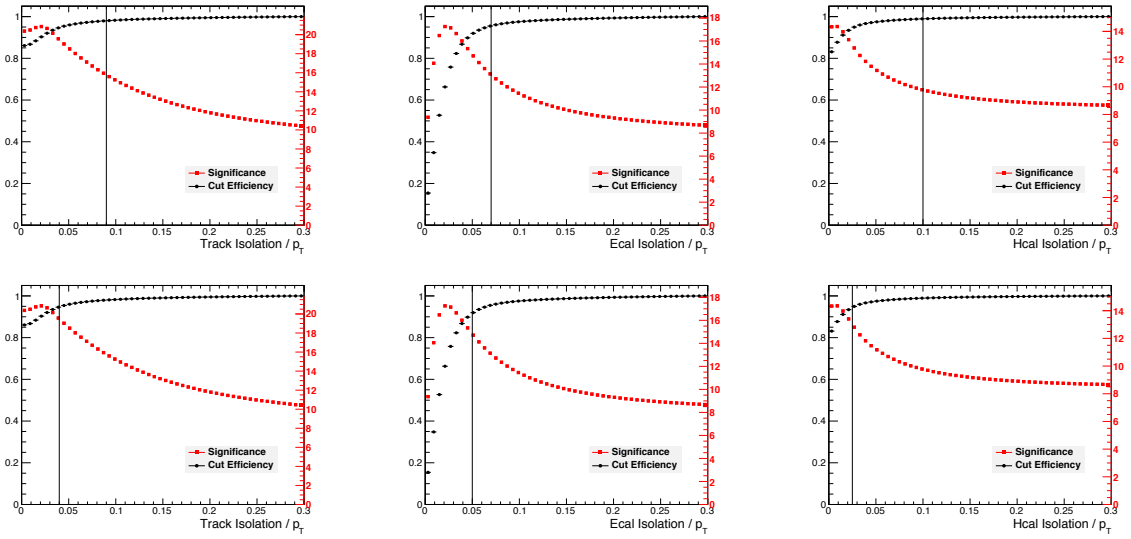
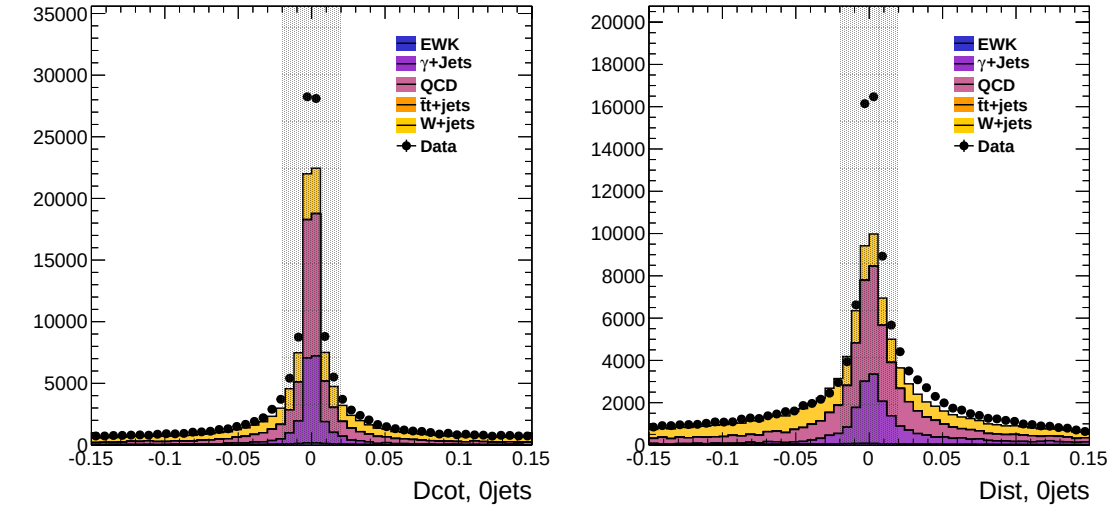


Figure 6.6: Isolation selection efficiency and $S/\sqrt{S+B}$ for electrons according to MadGraph and pythia MC.

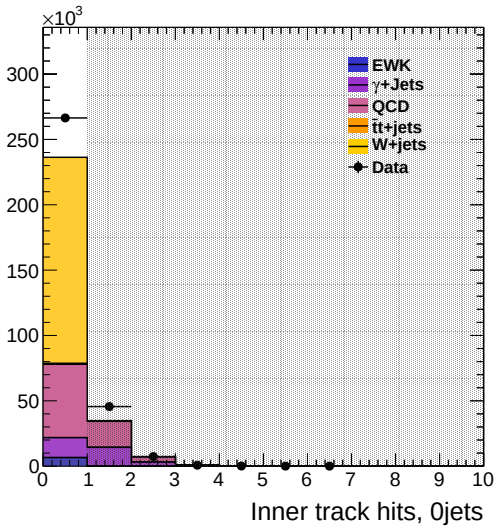
sign. $\Delta cot(\theta)$ refers to the difference in polar angle between the electron track and a nearby partner track. Dist refers to the distance of closest approach to a partner track. The distance variables are taken together and must both be less than 0.02 for an electron to be rejected as a converted photon [101]. These conversion variables are found plotted in Figure 6.7. They are shown with all cuts applied except for the cut on the variable plotted. There is a discrepancy between the data and MC that is explained by scaling of the QCD, which is not expected to be correct as this stage. The efficiency and significance of the distance cuts are shown in Figure 6.8, although since they are both taken together, the efficiency is actually higher than indicated. After the conversion rejection there are 328,701 remaining data events.

A table listing these ID, conversion, and isolation variables and the cut values used, collectively called "Working Point 80 (WP80)", is given in Table 6.3. Figure 6.9 presents the electron p_{T} and η after the full selection has been applied. The plots



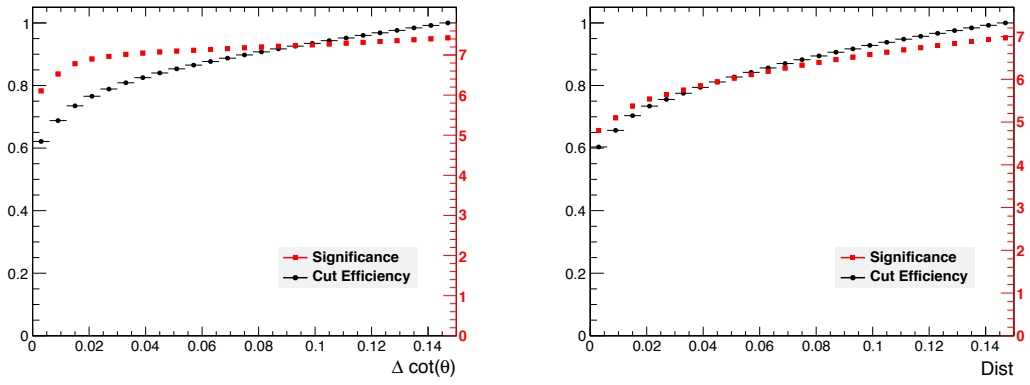
(a)

(b)



(c)

Figure 6.7: Luminosity normalized conversion rejection variables, $\Delta cot(\theta)$ (a), Dist (b), and Missing inner hits (c) in MadGraph and pythia MC along with data. Shaded area is the rejected events.



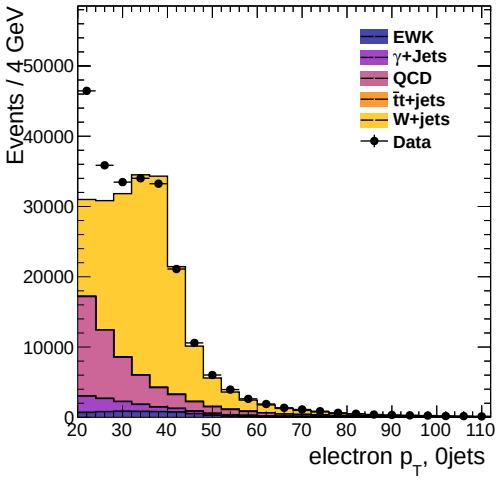
(a) (b)

Figure 6.8: Efficiency and $S/\sqrt{S+B}$ for $\Delta cot(\theta)$ (a) and Dist (b) according to MadGraph and pythia MC.

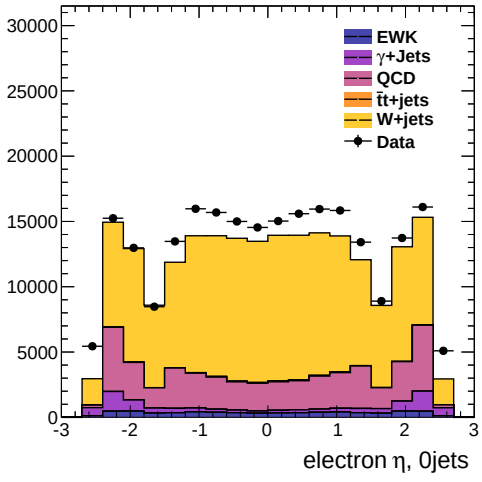
show clearly that the QCD MC cross section is underestimated. The lower (log) plots show better agreement after applying an additional requirement of $M_{\mathrm{T}} > 50$ GeV to enhance signal.

6.2 $\not\!\!E_T$ and Jet Selection

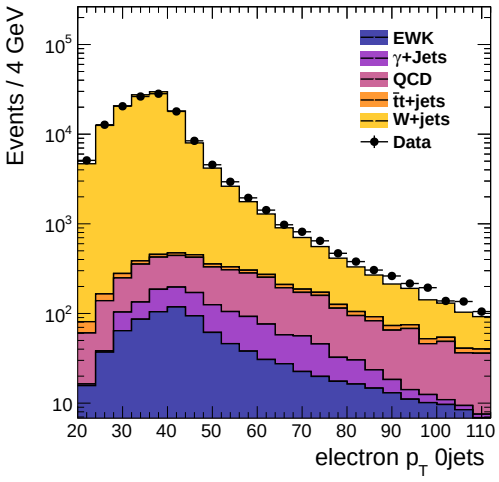
While no explicit cut is made on the $\not\!\!E_T$ in this analysis, it is still important to verify that the MCs used in this analysis match the data for this variable, because it is one of the two decay products of the W boson and is used when calculating the transverse mass of the W boson (which is used in signal extraction fits). This analysis uses "Particle Flow" (PFlow) $\not\!\!E_T$ exclusively, as was described in Section 5.6. Figure 6.10 shows the $\not\!\!E_T$ separated by the inclusive number of jets in the event after full selection has been applied. The agreement of the shape is good, although the pythia QCD MC underestimates the number of QCD events in the data. The roughly 10% uncertainty in the $\not\!\!E_T$ resolution is included as the uncertainty of the M_{T}



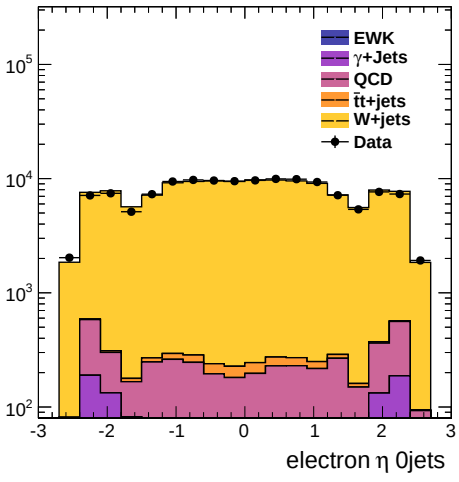
(a) Electron p_{T} for ≥ 0 jets



(b) Electron η for ≥ 0 jets



(c) Electron p_{T} for ≥ 0 jets



(d) Electron η for ≥ 0 jets

Figure 6.9: Electron p_{T} (a) and η (b) of data and luminosity normalized MC for $\geq$ 0 jets. Lower plots (log scale) show p_{T} (c) and η (d) with $M_{\mathrm{T}} > 50$ GeV to enhance signal.

	Barrel	Endcap
I_{trk}/E_T	0.09	0.04
I_{ECAL}/E_T	0.07	0.05
I_{HCAL}/E_T	0.10	0.025
$\sigma_{i\eta i\eta}$	0.01	0.03
$\Delta\phi_{in}$	0.06	0.03
$\Delta\eta_{in}$	0.004	0.007
H/E	0.04	0.025
Missing hits $\leq$	0	0
$\Delta cot(\theta)$	> 0.02	> 0.02
Dist	> 0.02	> 0.02

Table 6.3: WP80 selection cuts for electrons. All variables must be less than the value indicated except for the conversion rejection variables.

selection.

Jets from PFlow reconstruction are selected using $E_{\mathrm{T}} > 30$ GeV. The 30 GeV E_{T} cut has the disadvantage of low statistics at higher jet multiplicities but the advantages of avoiding jets from the underlying event and needing only small pile-up corrections. To take full advantage of the PFlow algorithm, jets are restricted to $|\eta| < 2.4$, where tracker information is available. Jets also have a loose ID applied to clean up stray noise and assure quality jets [102]. These cuts are listed in Table 6.4.

PFlow Jet ID variable	Selection
chargedEmEnergyFraction	< 0.99
neutralHadronEnergyFraction	< 0.99
neutralEmEnergyFraction	< 0.99
chargedHadronEnergyFraction	> 0
chargedMultiplicity	> 0

Table 6.4: Loose jet ID variables and cuts.

Since electrons are included in the list of reconstructed jets, care is also taken to

separate the electrons from the jets so that they are treated as distinct objects. The fully selected electron is removed from the jet candidate collection by rejecting any jet candidate within $\Delta R = 0.5$ of the electron candidate. Plots of the leading jet E_{T} and η spectrums comparing data and MC are shown in Figures 6.11 and 6.12. A further selection of $M_{\mathrm{T}} > 50$ GeV has been added to focus on the signal. The shape of the jet E_{T} and η spectrums are well modeled by the W+jets MadGraph Monte Carlo, although it tends to underestimate the cross sections for greater jet multiplicities.

6.3 Full Selection

In addition to the selection of electrons and definition of jets, three event cuts are made. To reduce the $Z \to e^+e^-$ background, events are rejected if there is a second electron in the acceptance that forms an invariant mass with the first electron that is comparable to the Z mass ($60 < m_{ee} < 120$). Any events with a reconstructed muon of $p_{\mathrm{T}} > 15$ GeV are discarded. In order to have a reasonable, data-driven fit, a cut on the W boson transverse mass is made. Signal extraction is done only with events of $M_{\mathrm{T}} > 20$ GeV. After all of the selection cuts there are 219,815 remaining data events. The effect on the number of remaining events for the full selection chain on data and MC is given in Table 6.5. The final W event yields, separated by number of jets, are shown in Table 6.6.

6.4 Selection Efficiency

The selection efficiency must be measured to convert final event yields into useable cross sections. The yields are divided by the efficiency to return to the original number of events produced within the acceptance. A data driven method has

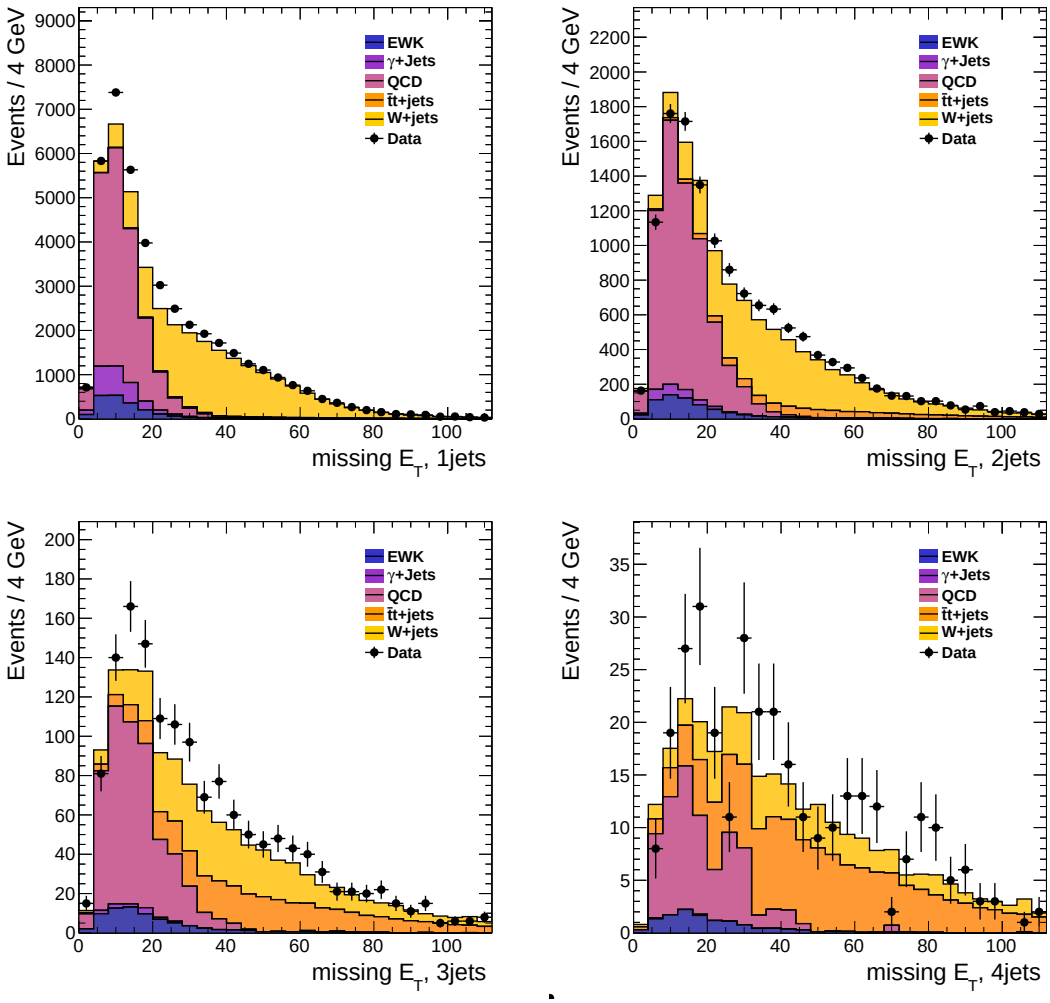


Figure 6.10: Luminosity normalized distribution of Missing E_T in data compared to MadGraph and pythia MC, by inclusive number of jets.

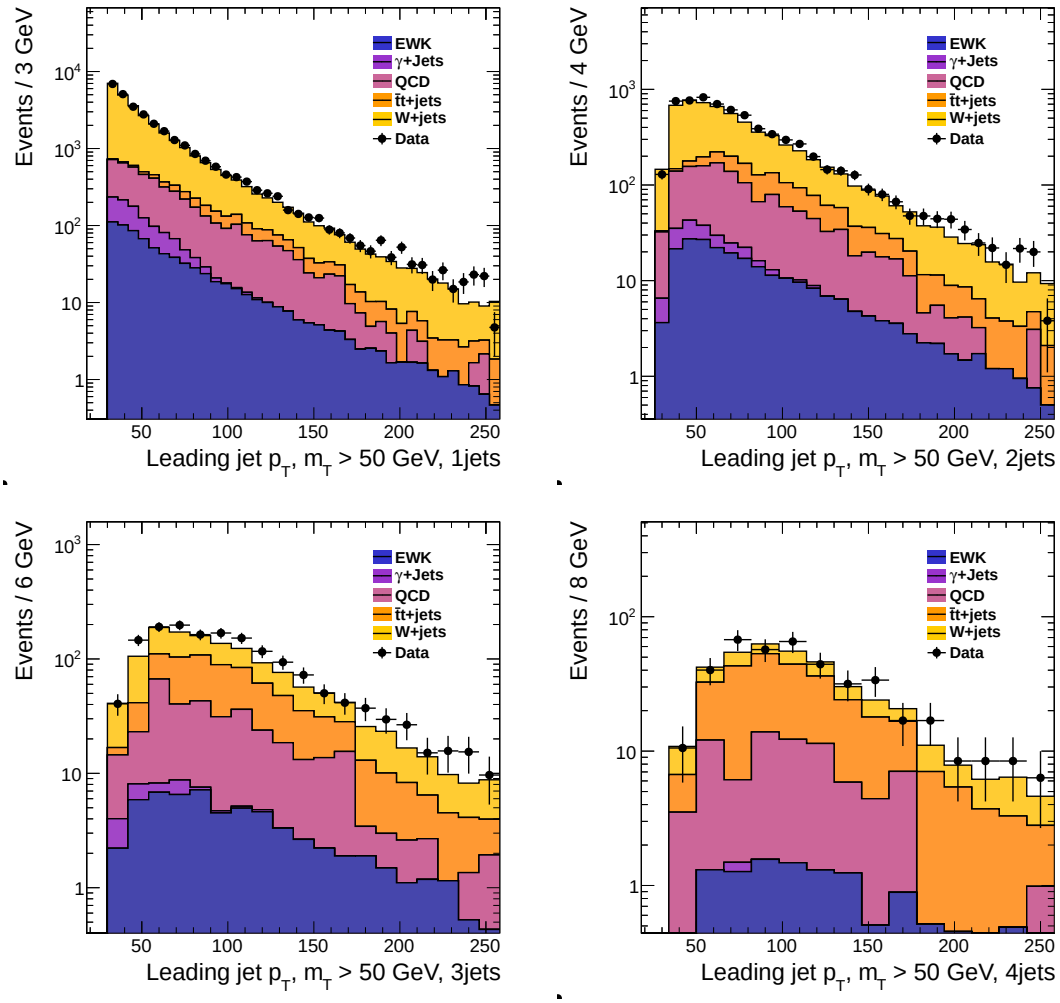


Figure 6.11: Luminosity normalized distribution of jet E_{T} compared to MadGraph and pythia MC, cut on 30 GeV for PFlow Jets and $Wm_T > 50$ GeV, by inclusive number of jets.

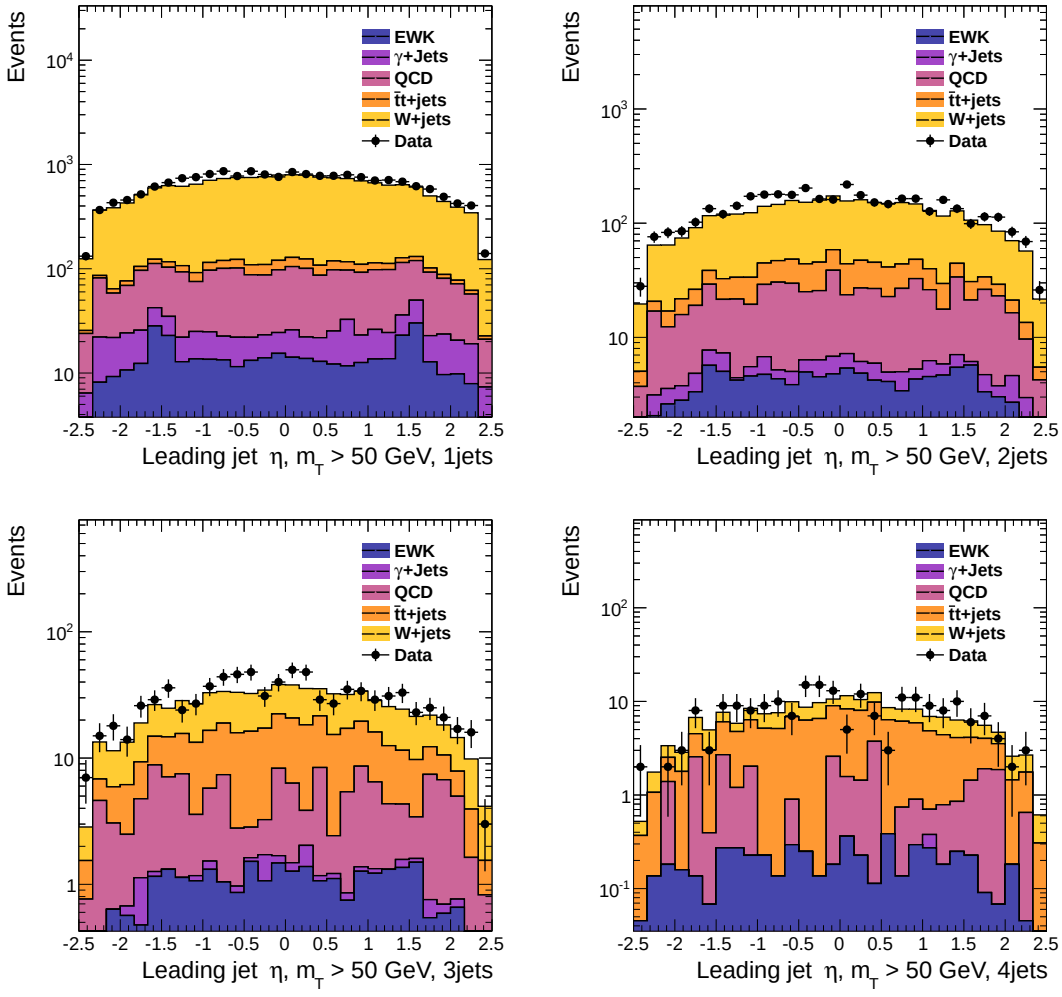


Figure 6.12: Luminosity normalized distribution of jet η in data compared to MadGraph and pythia MC, cut on 30 GeV for PFlow Jets and $Wm_T > 50$ GeV, by inclusive number of jets.

jets	W+jets	W (other)	Z+jets	top	QCD	γ+jets	Data
HLT	228366 (61%)	22716 (100%)	31246 (100%)	2392 (100%)	17054622 (100%)	146431 (100%)	10321965 (69%)
$e^{\pm}$ acceptance	199328 (89%)	17011 (74%)	30506 (99%)	2350 (99%)	16809180 (100%)	140464 (96%)	6823434 (66%)
electron ID	174835 (88%)	8108 (47%)	25600 (83%)	1040 (44%)	964944 (6%)	79886 (57%)	1205840 (18%)
isolation	163821 (94%)	7201 (88%)	24327 (95%)	871 (84%)	194171 (20%)	72894 (91%)	514511 (43%)
conv. rej.	159961 (98%)	6998 (97%)	23686 (97%)	850 (98%)	99400 (51%)	19292 (26%)	328701 (64%)
Z mass veto	158784 (99%)	6927 (98%)	11372 (48%)	759 (89%)	98747 (99%)	19009 (99%)	310950 (95%)
$\lvert\delta_{xy}\rvert \leq 350\mu m$	157844 (99%)	6874 (99%)	11291 (99%)	756 (100%)	85719 (87%)	15414 (81%)	282426 (91%)
muon veto	157753 (100%)	6846 (99%)	11180 (99%)	619 (82%)	85583 (100%)	15404 (100%)	281846 (100%)
W $m_T \geq 20$ GeV	156129 (99%)	6323 (92%)	6308 (56%)	559 (90%)	44838 (52%)	9176 (60%)	219815 (78%)

Table 6.5: Effect of selection cuts on data and signal and background Monte Carlo for ≥ 0 jet events. Efficiency percentages are given relative to previous cut. MC event counts are normalized to 36 pb^{-1}.

been used in conjunction with MC to avoid relying too heavily on MC. "Tag and probe" (T&P) [103] is a data driven method that relies on a relatively background free sample of electrons. The sample has two electrons that must form an invariant mass between 60 and 120 GeV, thus creating a reasonably pure Z/γ^* + jets data sample. One electron is treated as the "tag", satisfying all selection requirements and matched to the HLT triggered object. This requirement further reduces background processes. The other electron is treated as the "probe", selected with the variable being measured. The sample is divided into two subsamples depending on whether the probe passes or fails the desired selection criteria. Fits are performed on the invariant mass of the passing and failing samples, and the efficiency is determined from the level of signal in the pass versus fail subsamples. The tag and probe measurement is divided into three components: reconstruction efficiency, identification and isolation efficiency, and trigger efficiency.

The reconstruction efficiency, ϵ_{reco}, is the efficiency to fully reconstruct an electron. In this case, the probes are ECAL superclusters within the detector acceptance, defined above. Those that are matched within $\Delta R < 0.3$ to a reconstructed electron are considered passing probes, while the unmatched superclusters are failing probes. In terms of the number of events in the ECAL acceptance, $N_{\text{ECALacceptance}}$ and the number of events reconstructed, $N_{\text{reconstructed}}$, the reconstruction efficiency is $\epsilon_{\text{reco}} = N_{\text{reconstructed}}/N_{\text{ECALacceptance}}$. The ID and isolation efficiency, ϵ_{WP80}, is the efficiency of the full working point selection. The probes are reconstructed electrons in the acceptance (those passing the previous step), and the passing probes in this step are those which pass the WP80 selection, defined above. In terms of the number of events

passing WP80, N_{WP80}, the ID and isolation efficiency is $\epsilon_{\mathrm{WP80}} = N_{\mathrm{WP80}}/N_{\mathrm{reconstructed}}$. The trigger efficiency, ϵ_{trig}, is the efficiency of triggering on the electron. The probes are electrons passing WP80. Passing probes are those electrons matched within $\Delta R < 0.1$ to an HLT trigger object from the paths listed in Table 6.1. In terms of the number of events passing the trigger, $N_{\mathrm{triggered}}$, the trigger efficiency is $\epsilon_{\mathrm{trig}} = N_{\mathrm{triggered}}/N_{\mathrm{WP80}}$.

All efficiencies are calculated as a function of the number of jets in the event, with either a 15 or 30 GeV cut on the jet E_{T}. The isolation selection efficiency, in particular, is dependent on the number jets in the events. The fits to the invariant mass at the various selection stages are done using a Breigt-Wigner (BW) [104] convolved with a Crystal Ball (CB) [105, 106] function, separately for the passing and failing probes. The small remaining background is modeled with an exponential function. The signal line shape width parameters for events with ≥ 2 jets are taken from the 1-jet case, as the higher jet multiplicities are statistically limited. An example of the fits can be seen in Figure 6.13. A systematic estimate of the uncertainty for T&P comes from performing the same steps, but fitting the distributions with a Cruijff function instead (details on the Cruijff function, a modified gaussian, are given in Section 7.2, on signal extraction). The average of the two results is taken as the T&P central efficiency value, and the difference is the systematic uncertainty. Tag and Probe is also run on a Z+jets Monte Carlo sample, but instead of fitting, the passing and failing probes are simply counted. Each of the three tag and probe steps are combined for a total tag and probe efficiency $\epsilon_{\mathrm{T\&P}}$:

$$\epsilon_{\mathrm{T\&P}} = \epsilon_{\mathrm{reco}} \times \epsilon_{\mathrm{WP80}} \times \epsilon_{\mathrm{trig}} = N_{\mathrm{reco,\,WP80,\,trigger}}/N_{\mathrm{ECALacceptance}} \tag{6.1}$$

The tag and probe procedure has been performed on both data and Monte Carlo

n jets	W+jets	W (other)	Z+jets	top	QCD	γ+jets	Data
≥ 0	156129 $\pm$ 108	6323 $\pm$ 21	6308 $\pm$ 11	559 $\pm$ 1	44838 $\pm$ 281	9176 $\pm$ 188	219815 $\pm$ 468
≥ 1	18105 $\pm$ 36	862 $\pm$ 8	1970 $\pm$ 6	546 $\pm$ 1	16681 $\pm$ 170	2242 $\pm$ 52	40127 $\pm$ 200
≥ 2	3029 $\pm$ 15	157 $\pm$ 3	357 $\pm$ 2	471 $\pm$ 1	2865 $\pm$ 66	138 $\pm$ 6	7202 $\pm$ 84
≥ 3	458 $\pm$ 5	25 $\pm$ 1	68 $\pm$ 1	327 $\pm$ 1	467 $\pm$ 24	10 $\pm$ 1	1401 $\pm$ 37
≥ 4	75 $\pm$ 2	3 $\pm$ 0	11 $\pm$ 0	148 $\pm$ 0	68 $\pm$ 8	0 $\pm$ 0	310 $\pm$ 17

Table 6.6: Breakdown of events for the different jet multiplicities of signal and background events after the full W+jets selection is applied. Top means the sum of ttbar and single top t-channel, s-channel and tW-channel. The considered QCD is in the range $20 < p_T < 170$ GeV/c. MC event counts are normalized to 36 pb^{-1}. Particle Flow jets with $E_{\mathrm{T}} > 30$ GeV are considered for the counting.

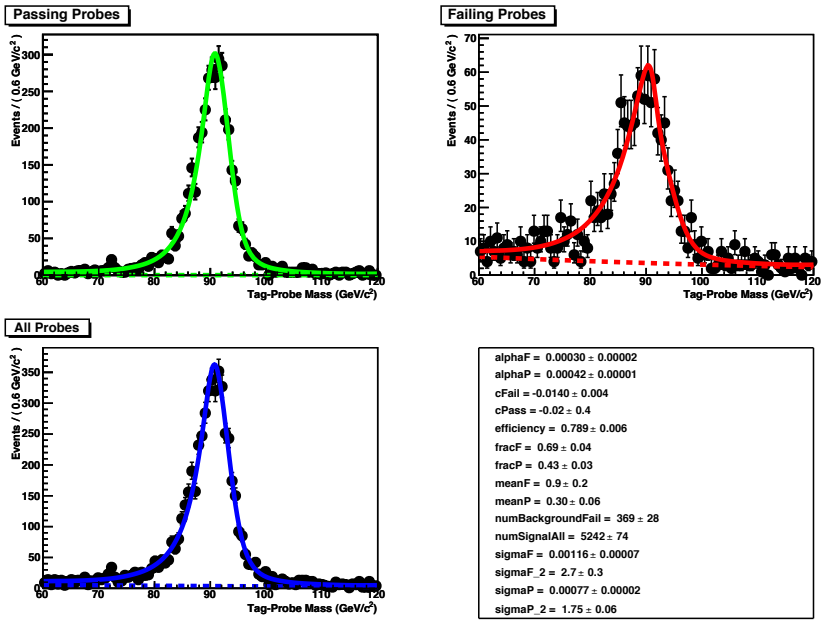


Figure 6.13: Example tag and probe fit results for WP80 selection. The dielectron mass distributions require one tag electron and a second electron with $p_T > 20$ GeV, for events with 1 jet of $E_{\mathrm{T}} > 15$ GeV. "Passing probes" pass the WP80 selection.

and the efficiency results for each step (using the BW+CB fit functions) are shown in Table 6.7. Because of insufficient statistics for the the tag and probe efficiencies, a looser jet E_{T} cut of 15 GeV is used. Studies have shown that the n-jet dependence for efficiency does not dramatically differ between thresholds of 15 and 30 GeV so the T&P results may be safely combined with Monte Carlo results using a 30 GeV jet E_{T} threshold. The data and MC T&P results from the Z+jets sample are used as a correction factor to the efficiencies calculated from the W+jets Monte Carlo. The MC efficiencies, ϵ_{MC}, are determined from the number of W+jets events passing the full selection, including the non-electron specific cuts such as M_{T}, relative to the number of events with a generated electron in the acceptance: $\epsilon_{\mathrm{MC}} = N_{\text{full selection}}/N_{\text{gen acceptance}}$. The tag and probe and Monte Carlo results are combined as follows:

$$\epsilon_{\mathrm{tot}} = \frac{\epsilon_{\mathrm{T\&P_{data}}}}{\epsilon_{\mathrm{T\&P_{MC}}}} \times \epsilon_{\mathrm{MC}} \tag{6.2}$$

By using a ratio of data versus MC T&P results as a scale factor to the W+jets MC efficiency, some of the systematic errors in the T&P calculation cancel out.

The T&P correction factors, the MC efficiency, and the final efficiencies used to correct the signal yields are given in Table 6.8. The tag and probe and full Monte Carlo efficiencies are shown in Figure 6.14. No acceptance corrections are made in this analysis; only the selection efficiency relative to the acceptance is calculated and corrected for. The acceptance cuts necessarily rely on a specific Monte Carlo, limiting the usefulness of the result one theoretical prediction. Correcting for acceptance also introduces theoretical uncertainties depending on the parameters used in the simulation, such as the renormalization scale and the choice of parton distribution function. All of the results in this paper are given within the acceptance, i.e., electron $p_{\mathrm{T}} > 20$ GeV

and $\eta < 2.5$ (excluding the gap region), and compared to MC simulations with the same acceptance requirements on the generated electrons.

	0 jets	1 jet	2 jets	3 jets	≥ 4 jets
Tag and Probe on Monte Carlo					
ϵ_{reco}	0.954 ± 0.000	0.960 ± 0.000	0.962 ± 0.000	0.962 ± 0.001	0.965 ± 0.002
ϵ_{WP80}	0.823 ± 0.000	0.817 ± 0.001	0.809 ± 0.001	0.802 ± 0.002	0.792 ± 0.004
ϵ_{trig}	0.933 ± 0.000	0.935 ± 0.000	0.936 ± 0.001	0.934 ± 0.001	0.940 ± 0.003
$\epsilon_{T\&P}$	0.732 ± 0.000	0.733 ± 0.001	0.729 ± 0.001	0.720 ± 0.002	0.719 ± 0.004
Tag and Probe on Data					
ϵ_{reco}	0.966 ± 0.002	0.960 ± 0.004	0.965 ± 0.008	0.96 ± 0.02	0.95 ± 0.03
ϵ_{WP80}	0.805 ± 0.003	0.789 ± 0.006	0.77 ± 0.01	0.79 ± 0.03	0.72 ± 0.05
ϵ_{trig}	0.975 ± 0.001	0.978 ± 0.002	0.985 ± 0.004	0.98 ± 0.009	1.0 ± 0.006
$\epsilon_{T\&P}$	0.753 ± 0.003	0.743 ± 0.007	0.722 ± 0.011	0.735 ± 0.033	0.693 ± 0.053
Systematics	$\pm$ 0.004	$\pm$ 0.002	$\pm$ 0.006	$\pm$ 0.010	$\pm$ 0.019

Table 6.7: Reconstruction, working point selection, and trigger efficiencies by jet multiplicity (requiring jet $p_T > 15$ GeV). The total data efficiency is the averaged efficiency from the two fitting strategies.

	0 jets	1 jet	2 jets	3 jets	≥ 4 jets
ϵ_{MC}	0.694 ± 0.000	0.646 ± 0.001	0.595 ± 0.002	0.540 ± 0.005	0.486 ± 0.010
T&P scale factor	1.03 ± 0.005	1.01 ± 0.009	0.990 ± 0.016	1.02 ± 0.046	0.976 ± 0.075
ϵ_{tot}	0.713 ± 0.003	0.655 ± 0.006	0.589 ± 0.010	0.551 ± 0.026	0.474 ± 0.039

Table 6.8: Final electron efficiencies used to correct the signal yields. MC counting, using a 30 GeV jet threshold, is scaled by the tag and probe results (using 15 GeV jet threshold). The errors given are statistical; the systematic errors are the same as for tag and probe only.

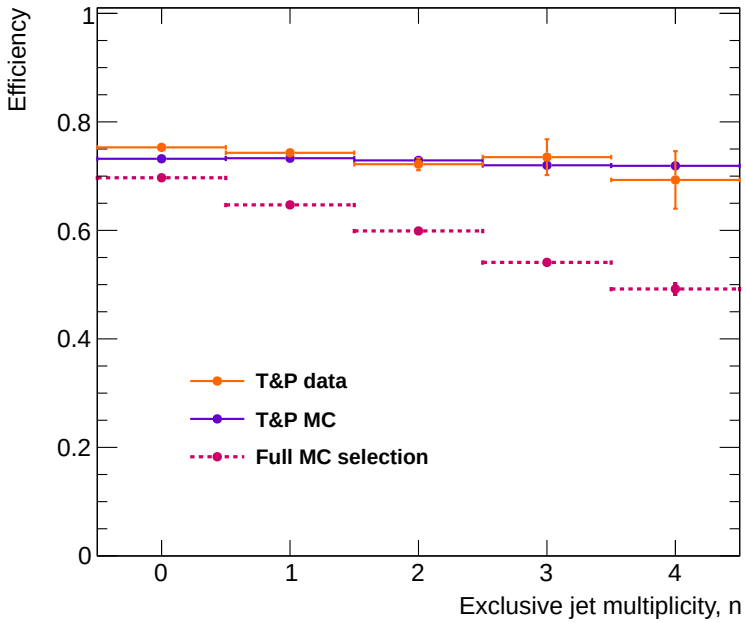


Figure 6.14: Tag and probe efficiency results for data and MC, and the full Monte Carlo selections efficiencies by number of jets.

Chapter 7

Analysis Method

In order to compare a final cross section measurement of W+jets to Monte Carlo, several more steps must take place. The cross section, within the acceptance, for $W + n$ jets is calculated from data using the equation

$$\sigma(W + n \text{ jets}) = \frac{\mathrm{N_W} \cdot \mathrm{U}}{\epsilon_{\mathrm{tot}} \cdot \mathrm{L}} \tag{7.1}$$

where N_W is the number of signal events extracted, U is the unfolding factor, ϵ is the selection efficiency, and L is the integrated luminosity. These values are determined for each n-jet bin exclusively (e.g., exactly 1 jet). The final results are then summed to obtain the inclusive (e.g., ≥ 1 jet) cross sections. The efficiency, ϵ, was discussed in Section 6.4. The integrated luminosity of 36.1 pb^{-1} $\pm$ 1.4 is provided by the LHC group.

One remaining element needed is the number of selected events that are signal events, N_W. The signal is extracted through a fitting method described in Section 7.2. Finally, the unfolding factor, U, is determined when unfolding the cross section by jet multiplicity, covered in Section 7.3. Instead of a determination of the absolute cross section, we will focus on the ratio of cross sections, $\sigma(W+ \geq n \text{ jets})/\sigma(W)$

and $\sigma(W+ \geq n \text{ jets})/\sigma(W+ \geq (n-1) \text{ jets})$. By taking the ratio, many systematic uncertainties cancel. This procedure follows closely the CMS V+jets analysis [107].

7.1 Comparison of Data and Monte Carlo

After selection it is important to verify that the data and Monte Carlo distributions are similar in order to reduce systematic uncertainties in the calculation of MC corrections. The MadGraph MC sample chosen for the W+jets signal models the data well, although it tends to underestimate the data slightly when using the NNLO cross section. The QCD MC sample scale is unreliable is determined by a fit to the M_{T} distribution, as discussed in section 7.2.1. Figure 7.1 shows the M_{T} distributions for the W boson, separated by the number of jets (using inclusive counting). The agreement with the M_{T} shape is good, although the overall normalization needs to be determined by the signal extraction procedure. The plots shown below, Figures 7.1 and 7.3, will be repeated after scaling the MC based on fit results for both signal and background. The distributions in Figures 7.2 and 7.3, showing the rapidity and p_{T} of the W, have an additional requirement of $M_{\mathrm{T}} > 50$ to remove more background and enhance the signal.

Based on the requirement of jet $E_{\mathrm{T}} > 30$ GeV, total event counts are binned by the inclusive number of jets and plotted in Figure 7.4. After placing the M_{T} cut of > 50 GeV, the signal MC shows good agreement with the data but slightly underestimates it for the higher jet multiplicities. Further background studies need to be done to get an accurate measurement of the number of signal events, and thus cross-section, by jet multiplicity. These studies are covered in the following sections.

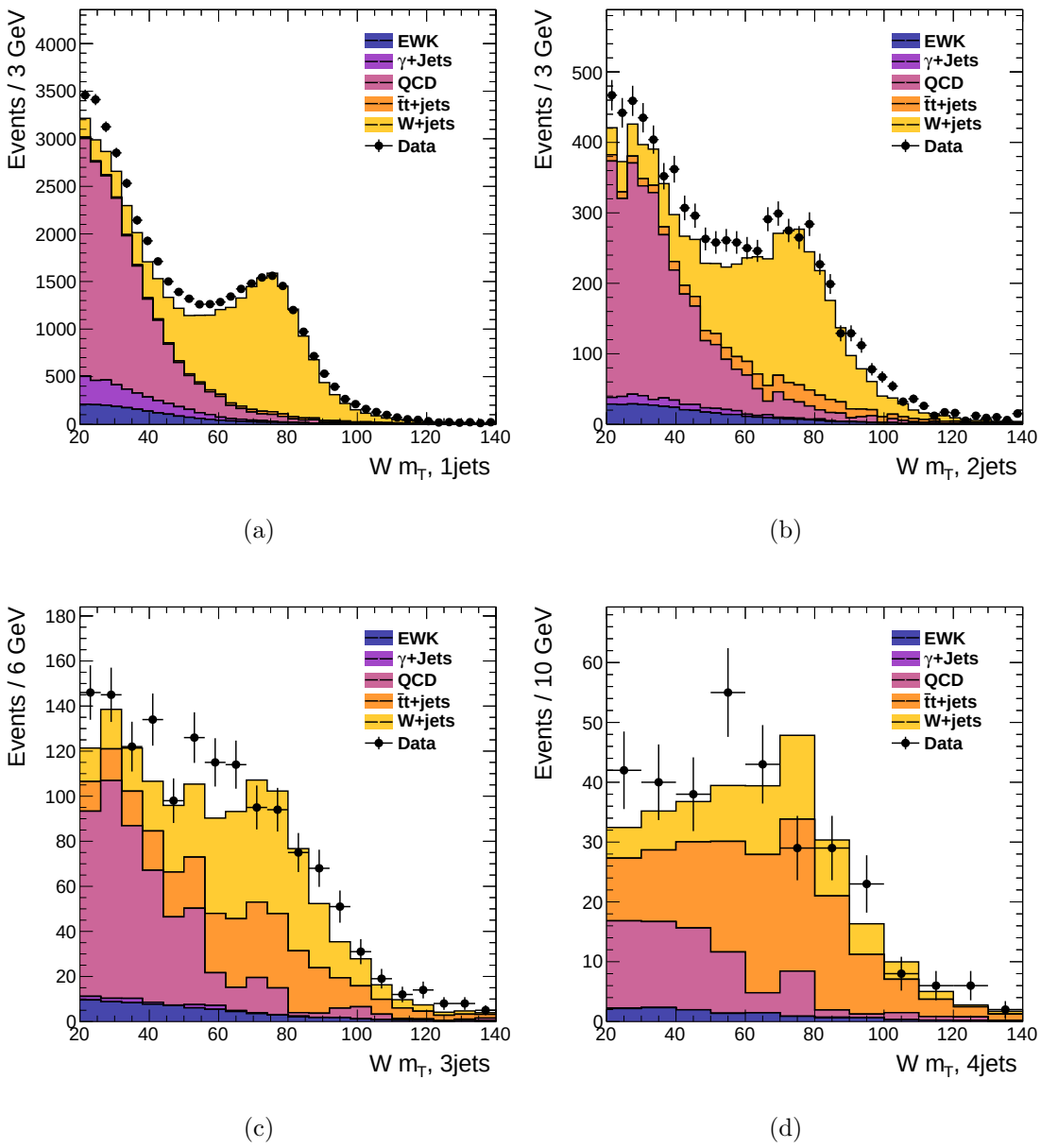
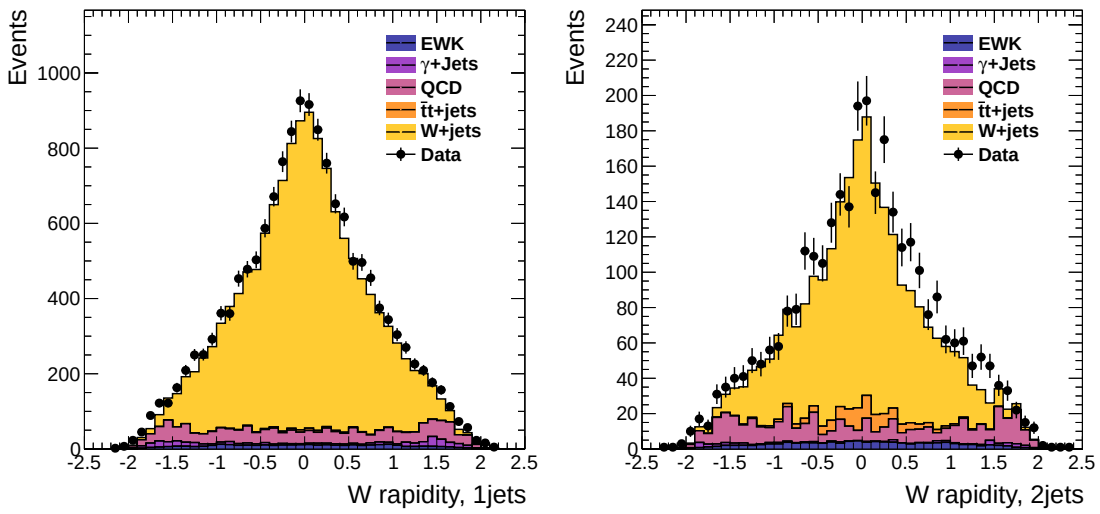
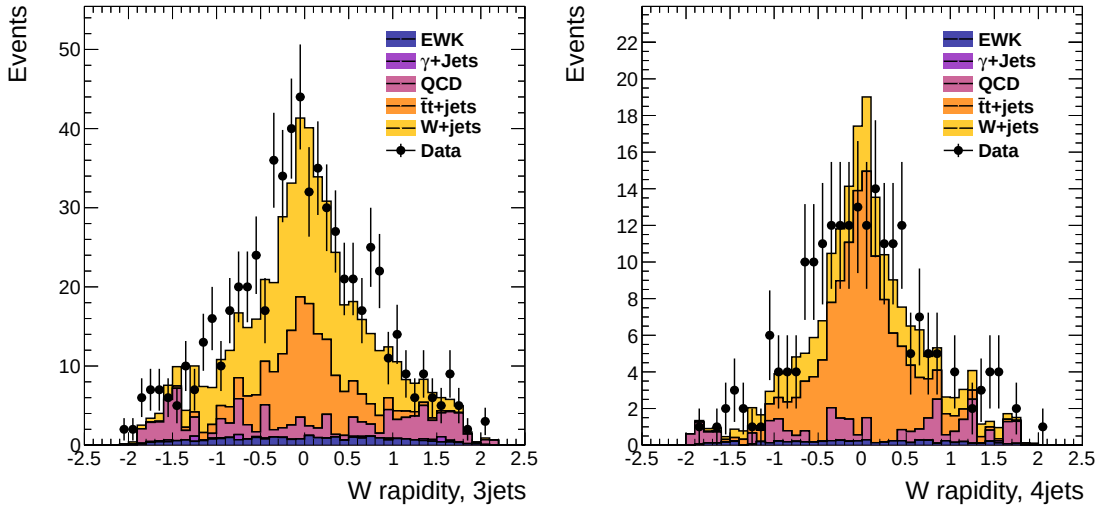


Figure 7.1: W m_T comparison between data and luminosity normalized MadGraph +pythia Z2 MC, for ≥ 1 (a), 2 (b), 3 (c), 4 (d) PF jets ($E_{\mathrm{T}} > 30$ GeV), after full event selection.



(a) (b)



(c) (d)

Figure 7.2: W rapidity comparison between data and luminosity normalized MadGraph +pythia Z2 MC, for ≥ 1 (a), 2 (b), 3 (c), 4 (d) PF jets ($E_{\mathrm{T}} > 30$ GeV), after full event selection and an extra cut of $M_{\mathrm{T}} > 50$ GeV for the W candidate.

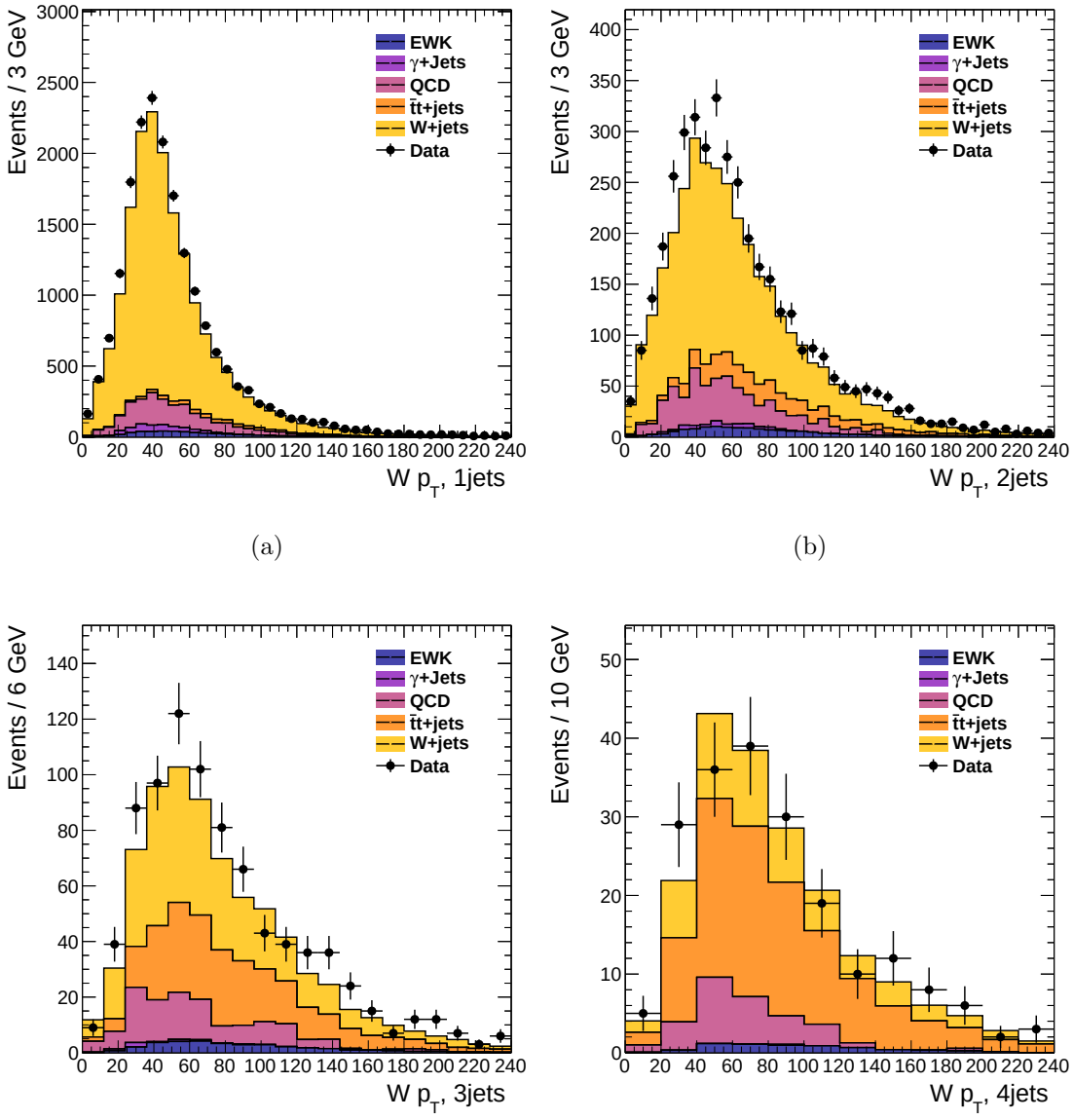


Figure 7.3: W p_{T} comparison between data and luminosity normalized MadGraph +pythia Z2 MC, for ≥ 1 (a), 2 (b), 3 (c), 4 (d) PF jets ($E_{\mathrm{T}} > 30$ GeV), with a cut of $M_{\mathrm{T}} > 50$ GeV for the W candidate.

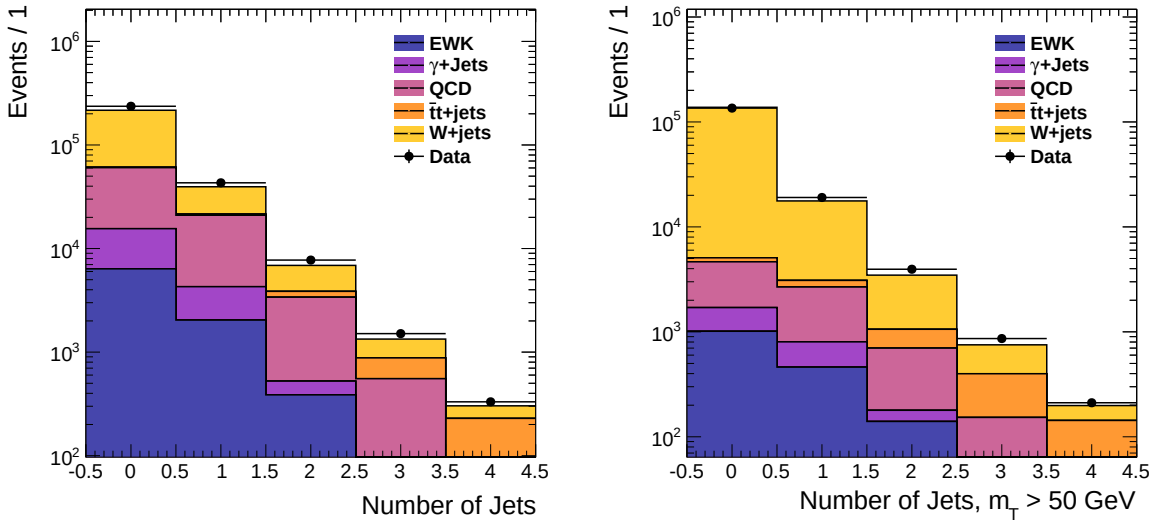


Figure 7.4: Distribution of jet multiplicity in data compared to luminosity normalized MadGraph and pythia MC after full event selection. Jet E_{T} threshold of 30 GeV for PFlow jets. An additional selection of $Wm_T > 50$ GeV is included on the right.

7.2 Background Studies

A significant amount of background remains after the full event selection. The background must be accounted for before calculating cross sections or looking at signal-only kinematic distributions. QCD is the main source of backgrounds for W+jets, mostly through a combination of the high cross section and the production of electrons in jets and in jets faking electrons. The electrons and $\not\!\!E_T$ in QCD differ from those in W+jets events and can therefore be used to distinguish between the two types of events. Fitting to the W transverse mass distribution, $M_{\mathrm{T}}(e, \not\!\!E_T)$, provides separation between the W and top events (those processes that peak in M_{T}) and the other backgrounds including QCD, γ+jets, Z+jets, and $W \to \tau\nu$+jets events.

The other major background to W+jets is top production. Top quarks decay into W bosons and by doing so create an irreducible background with a very similar

signature to W+jets. Because $t\bar{t}$ produces four jets, it does not become a significant background until one looks at the three and four jet multiplicities. One option to reduce this background is to use the presence of jets from b-quarks in the fit. Top production has a higher frequency of jets identified as originating from a b-quark than W+jets production, since top quarks always decays to a W and a b-quark. Therefore, distributions of the number of identified b jets, or "b-tagged" jets will significantly differ between top and non-top events. Because jets produced in association with W bosons directly can also be b-jets, the estimated signal yield using this strategy is only for W+light jets. The b-tagging process is described in Section 7.2.1.1.

7.2.1 Fitting Strategy

The M_{T} and b-tagging are fit simultaneously to extract the signal in one step. Unbinned extended maximum likelihood fits are performed on the distributions for each number of jets ("exclusive" counting, i.e., for events with = 0, 1, 2, 3, and ≥ 4 jets separately). Processes are divided into three species: signal (W + light jets), top ($t\bar{t}$ and single top), and four others (QCD, γ+jets, and EWK: Z+jets, $W \to \tau\nu$). A functional form, rather than a template from MC, is used for modeling M_{T} for all three processes. The MC is not expected to model the data perfectly, especially since it does not account for anomalous noise signals and imperfect modeling of the calorimeters. It also has incomplete descriptions of the underlying events and simplifications of the pile-up simulation. The flexibility of the functional fit, where the majority of the parameters are allowed to float, can take into account these differences between the data and Monte Carlo. For the W+jets and top, a Cruijff function is fit to a template from the MC M_{T} shape. The Cruijff function is a Gaussian with left- and right-handed

widths, σ(L, R), plus first order corrections to these widths, α(L, R), that are varied independently:

$$f(x; m, \sigma_L, \sigma_R, \alpha_L, \alpha_R) = N_s \cdot e^{-\frac{(x-m)^2}{2\sigma^2+\alpha(x-m)^2}} \tag{7.2}$$

where $\sigma = \sigma_L(\sigma_R)$ for $x < m(x > m)$ and $\alpha = \alpha_L(\alpha_R)$ for $x < m(x > m)$.

For signal events with zero or one jet, a second Cruijff is added to the first to account for the kinematic effect of the electron p_{T} cut. The lower resolution in the higher jet multiplicities masks this effect so that only one function is needed. For events with fewer than two jets, all parameters are floated for the first function due to sufficient statistical power. For the two and three jet samples, α_L and α_R are held constant. The top background for all jet multiplicities and signal parameters for ≥ 4 jets are fixed to the MC because of fewer statistics.

Monte Carlo simulations of QCD are not reliable and so a data-driven method is used to extract the QCD+others shape. The electron ID cut is inverted on data to create a background rich M_{T} distribution. The shape of this background distribution is similar to the MC distribution with and without the ID cut, as shown in Figure 7.5. An isolation-inverted cut was also investigated, but the shape did not match the full distribution as well. The ID cut is also less correlated with M_{T} than the isolation and thus interferes less with the shape. This inversion method also has the advantage of including detector effects not fully modeled by the MC, such as dead towers and anomalous signals. A Cruijff function is fit to the QCD enriched sample for the initial QCD+γjet+EWK parameterization. All of the QCD+others parameters are allowed to float when fitting to data. Monte Carlo distributions, and various methods of enriching the background in data, were also used for the initial parameterization as a

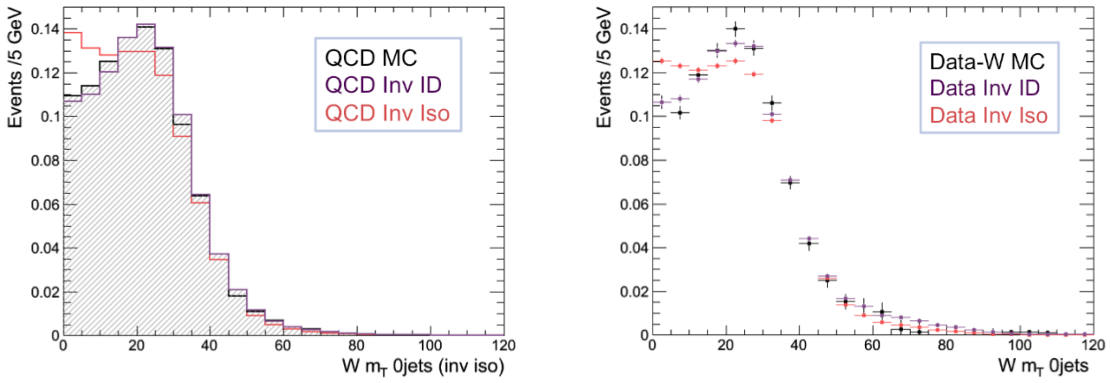


Figure 7.5: M_{T} distribution for events with ID or isolation cut inverted. On the left, QCD events, on the right, data compared to data with the W+Jets MC subtracted.

cross check. All methods gave similar results, well within the statistical uncertainty, and are included in the systematic estimate.

The Cruijff fits to MC, for the initial parameterization, are shown in Figures 7.6 through 7.8. The yields for each species are allowed to float when the combined background and signal functions are fit to the full data distribution.

7.2.1.1 B-tag Fit

Because the M_{T} distribution cannot easily distinguish the W+jets from top events, a second fit, using the number of b-tagged jets, is folded in with the M_{T} fit. The method used for tagging jets as coming from b-quarks is described in Section 5.5.1. Fitting the number of b-tagged jets is done using the following probability distribution

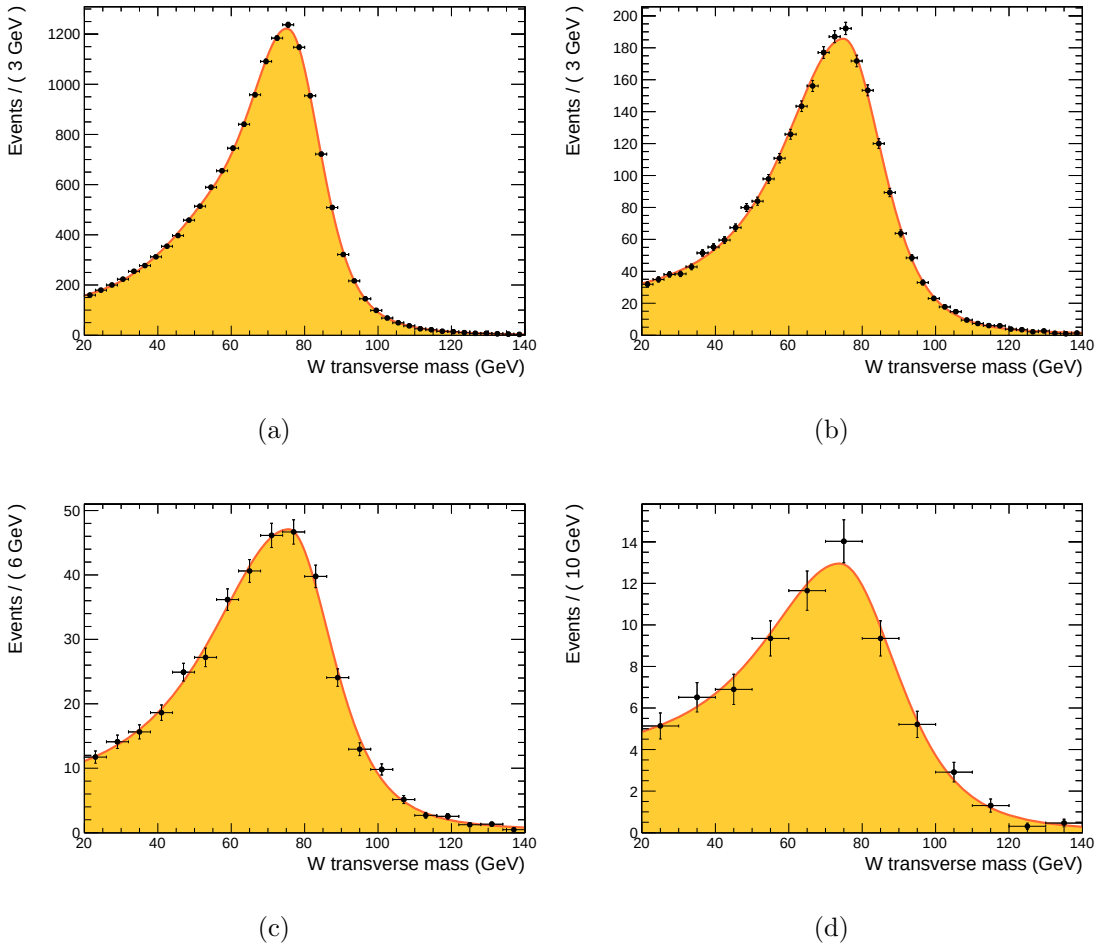


Figure 7.6: Fit of Cruijff function(s) to M_{T} distribution of W+jets MC for 1 (a), 2 (b), 3 (c), and ≥ 4 (d) jets.

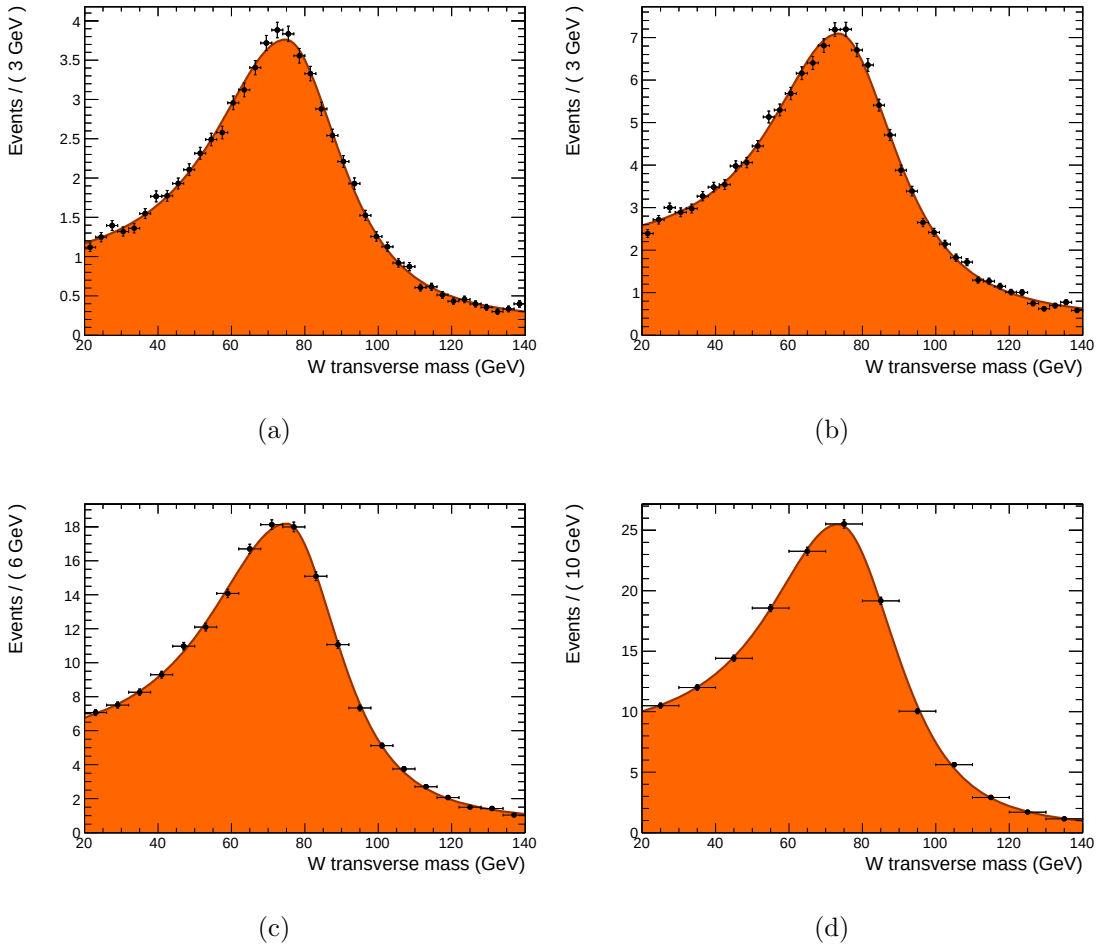


Figure 7.7: Fit of Cruijff function to M_{T} distribution of top MC for 1 (a), 2 (b), 3 (c), and ≥ 4 (d)jets.

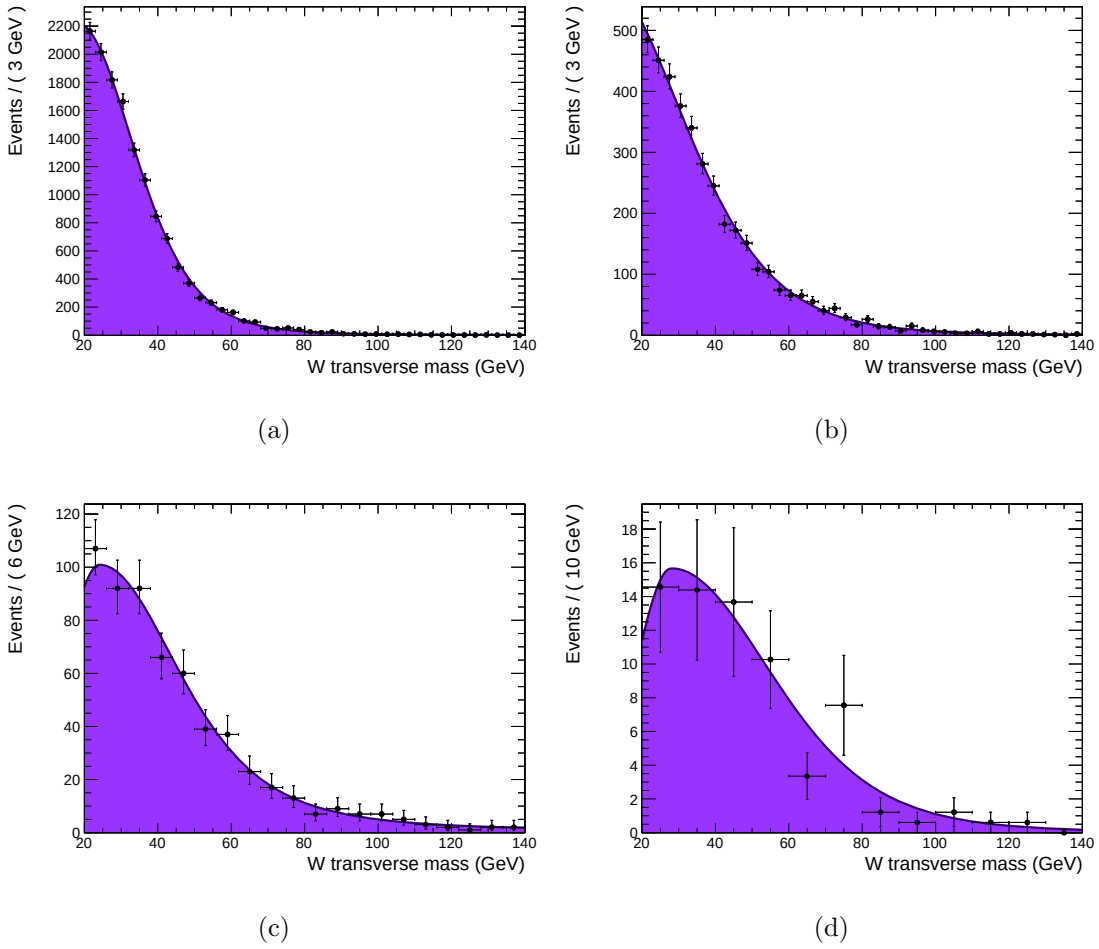


Figure 7.8: Fit of Cruijff function to M_{T} distribution of QCD MC for 1 (a), 2 (b), 3 (c), and ≥ 4 (d) jets.

function (pdf) for the number of b-tagged jets (n_b):

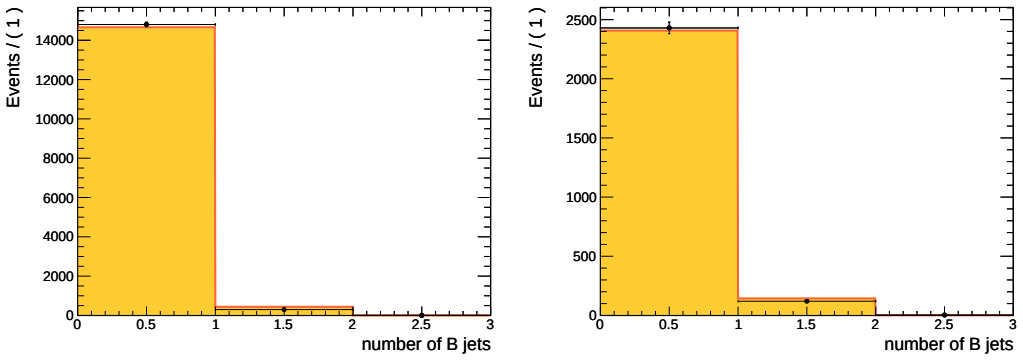
$$\begin{aligned} P(n_b = 0) &= (1 - \epsilon_{nob})^{n_j - n_{bj}} \cdot (1 - \epsilon_b)^{n_{bj}} \\ P(n_b = 1) &= (1 - \epsilon_{nob})^{n_j - n_{bj} - 1} \cdot (n_j - n_{bj}) \cdot (1 - \epsilon_b)^{n_{bj}} + \\ &\quad (1 - \epsilon_{nob})^{n_j - n_{bj}} \cdot (1 - \epsilon_b)^{n_{bj} - 1} \cdot \epsilon_b \cdot n_{bj} \\ P(n_b \geq 2) &= 1 - P(0) - P(1) \end{aligned} \tag{7.3}$$

where

- n_j is the number of jets
- n_{bj} is the number of b-jets in acceptance (i.e., the number of reconstructed jets originating from b hadronization). Each of them (0b/1b/2b) is treated as a sub-species of the fit for the top species.
- ϵ_{nob} is the b mistag rate
- ϵ_b is the b tag rate

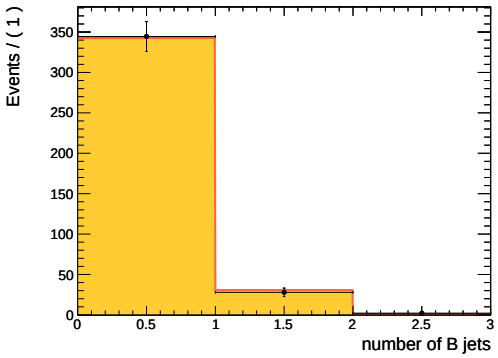
Fixed values of $\epsilon_{nob} = 0.029$ and $\epsilon_b = 0.63$ for the mis-tag and tag rates are based on Monte Carlo studies, but have also been validated with data [108].

The top species is further divided into three subspecies corresponding to the number of b-flavor jets in the acceptance (true b-jets, n_{bj}): 0, 1, and ≥ 2. Each subspecies is fit separately and has its own yield. The signal and other backgrounds are assumed to have $n_{bj} = 0$. Instead of using the pdf for the "other" backgrounds, which can be distinguished using M_{T}, the n_b distribution is simply taken from a MC template. Comparison of the n_b pdfs to MC are shown in Figures 7.9 through 7.11.

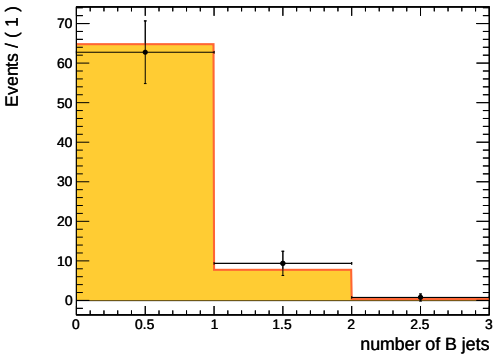


(a)

(b)



(c)



(d)

Figure 7.9: Comparison of n_b distribution of W+jets MC with n_b pdf (Eqn 7.3) for 1 (a), 2 (b), 3 (c), and ≥ 4 (d) jets.

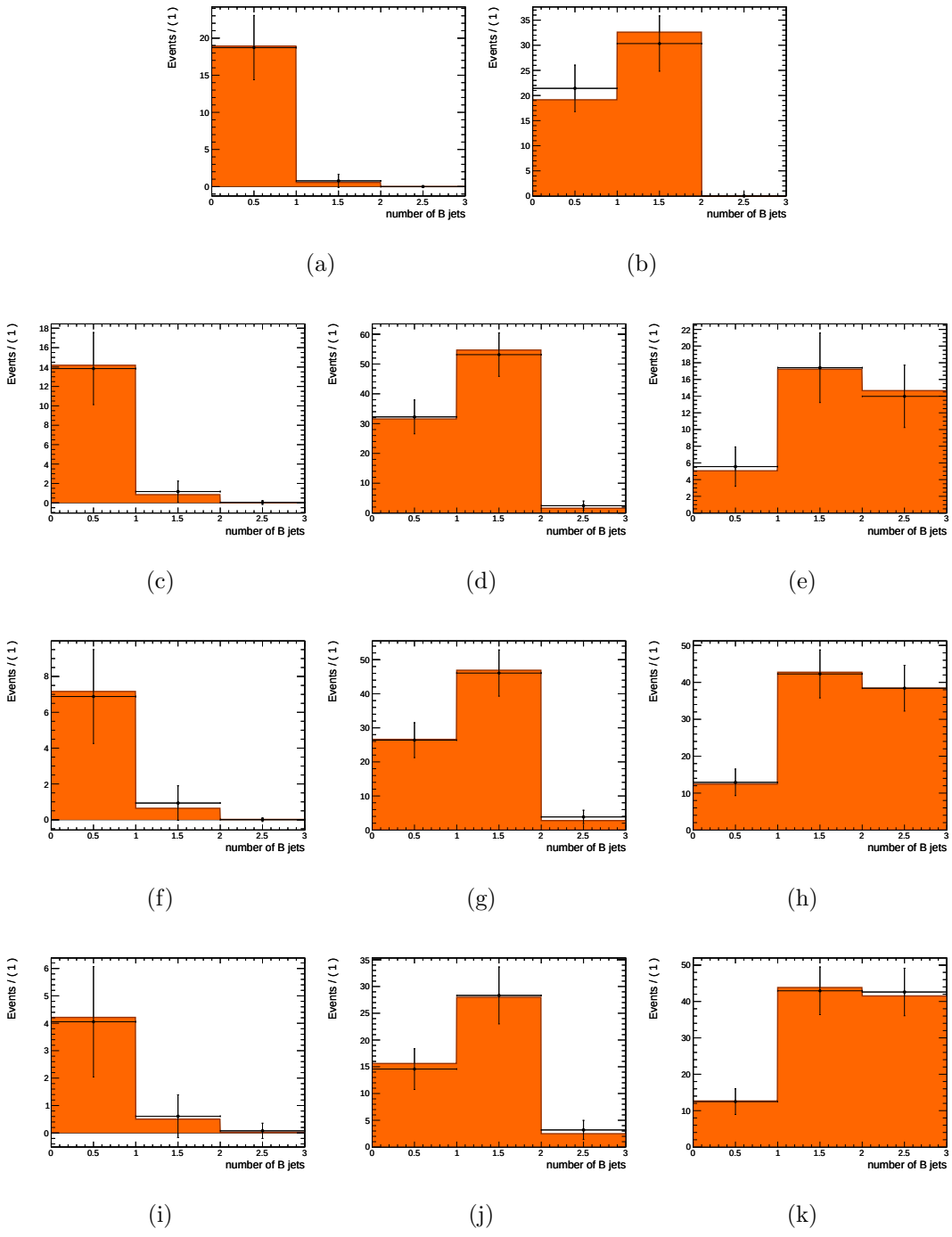
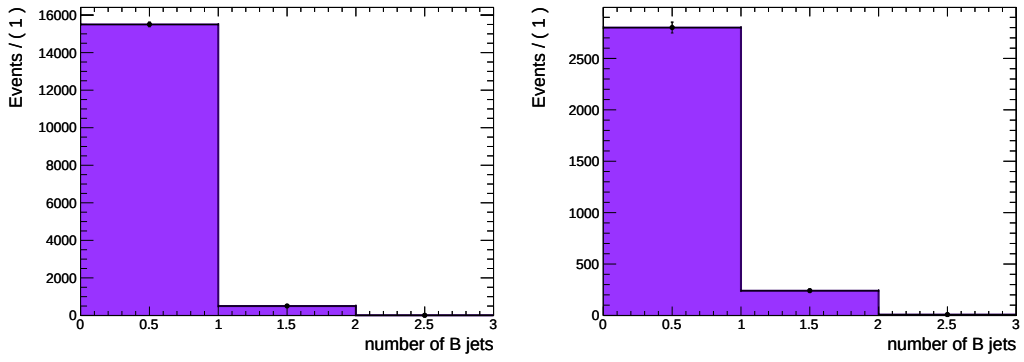
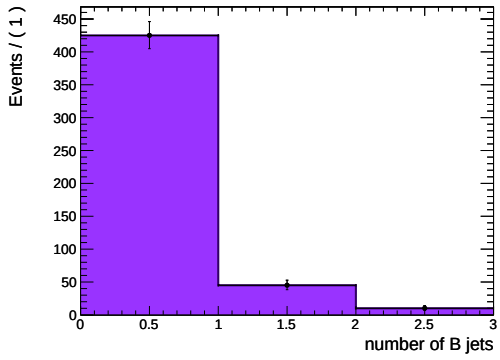


Figure 7.10: Comparison of n_b distribution of top MC with n_b pdf (Eqn 7.3) for 1, 2, 3, and ≥ 4 jets. Rows are number of b-jets, columns are number of jets.

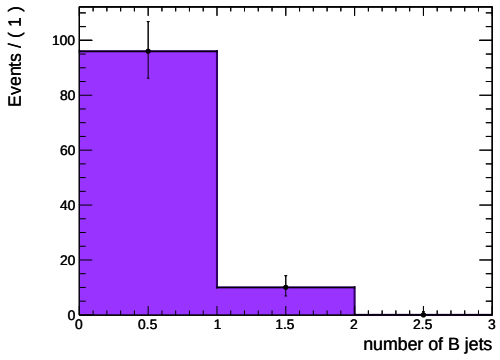


(a)

(b)



(c)



(d)

Figure 7.11: Comparison of n_b distribution of QCD MC to n_b pdf (Eqn 7.3) for 1 (a), 2 (b), 3 (c), and ≥ 4 (d) jets.

7.2.1.2 Fitting to data

The number of jets, n_j, is counted exclusively (inclusively for ≥ 4 jets) and M_{T} and n_b are fit together for each number of jets, 0 through ≥ 4. The two variables, M_{T} and n_b, are assumed uncorrelated and the likelihood fit function can be taken as the product of the two. The fit results give a reasonable estimate of the number of signal-only events. The results of the fit to 0-jet events can be seen in Figure 7.12. For the 0-jet case, where there are very few top events, the top yield is taken directly from MC. Figures 7.14 and 7.15 show the results of fitting M_{T} and n_b on data for 1, 2, 3, and ≥ 4 jets. The Cruijff and b-tagging pdfs appear to be good models for the shape of the data and give reasonable signal yields.

To validate the fit results, the pull of the signal yield was determined. The pull is the difference between the fit yield and the generated number of events, divided by the error on the fit:

$$pull = \frac{N^{fit}_{W+jets} - N^{gen}_{W+jets}}{N^{err}_{W+jets}}. \tag{7.4}$$

Six-hundred pseudo-data-samples were generated and fit to get a distribution for the pull. For an unbiased fit the mean and sigma should be 0 and 1 respectively. Figure 7.13 shows two sample pull plots for the signal yield. The pull sigma is as expected, but the mean is shifted slightly to the left, indicating the fit underestimates the yield relative to the number of generated events.

The results of the fit are yields for each of the three species and are used to scale the Monte Carlo. Of highest interest are the signal yields, $N_{\mathrm{obs}}(n\mathrm{jet})$, given in Table 7.1. The yields from the fit are corrected for the selection efficiency to recover the cross-section within the acceptance, $N_{\mathrm{effcorr}}(n\mathrm{jet})$.

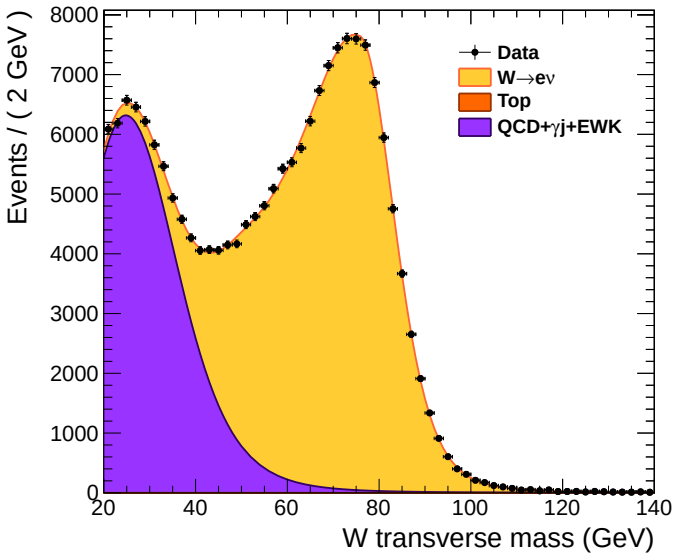
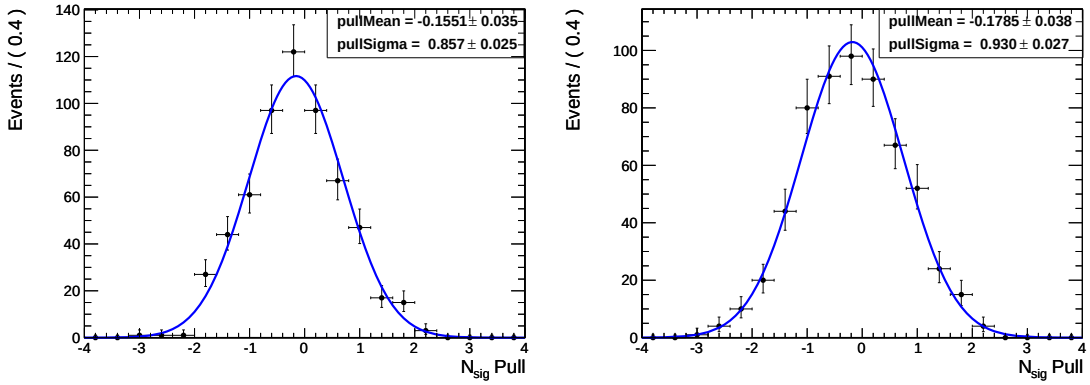


Figure 7.12: Functional fit to the data W m_T distribution. The solid distributions are the Cruijff fits for the signal, top (very small) and other backgrounds.



(a) (b)

Figure 7.13: Example pull plots for the signal yield for 1 jet (a) and 3 jets (b) . Plots are made from generating and fitting to 600 pseudo-datasets.

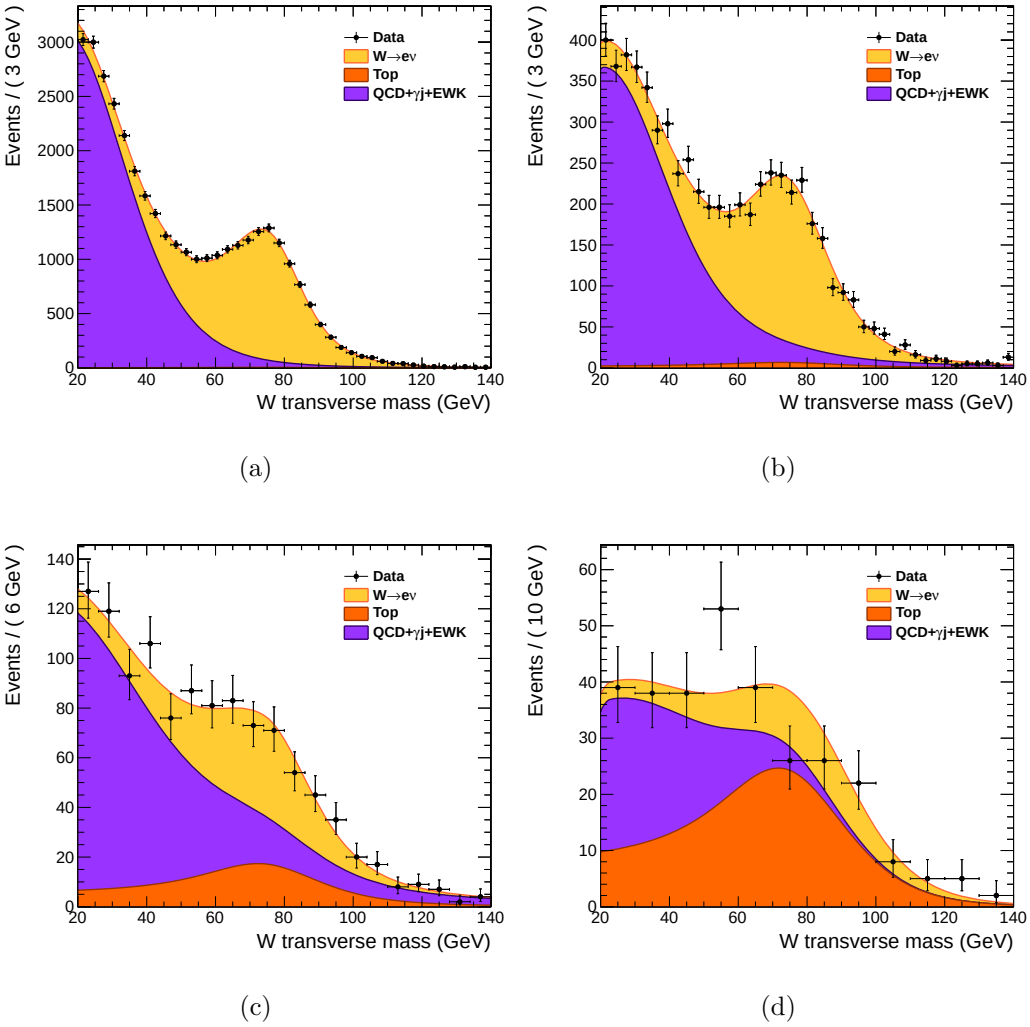


Figure 7.14: M_{T} distribution for 1 (a), 2 (b), 3 (c), and ≥ 4 (d) jets with the functional fit results superimposed, for signal (W+jets), top, and all others.

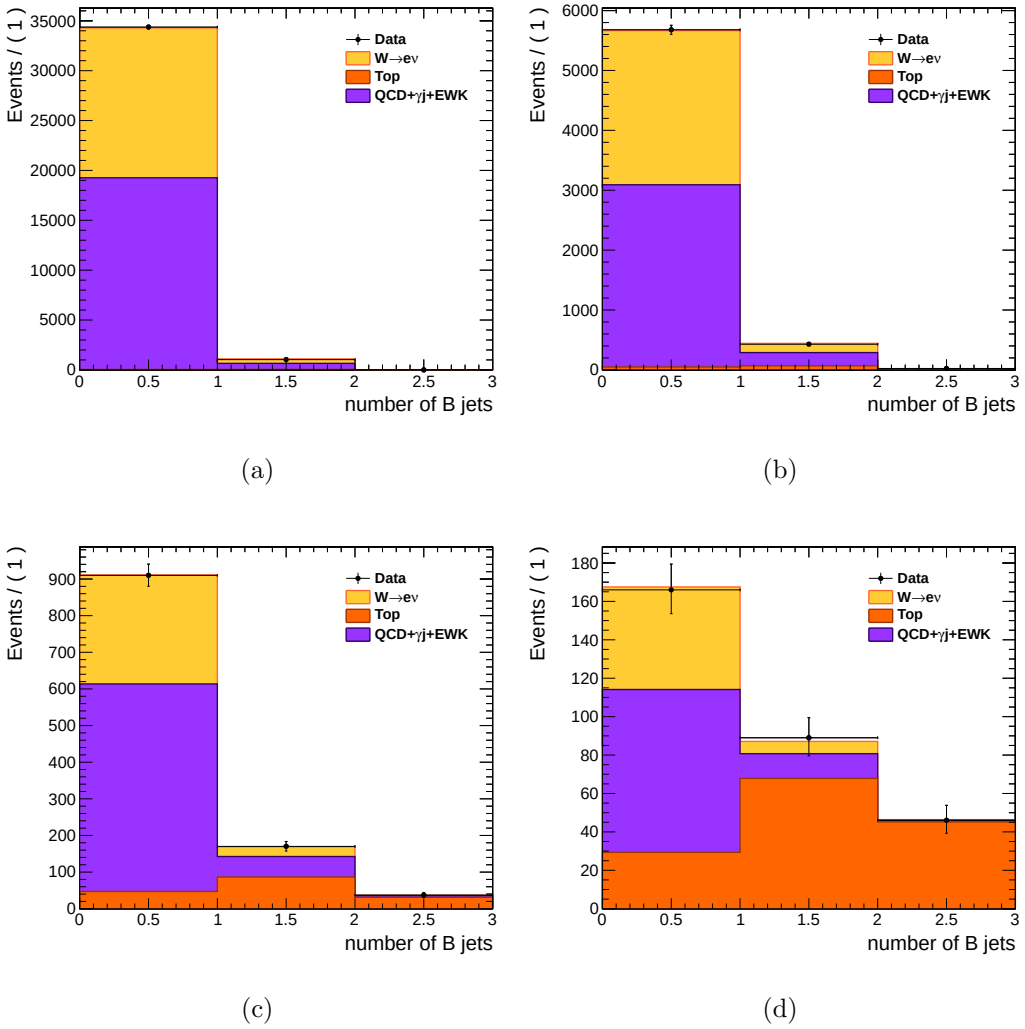


Figure 7.15: B-tagged jet distribution for 1 (a), 2 (b), 3 (c), and ≥ 4 (d) jets with the pdf fit results superimposed, for signal (W+jets), top, and all others.

jets	N_{obs}	$N_{effcorr}$
$= 0$	131376 $\pm$ 402	195566 $\pm$ 979
$= 1$	15476 $\pm$ 143	23138 $\pm$ 458
$= 2$	2730 $\pm$ 74	4289 $\pm$ 244
$= 3$	362 $\pm$ 35	709 $\pm$ 77
≥ 4	60.1 $\pm$ 17	150 $\pm$ 40

Table 7.1: Yields from the signal extraction, N_{obs}, and the efficiency corrected results, $N_{effcorr}$, by exclusive jet multiplicity, where the PF jet has $E_T > 30$ GeV and $|\eta| < 2.4$. Errors are statistical only.

7.3 Unfolding

The jet E_{T} is transformed and distorted by detector effects, such as the finite resolution and limited acceptance. This E_{T} smearing affects the jet counting. In order to arrive at the "true" distribution of W+jets events as a function of the jet multiplicity, the efficiency corrected distribution, $N_{\mathrm{effcorr}}(n\mathrm{jets})$, needs to be further corrected for the smearing effects. This un-smearing, or "unfolding", results in a distribution less dependent on the detector. The following unfolding methods have been investigated:

- singular value decomposition or SVD method as proposed in [109]. This method is used as the default for the nominal results.
- iterative or "Bayesian" method as proposed in [110]. This method is used as a systematic check.

Unfolding is performed by means of a migration matrix M_{ij} that maps the true distribution, T_j, onto the measured (reconstructed) distribution, R_i:

$$R_i = \sum_j M_{ij} T_j. \tag{7.5}$$

The migration matrix gives the fraction of events in bin T_j that are measured in bin R_i. The matrix is determined from a Monte Carlo training sample. The measured distribution is a convolution of the true distribution and detector effects. The unfolding procedure applies an inverted migration matrix to the measured data, while taking into account the uncertainties from statistical fluctuations in the data.

The iterative and SVD unfolding algorithms require a regularization parameter to prevent statistical fluctuations in the data from creating large oscillations in the true distribution. The regularization parameter determines the allowed amount of smoothing, placing either more weight on the data or on the training sample truth. For the iterative algorithm, the regularization parameter specifies the number of iterations (from zero iterations for training sample truth). The regularization parameter is chosen as $k_{\mathrm{Bayes}} = 4$, as it gives the best results on MC closure tests.

Choosing the regularization parameter for SVD, k_{SVD}, effectively determines the relative weight for the training sample and the data. Too small a value for k_{SVD} may bias the unfolding result towards the MC truth input, while a value that is too large may give a result that is dominated by statistical fluctuations with no physical basis. Numerically, the regularization parameter k_{SVD} has to lie between one and the number of bins. All unfolding results in this thesis use $k_{\mathrm{SVD}} = 5$ which corresponds to the number of bins in the distribution. With this value for k_{SVD}, no unphysical enhancement of the statistical fluctuations is observed; instead, the best error estimate is achieved. This case corresponds to a full inversion of the migration matrix. The full range of k_{SVD}, from 2 to 5, has been tested using two different MC samples. A value of $k_{\mathrm{SVD}} = 5$ gives the best recovery of MC truth from MC reconstructed.

The migration matrices used in the following unfolding procedure are derived from Monte Carlo samples with the following properties:

- MadGraph generator with Tune Z2 MC samples unless stated otherwise
- signal-only events ($W \to e\nu$ + jets) with pile-up plus pile-up subtraction applied, using the L1Fast algorithm used in data
- generated electron within the detector acceptance in η and $p_T^{gen} > 20$ GeV
- generated jet collection cleaned of jets that lie within $\Delta R = 0.3$ of generated electrons and passing the threshold in p_T^{gen} of 30 GeV
- reconstructed jets passing the ID requirements in Table 6.4, cleaned of jets lying within $\Delta R = 0.3$ of generated electrons and passing the threshold in p_T^{reco} of 30 GeV

These requirements ensure that the unfolding procedure mimics the data in which background has been subtracted and in which lepton efficiencies and pile-up have been corrected for. Figure 7.16 shows the migration matrix used to unfold the data. The bins correspond to $n = 0, 1, 2, 3$ and $n \geq 4$ with generated jets on the Y-axis and reconstructed jets on the X-axis. While the overall normalization of the migration matrix does not matter in the unfolding procedure, to improve readability the matrix is shown normalized to each row (number of generated jets) and normalized to the integrated luminosity.

The performance of the unfolding algorithms is first validated with the help of Monte Carlo studies. Next, the performance of the unfolding algorithm on data is investigated, showing that it does not introduce any additional bias when comparing

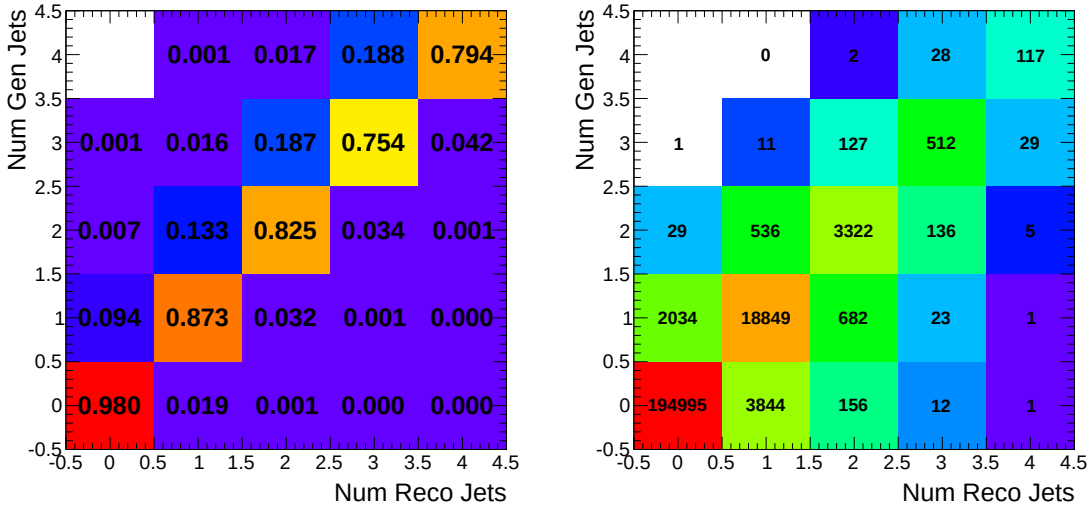


Figure 7.16: $W \to e\nu$ + jets migration matrix in bins of $n = 0, 1, 2, 3$ and $n \geq 4$ jets and for jet E_{T} cuts 30 GeV, scaled to the number of generated jets in each row (left) or scaled to luminosity, 36 pb^{-1} (right).

data before and after unfolding to signal Monte Carlo samples after reconstruction and at the generator level. Finally, the systematic uncertainty introduced by the unfolding procedure is studied by comparing a variety of unfolding algorithms and migration matrices from different signal Monte Carlo samples.

7.3.1 Validation of Unfolding Procedure Using Monte Carlo

Before applying the unfolding procedure to data, it must first be tested on similar Monte Carlo samples. Three different types of MC tests are employed. All make use of the migration matrices obtained as described above.

1. Unfold the reconstructed jet multiplicity distribution obtained from the same signal Monte Carlo that was used to derive the migration matrix (MadGraph TuneZ2).

   Two smaller subsamples were created from the MadGraph TuneZ2 MC.

One subsample, with roughly the same number of events as selected in the data, was used for the unfolded distribution, and a larger, independent subsample was used for the migration matrix.

2. Unfold the distribution obtained from a signal Monte Carlo that is different from the one used to derive the migration matrix.

3. Unfold a distribution from the selected MC events that have been efficiency-corrected in a similar manner as the data.

Two unfolding closure tests performed on Monte Carlo samples are shown in Figures 7.17. The uppermost of the joined plots compares the generated, the reconstructed and the unfolded jet multiplicity distribution for multiplicities $n = 0, 1, 2, 3$ and $n \geq 4$. The lower plot shows the ratio of reconstructed over generated distribution and the ratio of unfolded over generated distribution. The errors correspond to the statistical errors one would obtain if numerator and denominator were statistically independent. The result of the SVD unfolding algorithm, applied to a MadGraph TuneZ2 MC sample after the full selection has been made and after being corrected for the selection efficiency, is presented in Figure 7.17 on the left. The result in this case is near unity, as one would expect when using the same MC sample. Figure 7.17 on the right shows the result of the SVD unfolding algorithm on a MadGraph TuneD6T W+jets Monte Carlo sample using MadGraph TuneZ2 for migration matrix. As expected, the ratio of unfolded over generated multiplicity distribution is not perfect but is an improvement relative to the reconstructed over generated distribution. The closure tests demonstrate that the unfolding procedure does not introduce any additional bias.

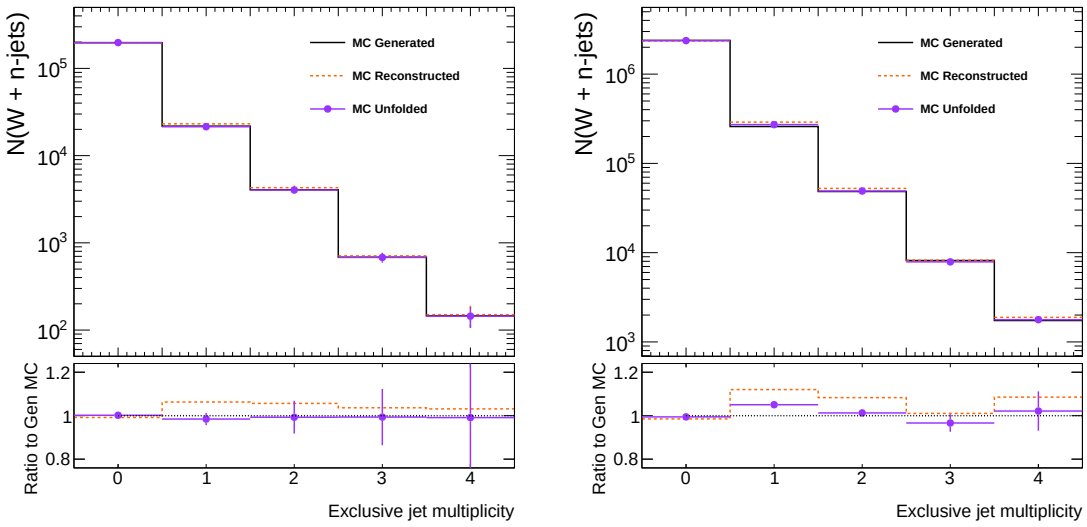


Figure 7.17: Unfolding validation on Monte Carlo samples: Results after SVD unfolding for jet $p_T > 30$ GeV. The migration matrix is derived from $W \to e\nu$ + jets MadGraph TuneZ2 signal Monte Carlo. Unfolding algorithm is applied to a data-sized $W \to e\nu$ + jets MadGraph TuneZ2 sample after full selection and efficiency corrections (left) and to a larger MadGraph TuneD6T signal sample (right).

7.3.1.1 Application of unfolding procedure to data

Unfolded data results are presented in Figure 7.18. The data, taken from Table 7.1, N_{effcorr}, are corrected for background, selection efficiencies and pile-up before unfolding. The uppermost of the two joined plots compares the Monte Carlo generated, the data reconstructed, and the data unfolded jet multiplicity distribution for multiplicities $n = 0, 1, 2, 3$ and $n \geq 4$. The lower plot shows the ratio of data reconstructed over Monte Carlo generated events and of data unfolded over Monte Carlo generated distributions. The errors correspond to the statistical errors one would obtain if numerator and denominator were statistically independent.

As an additional check that no bias is introduced by the unfolding procedure,

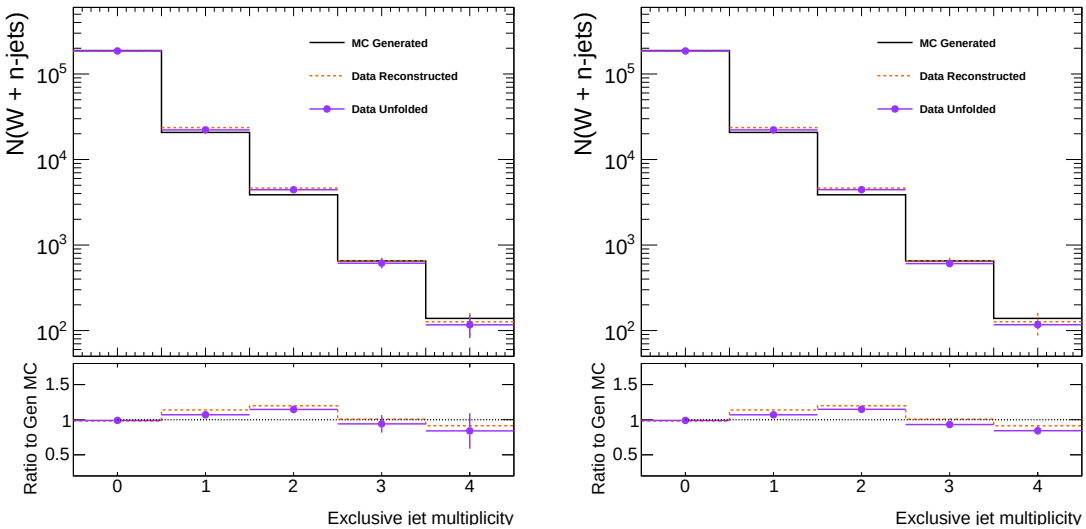


Figure 7.18: Unfolding validation on data: Results after SVD unfolding (left) and Bayes unfolding (right) for jet $p_T > 30$ GeV.

the χ^2 values are calculated for the deviation between unfolded data and generated Monte Carlo distribution, and for reconstructed data and reconstructed Monte Carlo distribution. The results are summarized in Table 7.2 and show that the values are of comparable magnitude.

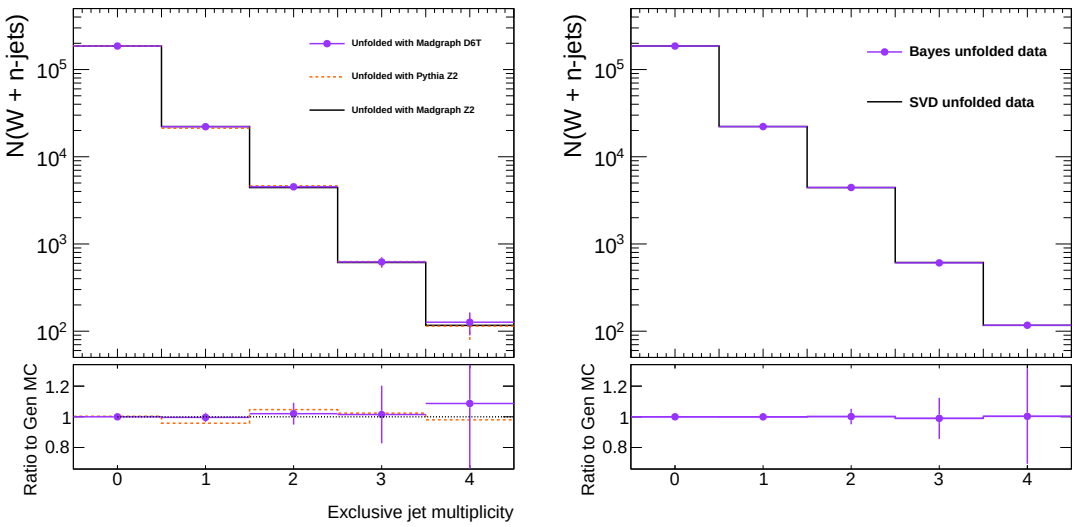
7.3.2 Systematic errors from unfolding

Systematic uncertainties in the unfolding are estimated by changing the tune or generator of the MC used to make the migration matrix, or by using different unfolding algorithms. Figure 7.19 shows the effect of using different unfolding algorithms and alternative Monte Carlo signal samples to derive the migration matrix. At a jet p_T cut of 30 GeV, only a weak dependence on the origin of the migration matrix is visible. A tendency to yield lower cross sections compared to the nominal results can be seen at higher jet multiplicities. The uppermost of the two joined plots on the left

PF jet $p_T > 30$ GeV		
MC Generator	χ^2/ndf values between data and Reco MC	χ^2/ndf values between unfolded data and Gen MC
MadGraph TuneZ2 SVD	134	175
MadGraph TuneZ2 Bayes	135	173
MadGraph TuneD6T SVD	158	252
pythia TuneZ2 SVD	681	1007

Table 7.2: χ^2/ndf values between fully reconstructed MC with only acceptance cuts and efficiency corrected data. χ^2/ndf values between generated MC with only acceptance cuts and unfolded, efficiency corrected data. Shown for various unfolding algorithms, and migration matrices from different generators, and tunes.

compares the data distributions that result after unfolding with migration matrices that are derived from three different signal Monte Carlo samples, MadGraph TuneZ2 [35, 71], MadGraph TuneD6T [35, 72] and pythia TuneZ2 [64, 71]. The lower of the two joined plots shows the ratio of the resulting data distributions when using a migration matrix derived from an alternative Monte Carlo sample over the distribution when using the MadGraph TuneZ2 derived migration matrix. The plots on the right have a similar format, but instead of changing the migration matrix, the unfolding algorithms SVD and Bayes are compared. Only a weak dependence on the unfolding algorithm is observed: SVD versus Bayes has less than a 1% effect. The results when using different regulation terms for the Bayes and SVD unfolding are shown in Appendix A. The effect of using pile-up (plus corrections) versus no-pile MC for the migration matrix is also shown in Appendix A.



(a) (b)

Figure 7.19: Unfolding systematics: (a) Data results after unfolding with migration matrices derived from different signal Monte Carlo samples compared to the nominal MadGraph TuneZ2. (b) Data results after unfolding with Bayes compared to unfolding with the nominal SVD algorithm, using migration matrix from MadGraph TuneZ2 for both. The two algorithms are nearly indistinguishable.

7.4 Theoretical Systematic Uncertainties

Some theoretical uncertainty estimates come from using Monte Carlo with different tunes for the underlying event. All of the distributions have been compared to MadGraph TuneZ2 and MadGraph TuneD6T. pythia has also been used for unfolding systematics, but in general does not deal sufficiently with multiple jets to be comparable to data. pythia does not include multiple partons in the hard scattering matrix element and so fewer jets pass the E_{T} threshold.

By taking ratios, any of the theoretical uncertainties that do not depend on jet multiplicity will cancel. Theoretical simulations are necessary to correct for the acceptance cuts, since events outside of the acceptance are undetected or poorly reconstructed. In this analysis the cross sections have not been adjusted for acceptance cuts in order to avoid dependence on the specific MC modeling used to make such corrections. In addition, leaving the result within the acceptance avoids the theoretical uncertainties that come with choosing specific parameters such as parton distribution function or renormalization scale.

7.5 Experimental Systematic Uncertainties

Systematic uncertainties for the cross-section versus number of jets can be divided into two types. Those uncertainties that affect the number of events selected, but do not shift events between n-jet bins, are considered uncorrelated. They include the uncertainty in the efficiency calculation, in the fitting procedure, and in the effect that pile-up has on event selection. Those uncertainties that result in shifting events between jet bins are considered correlated and are treated separately. They include

uncertainty in the jet energy corrections, in the pile-up effect on the jet energy, in the flavor of the jets, and in changes in the reconstruction code.

The systematic uncertainty in the efficiency comes from comparing tag and probe results when using two different line-shapes to describe the Z-peak. Details are given in Section 6.4. The systematic uncertainties in the fits are obtained by varying the fixed parameters of the functions (pdfs) and also by comparing the nominal fits to template-only fits. The floating parameter uncertainties are covered by the statistical error from the fit. The fixed b-tagging efficiency, ϵ_b, and mis-tag rate, ϵ_{nob}, present one of the larger uncertainties, but are mostly relevant in the three and four jet bins. The fits are performed after shifting ϵ_b and ϵ_{nob} up and down by their uncertainties. The differences in the signal yield from the nominal fit value are the uncertainties in the b-tag fit. For those parameters that are fixed in the M_{T} fits, i.e., all of those for the top distribution and some for the 2-jet through 4-jet bins for the W+jets distribution, the fitting is performed again after shifting the fixed parameters based on their uncertainties. The fits were also performed using different models for the QCD+others backgrounds with partial inversions of the ID variables and with the initial parameterization using Monte Carlo. The changes in the events yield from the alternative background modelling are included in the overall systematic uncertainty. Lastly, fits are done using MC templates for M_{T}, rather than fitting with functions. This method is much less flexible and relies on the shape of the QCD MC. Nonetheless, it gives reasonable results and is used for a conservative estimate of the uncertainty in the fit.

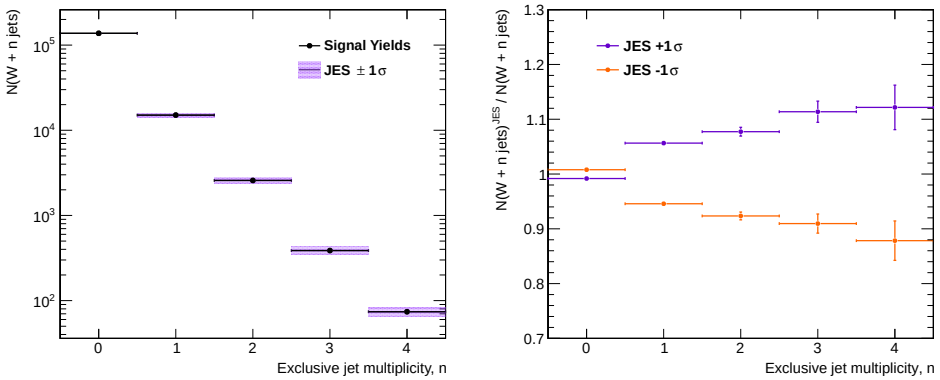
The effect on the n-jet counting due to uncertainties in the jet energy scale comes

jets	Signal M_{T}	Top M_{T}	b-tag	bkgd model	total
$= 0$	–	0.0%	0.00%	0.1%	0.10%
$= 1$	–	0.0%	0.05%	0.8%	0.80%
$= 2$	0.7%	0.2%	0.26%	1.0%	1.26%
$= 3$	1.1%	2.8%	2.87%	0.3%	4.16%
≥ 4	1.1%	5.7%	6.80%	0.5%	8.95%

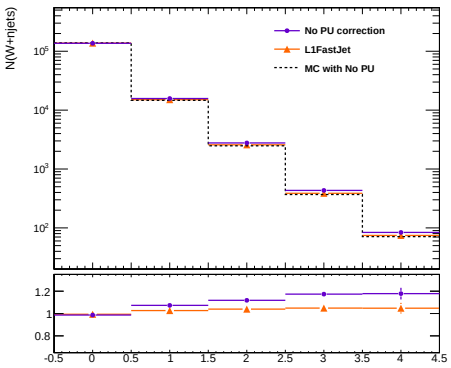
Table 7.3: Relative systematic uncertainties from the signal extraction fit.

from shifting the jet E_{T} by $\pm 1\sigma$. Part of the jet E_{T} uncertainty comes from uncertainty on the corrections made to the energy scale [95]. The correction uncertainty is p_{T} and η dependent and is roughly 2-3% for most jets. Uncertainty in the pile-up is included as a function of jet E_{T} and is around 1.2% for a 30 GeV jet. There is an additional uncertainty of 2-3% for b-flavored jets. A flat 1.5% uncertainty is assigned to all jets for changes in software release and calibrations. All of these uncertainties are added in quadrature, leading to a total jet E_{T} uncertainty of about 4% for a 30 GeV jet. They are used to adjust the jet E_{T} up or down in Monte Carlo. The number of jets with $E_{\mathrm{T}} > 30$ GeV are counted and the differences to the nominal jet multiplicity distribution are compared. An additional pile-up uncertainty has been added on top of the jet energy scale (JES) uncertainties to account for imperfect corrections using the L1FastJet algorithm. Figure 7.20 shows the n-jet distribution with error bands, the ratio of the $\pm 1\sigma$ scaling to the nominal, and the effect of pile-up corrections. The correlated and uncorrelated uncertainties in the n-jet distribution are shown in Table 7.4.

The 4% luminosity uncertainty cancels out when calculating the ratio of n-jets to (n-1)-jets and is not included in the final results.



(a) (b)



(c)

Figure 7.20: Jet multiplicity with JES uncertainties for jet multiplicity (a), the ratio of the shifted n-jet distribution relative to the nominal (b), and comparison between MC with pile-up and MC with pile-up plus L1FastJet corrections to MC with no pile-up (c).

jets	Efficiency	$\not\!\!E_T$	Fitting	JES +1 σ	JES -1 σ
$= 0$	0.53	0.1	0.10	1.02	1.06
$= 1$	0.27	0.3	0.80	6.2	6.5
$= 2$	0.83	0.5	1.26	9.0	9.0
$= 3$	1.36	0.5	4.16	10.6	12.9
≥ 4	2.74	1.4	8.95	13.1	14.4

Table 7.4: Relative uncertainty (%) on the exclusive jet multiplicity from efficiency, $\not\!\!E_T$ resolution, fit, and jet energy scale (JES) uncertainties.

Chapter 8

Results

8.1 Comparisons to Theory

After extracting the signal and background yields, the M_{T} distribution is presented again using information from the extraction to scale the Monte Carlo signal and backgrounds to the expected values. The data and MC are also corrected for the selection efficiency, and are shown in Figure 8.1. The W p_{T} and first, second, and third jet p_{T} are likewise scaled as seen in Figures 8.2, 8.3 and 8.4. The MC has been corrected using the MC selection efficiency, and the data has been corrected from the combined tag & probe and MC efficiencies, Table 6.8. With the scaling factors from the fits, the signal (background) MC describes the data well, although agreement drops slightly at higher jet multiplicities where the top events become more significant and the statistics are much lower. This generally good agreement indicates that the current theory, as implemented in MadGraph, provides a reasonable prediction of the behavior of W+jets. In addition, the agreement validates the use of the MadGraph MC in the efficiency calculation and fit parameterization used for the cross section measurement.

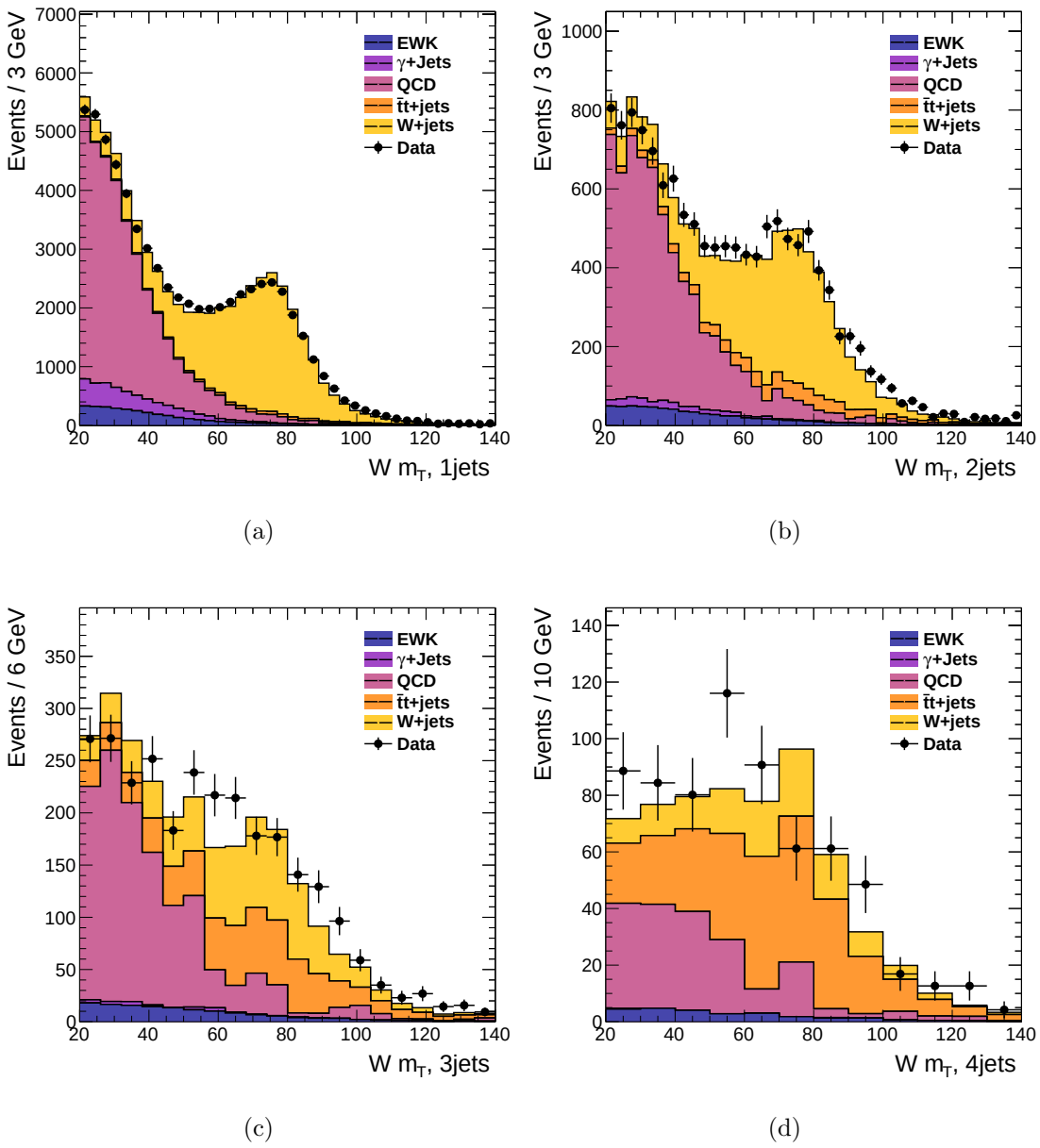


Figure 8.1: W m_T comparison between data and MadGraph +pythia Z2 MC scaled to the results of the fit, for $\geq$ 1, 2, 3, 4 PF jets ($E_{\mathrm{T}} > 30$ GeV), with a cut of $M_{\mathrm{T}} > 20$ GeV for the W candidate.

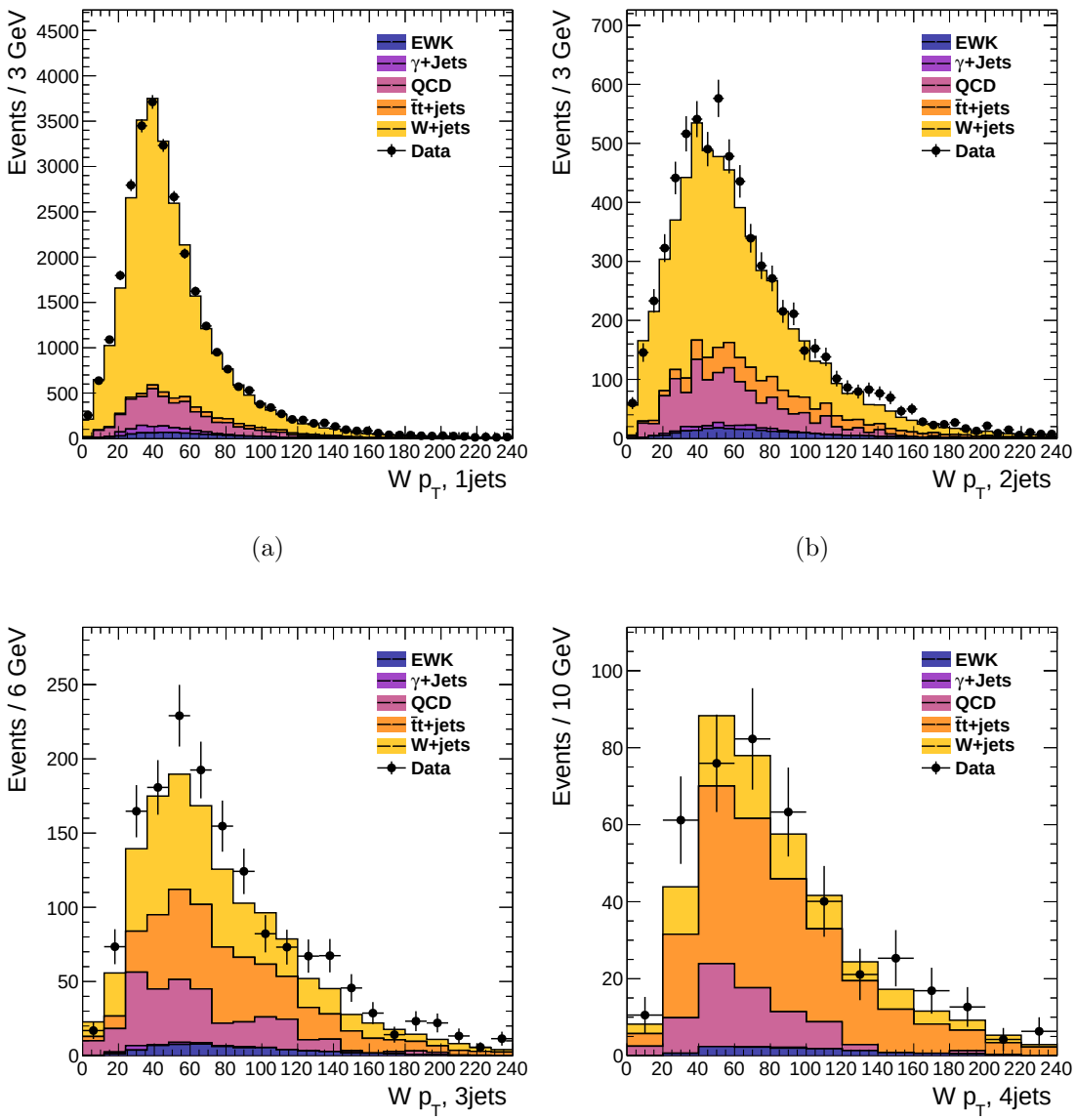


Figure 8.2: W p_{T} comparison between data and MadGraph +pythia Z2 MC scaled to the results of the fit, for $\geq$ 1, 2, 3, 4 PF jets ($E_{\mathrm{T}} > 30$ GeV), with a cut of $M_{\mathrm{T}} > 50$ GeV for the W candidate.

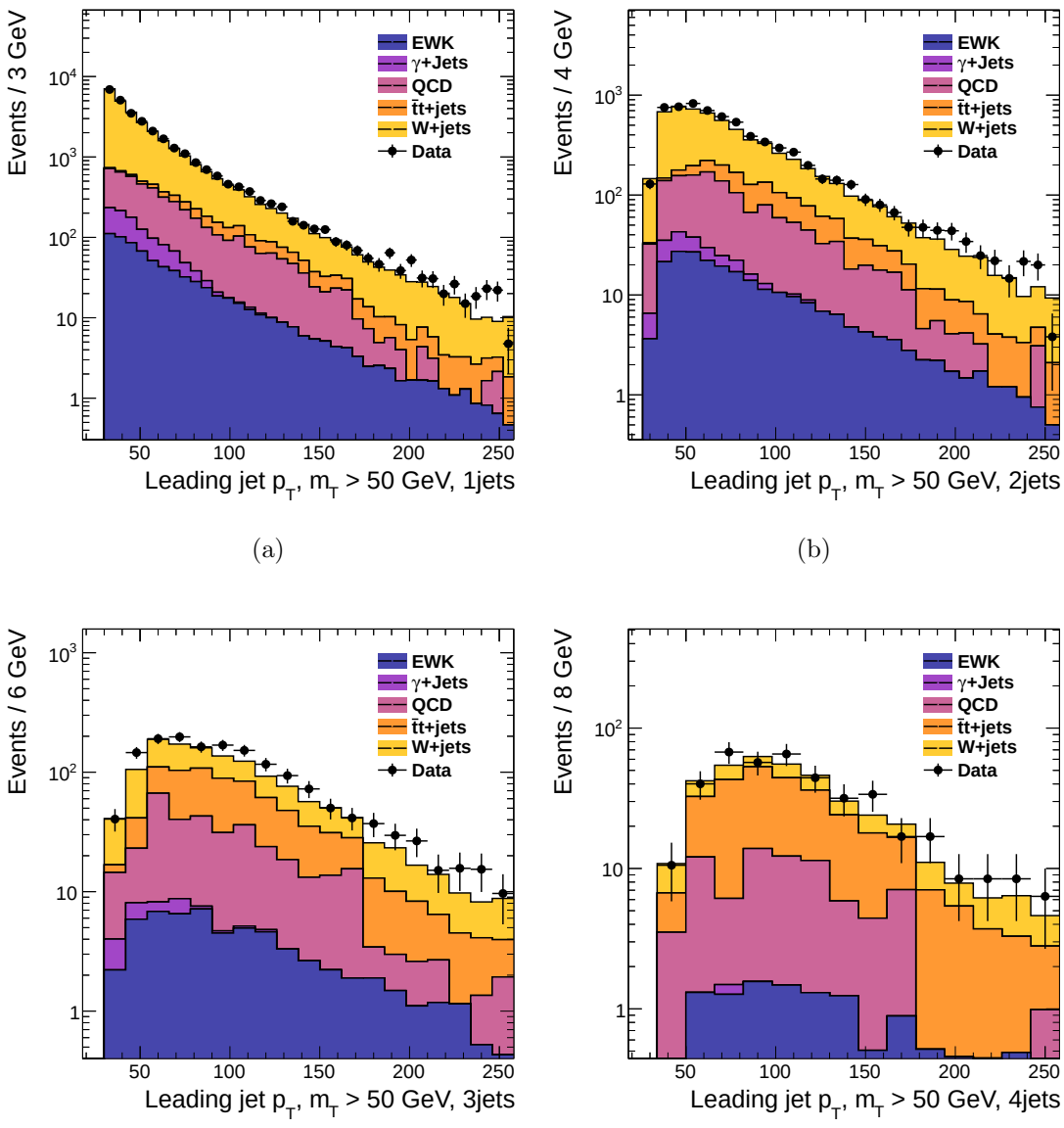
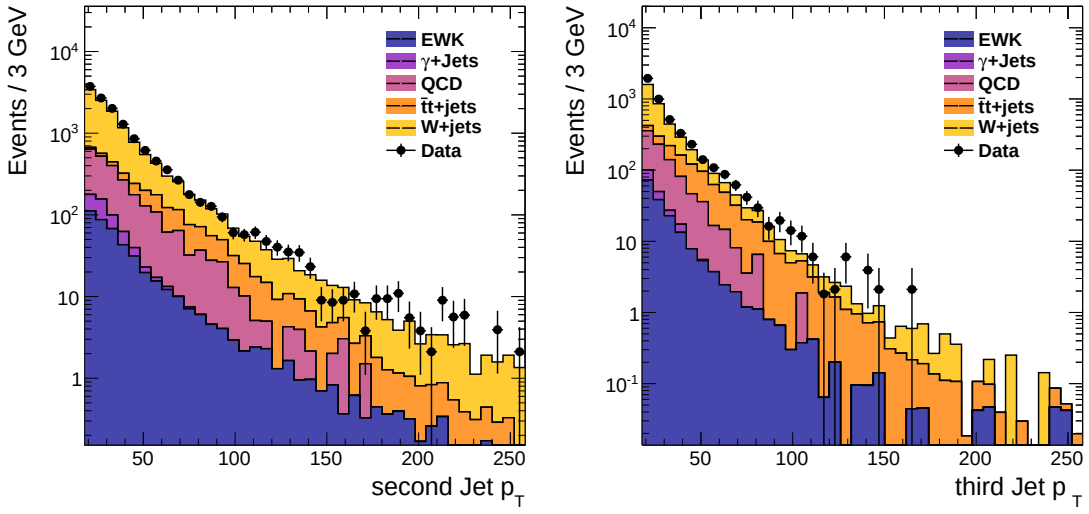


Figure 8.3: Jet p_{T} comparison between data and MadGraph +pythia Z2 MC scaled to the results of the fit, for $\geq$ 1, 2, 3, 4 PF jets ($p_{\mathrm{T}} > 30$ GeV), with a cut of $M_{\mathrm{T}} > 50$ GeV for the W candidate.



(a) (b)

Figure 8.4: Second (a) and third (b) jet p_{T} comparison between data and MadGraph +pythia Z2 MC scaled to the results of the fit, for ≥ 1 PF jets ($p_{\mathrm{T}} > 30$ GeV), with a cut of $M_{\mathrm{T}} > 50$ GeV for the W candidate.

8.2 Events and Cross section by Jet Multiplicity

Presented here are the final results of the W+jets cross section after selection, signal extraction, efficiency corrections, and unfolding. The cross sections are not corrected for the acceptance of the detector, which necessarily relies on the use of a specific Monte Carlo simulation. This strategy was chosen in part to avoid the theoretical uncertainty in the model used to calculate the acceptance and to present results without reliance on any one specific MC generator. The results are given as a ratio of $\sigma(W + n\text{-jets})$ to $\sigma(W)$ and $\sigma(W + n\text{-jets})$ to $\sigma((W + n\text{-}1)\text{-jets})$ to reduce systematic uncertainties. The luminosity uncertainty completely cancels because the overall normalization divides to one. Any errors in the electron selection efficiency that are not a function of the number of jets, but rather from the overall scale, will likewise

cancel. Much of the jet energy scale uncertainty also cancels. The cross sections are given within the acceptance, calculated according to

$$\sigma_{\rm acc} = \frac{N_{\rm sel} - N_{\rm bkgd}}{\epsilon_{\rm tot}\ L} \tag{8.1}$$

where $N_{\rm sel} - N_{\rm bkgd}$ is the signal yield extracted from the fits after event selection, $\epsilon_{\rm tot}$ is the selection efficiency relative to the acceptance and L is the integrated luminosity, 36.1 ± 1.4 pb^{-1}.

Table 8.1 presents the unfolded results in exclusive jet multiplicities, n-jets. $N_{\rm obs}$ is the number of W+jets candidates after event selection and background subtraction. The error for $N_{\rm obs}$ includes both statistical and systematic uncertainties provided by the signal extraction fit. The electron efficiencies are taken from Table 6.8 and the error is the combined statistical and systematic. $N_{\rm effcor}$ is the result obtained when correcting $N_{\rm obs}$ for efficiencies. The errors on $N_{\rm effcor}$ take the statistical and systematic uncertainties of $N_{\rm obs}$, as well as the uncertainties from the efficiency correction factors, into account. $N_{\rm unf}$ is the result obtained after applying the SVD unfolding algorithm with a response matrix derived from the MadGraph TuneZ2 signal Monte Carlo sample. The errors are the result of propagating the statistical and uncorrelated systematic errors through the unfolding algorithm. The last three columns show the deviations from the nominal result when using an alternative unfolding algorithm or when using response matrices derived from alternative Monte Carlo generator and tune signal samples. The uncertainty quoted is based on the difference in the result between using these matrices and using the nominal migration matrix, MadGraph TuneZ2 . The unfolded cross section results (within acceptance), $\sigma_{\rm acc}$, are also given by inclusive jet multiplicity in Table 8.2, with statistical, uncorrelated systematic,

and correlated systematic uncertainties. For events with at least one jet, the largest uncertainty in the cross section ratios comes from uncertainty in the jet energy scale (JES). For the inclusive cross section the statistical error dominates.

Tables 8.3 and 8.4 provide the final cross section ratios and all of the uncertainties that have been considered. Figure 8.5 displays a plot of the cross section ratios given in the tables. Both ratios are compared to MadGraph TuneZ2, MadGraph TuneD6T, and pythia TuneZ2 Monte Carlo simulations. The cross sections for $\geq$ 1-jet and $\geq$ 2-jet events relative to the total cross section are slightly overestimated in data relative to MadGraph MC, while the $\geq$ 3, 4-jet events are slightly underestimated. The low statistics and high top quark contamination for the higher jet multiplicities made both signal extraction and data driven efficiency calculations more difficult, hence the presence of larger uncertainties. Overall, good agreement is seen between the data and MadGraph MC. Agreement with pythia alone is not expected, since pythia does not handle multiple jets in the matrix element. The observed agreement between MadGraph and this Standard Model measurement indicates the usefulness of MadGraph in future searches for new physics.

PF jet $p_T > 30$ GeV							
					Unfolding systematic deviation		
n jets	N_{obs}	ϵ_{tot}	N_{effcor}	N_{unf}	SVD - Bayes	MC generator	MC tune
0	131376 ± 423	0.713 ± 0.0049	184258 ± 1399	185946 ± 1525	4.0	697.0	-26.0
1	15476 ± 189	0.655 ± 0.00624	23627 ± 366	22198 ± 473	-7.2	-926.8	-84.9
2	2730 ± 81.6	0.589 ± 0.0115	4635 ± 165	4433 ± 217	7.6	208.1	90.4
3	362 ± 38.1	0.551 ± 0.0269	657 ± 76	613 ± 81	-6.2	14.7	9.1
4	60 ± 17.8	0.474 ± 0.0421	127 ± 39	117 ± 35	0.4	-2.3	10.1

Table 8.1: N_{obs} are the results from the signal extraction, N_{effcor} are the results after correcting for electron efficiency, ϵ_{tot}, and N_{unf} are the results after unfolding, all with with exclusive jet counting. The last three columns represent the deviation from the nominal unfolding results when changing the algorithm, the MC generator, and the MC tune, respectively.

PF jet $p_T > 30$ GeV								
n jets	σ	stat	stat+sys	JES syst error		Unfolding systematic deviation		
	in acceptance			(±)		SVD - Bayes	MC generator	MC tune
≥ 0 jets	5909	33.4	44.7	2.50	2.92	-0.04	-0.26	-0.04
≥ 1 jets	758	12.8	14.6	60.0	62.7	-0.15	-19.6	0.68
≥ 2 jets	143	5.92	6.49	14.2	14.6	0.05	6.11	3.04
≥ 3 jets	20.2	2.30	2.44	2.36	2.88	-0.16	0.34	0.53
≥ 4 jets	3.23	0.91	0.97	0.44	0.51	0.01	-0.06	0.28

Table 8.2: Results for cross section $\sigma\,(\geq n\,\text{jets})$ within the acceptance with inclusive jet counting. Sources of uncertainty shown are statistical, statistical + uncorrelated systematics (fit and efficiency), correlated systematics (jet energy scale, JES), and deviations when using different unfolding methods (algorithm, generator, and tune). There is also an overall 4% uncertainty for the luminosity.

PF jet $p_T > 30$ GeV								
n jets	σ ratio	stat	stat+sys	JES syst error		Unfolding systematic deviation		
	in acceptance			(±)		SVD - Bayes	MC generator	MC tune
≥ 1 / ≥ 0 jets	0.128	0.002	0.00234	0.0101	0.0106	-2.47e-05	-0.00331	0.000117
≥ 2 / ≥ 0 jets	0.0242	0.000987	0.00109	0.00239	0.00246	8.33e-06	0.00103	0.000514
≥ 3 / ≥ 0 jets	0.00342	0.000388	0.000413	0.000397	0.000486	-2.75e-05	5.83e-05	9.02e-05
≥ 4 / ≥ 0 jets	0.000547	0.000155	0.000164	7.35e-05	8.63e-05	1.73e-06	-1.08e-05	4.75e-05

Table 8.3: Results for cross section ratio $\sigma(W+ \geq n\,\text{jets})/\sigma(W)$ within the acceptance with inclusive jet counting. Sources of uncertainty shown are statistical, statistical + uncorrelated systematics (fit and efficiency), correlated systematics (jet energy scale, JES), and deviations when using different unfolding methods (algorithm, generator, and tune).

PF jet $p_T > 30$ GeV								
n jets	σ ratio in acceptance	stat	stat+sys	JES syst error (±)		Unfolding systematic deviation		
						SVD - Bayes	MC generator	MC tune
≥ 1 / ≥ 0 jets	0.128	0.002	0.00234	0.0101	0.0106	-2.47e-05	-0.00331	0.000117
≥ 2 / ≥ 1 jets	0.189	0.00694	0.00767	0.00351	0.004	0.000101	0.0133	0.00383
≥ 3 / ≥ 2 jets	0.141	0.0148	0.0158	0.00223	0.00636	-0.00118	-0.00349	0.000708
≥ 4 / ≥ 3 jets	0.16	0.0415	0.044	0.0026	0.00292	0.0018	-0.00577	0.00941

Table 8.4: Results for cross section ratio $\sigma(W+ \geq n \text{ jets})/\sigma(W+ \geq (n-1) \text{ jets})$ within the acceptance with inclusive jet counting. Sources of uncertainty shown are statistical, statistical + uncorrelated systematics (fit and efficiency), correlated systematics (jet energy scale, JES), and deviations when using different unfolding methods (algorithm, generator, and tune).

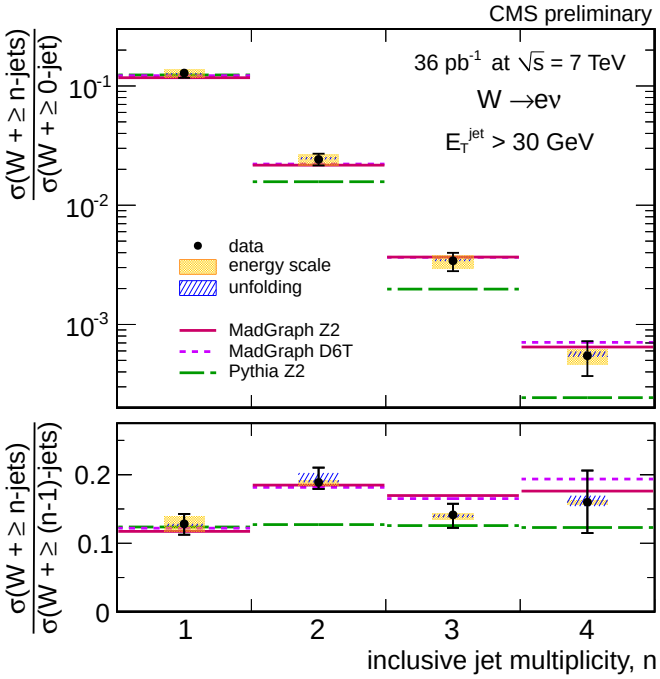


Figure 8.5: The cross section ratios $\sigma(W+ \geq n \text{ jets})/\sigma(W+ \geq (n-1) \text{ jets})$ and $\sigma(W+ \geq n \text{ jets})/\sigma(W)$ compared to expectations from MadGraph and pythia.

Chapter 9

Conclusion

There is a theory which states that if ever anybody discovers exactly what the Universe is for and why it is here, it will instantly disappear and be replaced by something even more bizarre and inexplicable. There is another theory which states that this has already happened.

— Douglas Adams

9.1 Summary

A complete strategy for reconstructing, selecting, analyzing, and calculating the n-jet cross section ratios for W+jets events is presented. Data events are fully reconstructed with CMS software and compared to theoretical models. The reconstruction algorithm focuses on high energy electrons such as those from W bosons. The $\not\!\!E_T$ is reconstructed with the best possible resolution for an accurate measurement of the W transverse mass. Jets are reconstructed with an algorithm designed to use all aspects of the CMS detectors and are corrected using data-driven techniques for a precise measurement of the energy of the original parton. The electron selection process is designed specifically for $W \to e\nu$ events and removes over 99.1% of the background events. The full selection, including a Z boson veto and $M_T^W \geq 20$ GeV, removes over

99.6% of the background events, with a total selection efficiency of approximately 68% for the signal. The remaining background is estimated using a functional fit to the M_T^W and b-tagged distributions for exclusive jet bins of 0-4 jets. It provides a reasonable measurement of the number of W+jets, top quarks, and a combination of QCD, γ+jet, and electroweak events. Efficiency corrections to the signal yield were made using a combination of Monte Carlo and data-driven techniques to recover the cross section within the acceptance. The cross section as a function of the number of jets was then unfolded for detector effects to recover an n-jets cross section that is not detector dependent.

Finally, the cross section ratios $\sigma(W+ \geq n \text{ jets})/\sigma(W+ \geq (n-1) \text{ jets})$ and $\sigma(W+ \geq n \text{ jets})/\sigma(W)$ are presented and compared to theoretical models. There is a lack of agreement with the LO PYTHIA, as expected, because PYTHIA does not include multiple hard partons in the matrix element calculation. The measured results show good agreement with the MADGRAPH TuneZ2 predictions within the statistical and systematic uncertainties. The agreement of the n-jet cross section predictions with the data indicates that current perturbative QCD simulations are reasonable.

9.2 Outlook

Future studies with higher statistics will make possible precision measurements of the absolute cross section of the number of jets produced in association with W bosons. More events will mean smaller errors and less need to use ratios to cancel out many of the systematics. Measurements involving higher jet multiplicities will be less limited by statistics. It will also be possible to show the unfolded cross section as a function of jet E_{T}, for events with 1-4 jets, as has been done at lower energies

at previous experiments [41, 42]. The study of W+jets serves as a starting point for searches for new physics, since it presents one of the larger backgrounds for many supersymmetric models and Higgs studies. More accurate measurements of Standard Model processes will help differentiate any new physics signals.

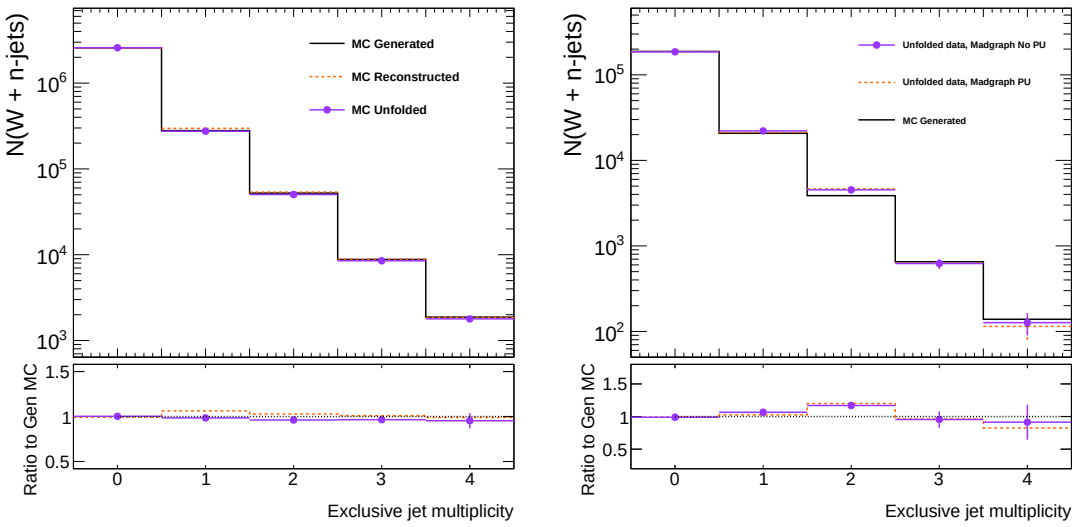
Appendix A

Unfolding closure tests on Monte Carlo

This section contains further unfolding closure tests and systematics for $W \to e\nu$+jets, performed on signal Monte Carlo samples generated with PYTHIA [64] and MADGRAPH [35], with Tunes Z2 [70] and D6T [72] and with and without pile-up.

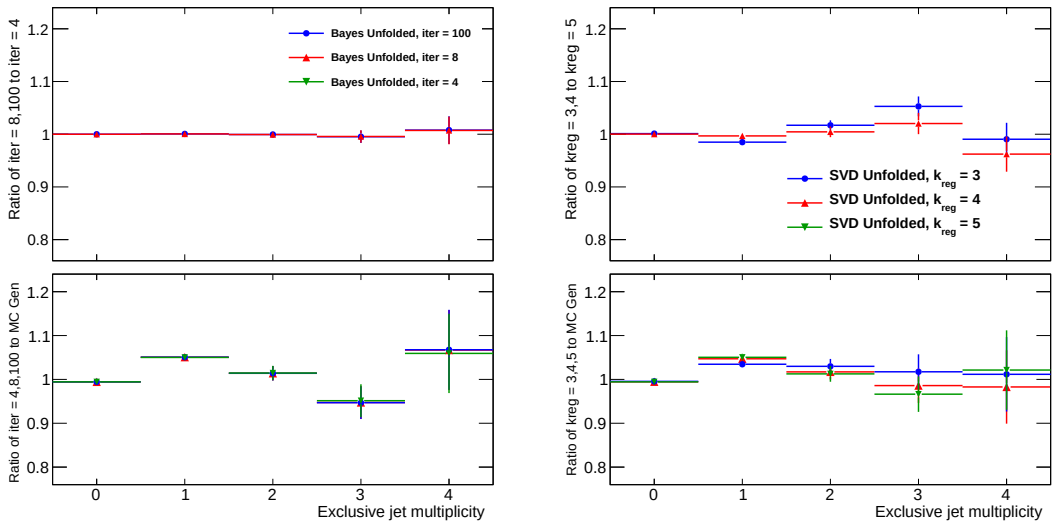
A study of the effect of pile-up on unfolding is shown in Figure A.1. The left plot (a) starts from a matrix with pile-up plus pile-up corrections and unfolds a MC distribution without pile-up. The upper plot shows the n-jet distribution before and after unfolding along with the generator level jet distribution. The lower plot shows a comparison of the pre and post unfolded distributions compared to the generator level. The pile up has a very small effect. The right plot (b) compares the unfolded data distribution to the generator level when using a response matrix with pile-up+corrections and when using a matrix without pile-up. The difference in unfolded results between using a pile-up and a no pile-up response matrix remains small over all jet bins, increasing to about an 8% difference in the 4th jet bin. The bins correspond to $n = 0, 1, 2, 3$ and $n \geq 4$ jets.

As a test of how the parameters of the Bayes (SVD) unfolding algorithms (see Section 7.3) affect the result, we have changed the regularization parameter to 8 and 100 (3 and 4) and compared them to the results using $k_{\text{reg}} = 4$ (5). Figures A.2 and A.3 shows the results of these studies on MC and data, respectively. The uppermost



(a) (b)

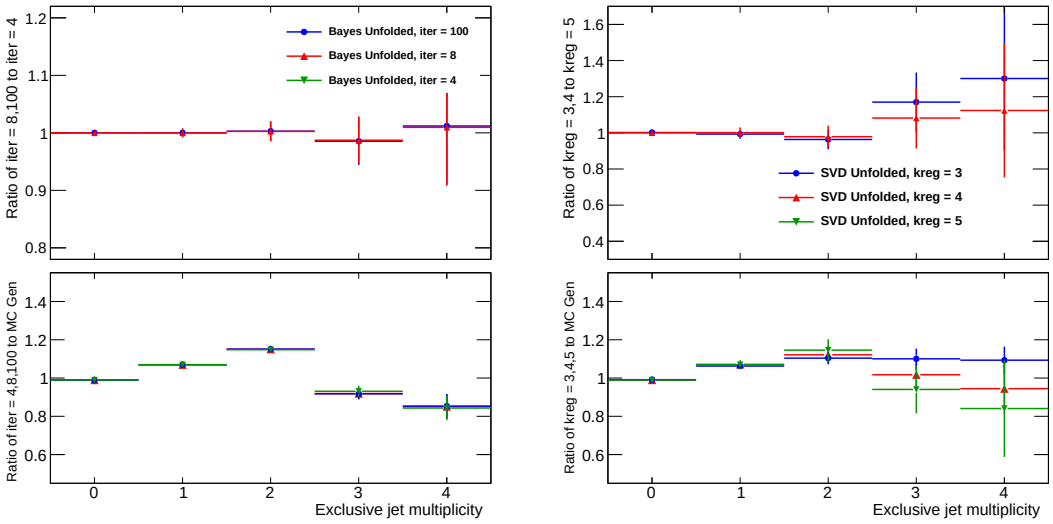
Figure A.1: Pile-up effects in unfolding: (a) unfolding a distribution without pile-up using a response matrix with pile-up+corrections, (b) unfolding data using a matrix with pile-up+corrections and without pile-up. Magraph TuneZ2 Monte Carlo was used for all matrices.



(a) (b)

Figure A.2: (a) Unfolding systematics from changing number of iterations when using the Bayes algorithm on MC. Upper plot compares 8 or 100 iterations to 4 iterations. Lower plot compares 4, 8, and 100 iterations to generator level MC. (b) Unfolding systematics from changing k_{SVD} when using the SVD algorithm on MC. Upper plot compares $k_{\mathrm{SVD}} = 3$ and 4 to $k_{\mathrm{SVD}} = 5$ iterations. Lower plot compares 3, 4, and 5 iterations to generator level MC. Response matrix is derived from Madgraph TuneZ2 and the unfolded distribution is from Madgraph TuneD6T in both cases.

of the two joined plots is a ratio of the data results unfolded with Bayes (SVD) with either 8 or 100 iterations ($k_{\mathrm{SVD}} = 3$ or 4), and results unfolded with Bayes (SVD) with 4 iterations ($k_{\mathrm{SVD}} = 5$). The lower plot is the ratio of the results unfolded and the MC generator prediction. The effect of changing the number of iterations when unfolding with Bayes or the k_{SVD} when unfolding with SVD is on the order of 2% for electrons.



(a) (b)

Figure A.3: (a) Unfolding systematics from changing number of iterations when using the Bayes algorithm on data. Upper plot compares 8 and iterations to 4 iterations. Lower plot compares 4, 8, and 100 iterations to generator level MC. (b) Unfolding systematics from changing k_{SVD} when using the SVD algorithm on data. Upper plot compares $k_{\mathrm{SVD}} = 3$ and 4 to $k_{\mathrm{SVD}} = 5$ iterations. Lower plot compares 3, 4, and 5 iterations to generator level MC. Response matrix is derived from Madgraph TuneZ2 in both cases.

Bibliography

[1] Glashow, S. L., Iliopoulos, J. & Maiani, L. Weak interactions with lepton-hadron symmetry. *Phys. Rev. D* **2**, 1285–1292 (1970).

[2] Barr, S. & Langacker, P. Superweak gauge theory of *cp*-invariance violation. *Phys. Rev. Lett.* **42**, 1654–1658 (1979).

[3] Salam, A. Weak and Electromagnetic Interactions Originally printed in *Svartholm: Elementary Particle Theory, Proceedings Of The Nobel Symposium Held 1968 At Lerum, Sweden*, Stockholm 1968, 367-377.

[4] Nakamura, K. *et al.* (Particle Data Group). *J. Phys. G* **37**, 075021 (2010).

[5] Tomonaga, S. On a relativistically invariant formulation of the quantum theory of wave fields. *Progress of Theoretical Physics* **1**, 27–42 (1946). URL `http://ptp.ipap.jp/link?PTP/1/27/`.

[6] Feynman, R. P. Mathematical formulation of the quantum theory of electromagnetic interaction. *Phys. Rev.* **80**, 440–457 (1950).

[7] Weinberg, S. A Model of Leptons. *Phys. Rev. Lett.* **19**, 1264–1266 (1967).

[8] Salam, A. Gauge unification of fundamental forces. *Rev. Mod. Phys.* **52**, 525–538 (1980).

[9] Glashow, S. L. Towards a Unified Theory: Threads in a Tapestry. *Rev. Mod. Phys.* **52**, 539–543 (1980).

[10] Higgs, P. W. Broken symmetries, massless particles and gauge fields. *Phys. Lett.* **12**, 132 (1964).

[11] Englert, F. & Brout, R. BROKEN SYMMETRY AND THE MASS OF GAUGE VECTOR MESONS. *Phys. Rev. Lett.* **13**, 321–322 (1964).

[12] Guralnik, G. S., Hagen, C. R. & Kibble, T. W. B. GLOBAL CONSERVATION LAWS AND MASSLESS PARTICLES. *Phys. Rev. Lett.* **13**, 585–587 (1964).

[13] Griffiths, D.J. *Introduction to Elementary Particles* (New York: John Wiley & Sons, 1987).

[14] Wilson, K. G. CONFINEMENT OF QUARKS. *Phys. Rev.* **D10**, 2445–2459 (1974).

[15] F. Halzen, A. Martin. *Quarks and Leptons: An Introductory Course in Modern Particle Physics* (John Wiley & Sons, Inc., 1984).

[16] Dorn, G., Pervushin, V. N. & Ebert, D. Contributions of sea quarks to deep inelastic scattering in two-dimensional quantum chromodynamics. *Theoretical and Mathematical Physics* **39**, 486–495 (1979). URL `http://dx.doi.org/10.1007/BF01017995`. 10.1007/BF01017995.

[17] Weinberg, S. *The Quantum Theory of Fields* (Cambridge University Press., 1995).

[18] Campbell, J. M., Huston, J. W. & Stirling, W. J. Hard Interactions of Quarks and Gluons: A Primer for LHC Physics. *Rept. Prog. Phys.* **70**, 89 (2007). Updated plot at http://projects.hepforge.org, `hep-ph/0611148`.

[19] Arnison, G. *et al.* Recent Results on Intermediate Vector Boson Properties at the CERN Super Proton Synchrotron Collider. *Phys. Lett.* **B166**, 484–490 (1986).

[20] Ansari, R. *et al.* Measurement of the Standard Model Parameters from a Study of W and Z Bosons. *Phys. Lett.* **B186**, 440–451 (1987).

[21] Coleman, S. & Weinberg, E. Radiative corrections as the origin of spontaneous symmetry breaking. *Phys. Rev. D* **7**, 1888–1910 (1973).

[22] Kane, G. L., Repko, W. W. & Rolnick, W. B. The Effective W+-, Z0 Approximation for High-Energy Collisions. *Phys. Lett.* **B148**, 367–372 (1984).

[23] Kauer, N., Plehn, T., Rainwater, D. L. & Zeppenfeld, D. H –¿ W W as the discovery mode for a light Higgs boson. *Phys. Lett.* **B503**, 113–120 (2001). hep-ph/0012351.

[24] Ellis, R. K., Georgi, H., Machacek, M., Politzer, H. D. & Ross, G. G. Perturbation theory and the parton model in qcd. *Nuclear Physics B* **152**, 285 – 329 (1979). URL http://www.sciencedirect.com/science/article/pii/0550321379901056.

[25] V. Barger, R. Phillips. *Collider Physics, Updated Edition* (Westview Press, 1991).

[26] Bjorken, J. D. & Paschos, E. A. Inelastic electron-proton and γ-proton scattering and the structure of the nucleon. *Phys. Rev.* **185**, 1975–1982 (1969).

[27] Altarelli, G. & Parisi, G. Asymptotic freedom in parton language. *Nuclear Physics B* **126**, 298 – 318 (1977). URL http://www.sciencedirect.com/science/article/pii/0550321377903844.

[28] Lai, H.-L. *et al.* New parton distributions for collider physics. *Phys. Rev.* **D82**, 074024 (2010). hep-ph/1007.2241.

[29] Wilson, R. R. The Tevatron. *Phys. Today* **30N10**, 23–30 (1977).

[30] D0 Collaboration. Study of the Strong Coupling Constant Using W1 Jet Processes. *Phys. Rev. Lett.* **75**, 18 (1995).

[31] CDF Collaboration. Evidence for top quark production in $p\bar{p}$ collisions at $\sqrt{s}$=1.8 TeV. *Phys. Rev. D* **50**, 2966–3026 (1994).

[32] CDF Collaboration. Observation of Top Quark Production in $p\bar{p}$ Collisions with the Collider Detector at Fermilab. *Phys. Rev. Lett.* **74**, 2626–2631 (1995).

[33] Cabibbo, N. Unitary Symmetry and Leptonic Decays. *Phys. Rev. Lett.* **10**, 531–533 (1963).

[34] Kobayashi, M. & Maskawa, T. CP Violation in the Renormalizable Theory of Weak Interaction. *Prog. Theor. Phys.* **49**, 652–657 (1973).

[35] Maltoni, F. & Stelzer, T. MadEvent: Automatic event generation with MadGraph. *JHEP* **02**, 027 (2003). hep-ph/0208156.

[36] Abachi, S. *et al.* The D0 Detector. *Nucl. Instrum. Meth.* **A338**, 185–253 (1994).

[37] Abe, F. *et al.* The CDF detector: an overview. *Nucl. Instr. Meth.* **A271**, 387–403 (1988).

[38] Evans, L. & Bryant, P. Lhc machine. *Journal of Instrumentation* **3**, S08001 (2008). URL http://stacks.iop.org/1748-0221/3/i=08/a=S08001.

[39] CMS Collaboration. The CMS experiment at the CERN LHC. *Journal of Instrumentation* **3**, S08004 (2008).

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

[41] CDF Collaboration. Measurement of the cross section for W-boson production in association with jets in $p\bar{p}$ collisions at $\sqrt{s}$=1.96 TeV. *Phys. Rev. D* **77**, 011108 (2008).

[42] D0 Collaboration. Measurements of inclusive W+jets production rates as a function of jet transverse momentum in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV. *Submitted to PLB* (2011). hep-ph/1106.1457v1.

[43] CERN. The accelerator complex. http://public.web.cern.ch/public/en/research/AccelComplex-en.html.

[44] CMS Collaboration. CMS Physics Technical Design Report Volume 1: Detector performance and software. *CERN/LHCC 2006-021, CMS TDR 8.2* (2006).

[45] Collaboration, C. CMS Detector. http://cms.web.cern.ch/cms/Media/Images/Detector/index.html.

[46] Collaboration, C. Transverse Slice of the Compact Muon Solenoid. http://cms.web.cern.ch/cms/Detector/FullDetector/index.html.

[47] CMS Collaboration. The Tracker Project Technical Design Report. *CERN/LHCC* **98-006** (1998).

[48] CMS Collaboration. Addendum to the CMS tracker TDR. *CERN-LHCC* **2000-016** (2000).

[49] Das, S. Status and performance of the Compact Muon Solenoid pixel detector. *Nucl. Instrum. Meth.* **A623**, 147–149 (2010).

[50] Cittolin, S. CMS Detector posters. `https://cms-docdb.cern.ch/cgi-bin/PublicDocDB/ShowDocument?docid=4227` (2000).

[51] CMS Collaboration. CMS: The electromagnetic calorimeter. Technical design report. *CERN-LHCC* **97-336** (1997).

[52] CMS Collaboration. Addendum to the CMS ECAL technical design report: Changes to the CMS ECAL electronics. *CERN-LHCC* **2002-027** (2002).

[53] CMS Collaboration. HCAL Technical Design Report. *CERN-LHCC* **97-31** (1997). URL `http://uscms.fnal.gov/uscms/Subsystems-old/HCAL/hcal_tdr/hcal_tdr.html`.

[54] CMS, the Compact Muon Solenoid. Muon technical design report. *CERN-LHCC* **97-32** (1997).

[55] CMS Collaboration. CMS TriDAS project: Technical Design Report; 1, the trigger systems. *CERN-LHCC* **2000-038** (2000).

[56] Chumney, P. *et al.* Level-1 Regional Calorimeter Trigger System for CMS. *Proceedings of Computing in High Energy Physics and Nuclear Physics* (2003). URL `http://arxiv.org/abs/hep-ex/0305047`.

[57] Adam, W. *et al.* The CMS high level trigger. *Eur. Phys. J.* **C46**, 605–667 (2006). `hep-ex/0512077`.

[58] Sphicas, e. ., P. CMS: The TriDAS project. Technical design report, Vol. 2: Data acquisition and high-level trigger. *CERN-LHCC* **2002-026** (2002).

[59] CMS Collaboration. Electron reconstruction and identification at $\sqrt{s} = 7$ TeV. *CMS PAS* **EGM-2010-004** (2010).

[60] Metropolis, N., Ulam, S. The Monte Carlo Method. *Journal of the American Statistical Association* **44:247**, 335–341 (1949).

[61] Dobbs, M. A. *et al.* Les Houches guidebook to Monte Carlo generators for hadron collider physics (2004). hep-ph/0403045.

[62] Gleisberg, T. *et al.* Event generation with SHERPA 1.1. *JHEP* **02**, 007 (2009). 0811.4622.

[63] G. Corcella, I.G. Knowles, G. Marchesini, S. Moretti, K. Odagiri, P. Richardson, M.H. Seymour and B.R. Webber. HERWIG 6.5: an event generator for Hadron Emission Reactions With Interfering Gluons. *JHEP* **0101**, 010 (2001). hep-ph/0011363.

[64] Sjöstrand, T., Mrenna, S. & Skands, P. Z. PYTHIA 6.4 Physics and Manual. *JHEP* **05**, 026 (2006). hep-ph/0603175.

[65] Catani, S., Krauss, F., Kuhn, R. & Webber, B. R. QCD Matrix Elements + Parton Showers. *JHEP* **11**, 063 (2001). hep-ph/0109231.

[66] Mangano, M. L., Moretti, M., Piccinini, F. & Treccani, M. Matching matrix elements and shower evolution for top- quark production in hadronic collisions. *JHEP* **01**, 013 (2007). hep-ph/0611129.

[67] Sjostrand, T. A Model for Initial State Parton Showers. *Phys. Lett.* **B157**, 321 (1985).

[68] Andersson, B., Gustafson, G., Ingelman, G. & Sjostrand, T. Parton Fragmentation and String Dynamics. *Phys. Rept.* **97**, 31–145 (1983).

[69] Mangano, M. L., Moretti, M., Piccinini, F., Pittau, R. & Polosa, A. D. ALPGEN, a generator for hard multiparton processes in hadronic collisions. *JHEP* **07**, 001 (2003). hep-ph/0206293.

[70] CMS Collaboration. Measurement of the Underlying Event Activity at the LHC with $\sqrt{s} = 7$ TeV and Comparison with $\sqrt{s} = 0.9$ TeV. *CMS PAS* **QCD-2010-010** (2010).

[71] Field, R. Early LHC Underlying Event Data - Findings and Surprises (2010). 1010.3558.

[72] Skands, P. Z. Tuning Monte Carlo Generators: The Perugia Tunes. *Phys. Rev.* **D82**, 074018 (2010). 1005.3457.

[73] Lai, H. L. *et al.* Global QCD analysis of parton structure of the nucleon: CTEQ5 parton distributions. *Eur. Phys. J.* **C12**, 375–392 (2000). hep-ph/9903282.

[74] Campbell, J. M., Frederix, R., Maltoni, F. & Tramontano, F. Next-to-Leading-Order Predictions for t-Channel Single-Top Production at Hadron Colliders. *Phys. Rev. Lett.* **102**, 182003 (2009). 0903.0005.

[75] Gavin, R., Li, Y., Petriello, F. & Quackenbush, S. FEWZ 2.0: A code for hadronic Z production at next-to- next-to-leading order (2010). 1011.3540.

[76] Campbell, J. & Ellis, R. MCFM - Monte Carlo for FeMtobarn processes URL http://mcfm.fnal.gov/.

[77] Melnikov, K. & Petriello, F. Electroweak gauge boson production at hadron colliders through O(alpha(s)**2). *Phys. Rev.* **D74**, 114017 (2006). hep-ph/0609070.

[78] Melnikov, K. & Petriello, F. The W boson production cross section at the LHC through $O(\alpha_s^2)$. *Phys. Rev. Lett.* **96**, 231803 (2006). hep-ph/0603182.

[79] Agostinelli, S. *et al.* GEANT4: A simulation toolkit. *Nucl. Instrum. Meth.* **A506**, 250–303 (2003).

[80] Allison, J. *et al.* Geant4 developments and applications. *IEEE Trans. Nucl. Sci.* **53**, 270 (2006).

[81] CMS Collabortation. Track Reconstruction in the CMS tracker (in preparation). *CMS PAS* **TRK-09-001** (2009).

[82] CMS Collaboration. The cms high level trigger. *The European Physical Journal C - Particles and Fields* **46**, 605–667 (2006). URL http://dx.doi.org/10.1140/epjc/s2006-02495-8. 10.1140/epjc/s2006-02495-8.

[83] L. Agostino, M. Pieri. High Level Trigger Selection of Electrons and Photons. *CMS Note* **2006/078** (2006).

[84] CMS Collaboration. ParticleFlow Event Reconstruction in CMS and Performance for Jets, Taus, and $\not\!\!E_T$. *CMS PAS* **PFT-09-001** (2009).

[85] F. Beaudette, D. Benedetti, P. Janot, and M. Pioppi. Electron Reconstruction within the Particle Flow Algorithm. *CMS AN* **2010-034** (2010).

[86] W. Adam, S. Baffioni, F. Beaudette, *et al.* . Electron Reconstruction in CMS. *CMS AN* **2009-164** (2009).

[87] W. Adam, R. Fruhwirt. Reconstruction of Electron Tracks With the Gaussian-Sum Filter in the CMS tracker at LHC. *J. Phys. G: Nucl. Part. Phys.* **31**, N9–N20 (2005).

[88] Kovalskyi, D. *et al.* Fireworks: A physics event display for CMS. *J. Phys. Conf. Ser.* **219**, 032014 (2010).

[89] Baffioni, S. *et al.* Electron reconstruction in CMS. *Eur. Phys. J.* **C49**, 1099–1116 (2007).

[90] Cacciari, M., Salam, G. P. & Soyez, G. The anti-k_t jet clustering algorithm. *Journal of High Energy Physics* **2008**, 063 (2008). URL `http://stacks.iop.org/1126-6708/2008/i=04/a=063`.

[91] Ellis, S. D. & Soper, D. E. Successive combination jet algorithm for hadron collisions. *Phys. Rev.* **D48**, 3160–3166 (1993). `hep-ph/9305266`.

[92] M. Cacciari and G.P. Salam. Pileup subtraction using jet areas. *Physics Letters* **B 659**, 119–126 (2008). `hep-ph/0707.1378`.

[93] M. Cacciari, G.P. Salam and G. Soyez. The catchment area of jets (hep-ph/0802.1188). *JHEP* **0804**, 005 (2008).

[94] CMS Collaboration. Jet Performance in pp Collisions at 7 TeV. *CMS PAS* **JME-10-003** (2010).

[95] CMS Collaboration. Jet Energy Corrections determination at $\sqrt{s}$=7 TeV. *CMS PAS* **JME-10-010** (2010).

[96] CMS Collaboration. Algorithms for b Jet Identification in CMS. *CMS PAS* **BTV-09-001** (2009).

[97] CMS Collaboration. Commissioning of b-jet identification with pp collisions at $\sqrt{s} = 7$ TeV. *CMS PAS* **BTV-10-001** (2010).

[98] CMS Collaboration. Missing Transverse Energy Performance in Minimum-Bias and Jet Events from Proton-Proton Collisions at $\sqrt{s} = 7$ TeV. *CMS PAS* **JME-10-004** (2010).

[99] CMS Collaboration. Measurement of CMS Luminosity. *CMS PAS* **EWK-10-004** (2010). URL `http://cdsweb.cern.ch/record/1279145`.

[100] CMS Collaboration. Measurements of Inclusive W and Z Cross Sections in pp Collisions at $\sqrt(s) = 7$ TeV. *CMS PAS* **2010/005** (2010).

[101] Barge, D., *et al.* Study of photon conversion rejection at CMS. *CMS AN* **2009-159** (2009).

[102] Chatrchyan, S. *et al.* Identification and Filtering of Uncharacteristic Noise in the CMS Hadron Calorimeter. *JINST* **5**, T03014 (2010). `0911.4881`.

[103] N. Adam, J. Berryhill, V. Halyo, A. Hunt, K. Mishra. Generic Tag and Probe Tool for Measuring Efficiency at CMS with Early Data. *CMS AN* **2009-111** (2009).

[104] G. Breit, E. W. Capture of slow neutrons. *Phys. Rev.* **49** (1936).

[105] Oreglia, M. A study of the reactions $\psi' \to \gamma\gamma\psi$. SLAC-0236.

[106] Gaiser, J. Charmonium spectroscopy from radiative decays of the J/ψ and ψ'. SLAC-0255.

[107] CMS Collaboration. Rates of Jets Produced in Association with W and Z Bosons in pp Collisions at $\sqrt(s) = 7$ TeV. *CMS PAS* **EWK-10-012** (2010).

[108] Vanelderen, L. *et al.* Evaluation of the b-tag performance for the $W \to l\nu$ + jets measurements. *CMS Note* **AN-10-461** (2010).

[109] Hocker, A. & Kartvelishvili, V. SVD Approach to Data Unfolding. *Nucl. Instrum. Meth.* **A372**, 469–481 (1996). `hep-ph/9509307`.

[110] D'Agostini, G. A Multidimensional unfolding method based on Bayes' theorem. *Nucl. Instrum. Meth.* **A362**, 487–498 (1995).