University of California
Santa Barbara

A search for Lorentz-boosted Higgs bosons decaying to charm quarks in the CMS experiment using deep neural networks

A dissertation submitted in partial satisfaction
of the requirements for the degree

Doctor of Philosophy
in
Physics

by

Huilin Qu

Committee in charge:

Professor Joseph Incandela, Chair
Associate Professor Nathaniel Craig
Professor Jeffrey Richman

September 2019

The Dissertation of Huilin Qu is approved.

Associate Professor Nathaniel Craig

Professor Jeffrey Richman

Professor Joseph Incandela, Committee Chair

September 2019

A search for Lorentz-boosted Higgs bosons decaying to charm quarks in the CMS experiment using deep neural networks

by

Huilin Qu

Acknowledgements

The pursuit of my Ph.D. in particle physics is a long journey with many ups and downs. I am grateful to so many people who showed up along this way, influenced me greatly with their knowledge and character, and provided me with invaluable help and support. I apologize I cannot name every one of them here – the list is too long. But I have my sincere gratitude to all of them deep in my heart.

First of all, I would like to express my deepest gratitude to my research advisor, Prof. Joseph Incandela, for being always so supportive and inspiring, for allowing a lot of freedom in my research, and for providing me with countless valuable suggestions not only about my research but also on how to grow as a scientist. Talking with him is like magic – his words are so inspiring, and can always clear up my doubts and leave me motivated. I have learned certainly an enormous amount as a Ph.D. student of his, and I hope I will have the chance to learn even more from him in the future.

I would also like to thank Prof. Jeffrey Richman and Prof. Nathaniel Craig. As professors at UC Santa Barbara, they teach excellent courses and helped me build a solid foundation in particle physics and quantum field theory. As my research advisors, they have always been kind and helpful, providing me with valuable guidance at all stages of my graduate studies.

A big thank you should go to Loukas Gouskos, who has been my closest collaborator in the past five years. He is my supervisor on several projects, but also a great friend and an invaluable mentor. He has influenced me in almost all aspects as a physicist: from reading and understanding a plot, to designing a robust physics analysis, and to thinking strategically and planning well ahead for the future. I am deeply indebted to him for his invaluable help, guidance, and support.

I am also grateful to many other members of the Incandela's group: Owen Colegrove,

Giulia Collura, Valentina Dutta, Nick McColl, Sam Mullin, Alex Patterson, Melissa Quinnan. With some of them, I have worked on the 0-lepton top squark search, which was my first research project as a Ph.D. student. I have learned so much from all of them. Some of the tools they developed have actually become an indispensable part of my arsenal. With the others, though I have not worked with them directly, I really enjoyed the conversations with them and the help and advice I got from them. It is such an unforgettable memory being together with this friendly and creative group.

As a member of the CMS Collaboration, I enjoy the privilege of working together with so many outstanding physicists, and I have benefited a lot from the collaboration. I would like to thank all of the (sub)group conveners, analysis review committees, and members at large for their constructive comments. It is hard to make a small list, but I am grateful to these people in particular for their insightful comments and helpful advice: Wolfgang Adam, Claudio Campagnari, Zeynep Demiragli, Pieter Everaerts, Frank Golf, Andreas Hinzmann, Gregor Kasieczka, Clemens Lange, Petar Maksimovic, Pablo Martinez, Filip Moortgat, Kevin Nash, Giovanni Petrucciani, Salvatore Rappoccio, Roberto Salerno, Matthias Schröder, Seema Sharma, Lesya Shchutska, Nhan Tran. I am deeply grateful to Loukas Gouskos, Jan Kieseler, Markus Stoye and Mauro Verzetti for initiating the DeepJet and DeepAK8 efforts, and also Huang Huang, Matthias Mozer and Pantelis Kontaxakis for making the DeepAK8 team more fun and productive. Thank you to Gaelle Boudoul, Erik Butz, Giovanni Franzoni, Giacomo Govi, Arun Kumar and Marco Musich! I really enjoyed the days working with them on the tracker O2O project, and I have learned a lot in that process.

Thank you, my parents! Words are not enough to express my sincere gratitude to them for all their love and support over so many years. I know I owe them too much these years when I am 8000 kilometers away from them. I have to meet them in person to express my deepest gratitude to them!

I cannot conclude this thesis without thanking my girlfriend, Weidi Yuan, without whom this thesis cannot materialize (at least not within another six months). I am deeply grateful to her for always being by my side, from Beijing to Santa Barbara, Geneva, and wherever we will be together in the future. Her company is like the Californian sunshine, brightening and warming up my life every day. And her insights and advice are always just to the point. Knowing that I will always have her support is one of the most important sources of my courage to take on yet a new adventure in my life.

Curriculum Vitæ

Huilin Qu

Education

2019	Ph.D. in Physics (Expected), University of California, Santa Barbara.
2016	M.A. in Physics, University of California, Santa Barbara.
2014	B.S. in Physics, Peking University.

Selected publications in refereed journals

- G. Kasieczka *et al.*, "The Machine Learning Landscape of Top Taggers," SciPost Phys. 7, 014 (2019).
- H. Qu and L. Gouskos, "ParticleNet: Jet Tagging via Particle Clouds," arXiv:1902.08570 [hep-ph] (*Submitted to Phys. Rev. D*).
- A. M. Sirunyan *et al.* [CMS Collaboration], "Search for direct production of supersymmetric partners of the top quark in the all-jets final state in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$," JHEP 1710, 005 (2017).
- A. M. Sirunyan *et al.* [CMS Collaboration], "Searches for pair production of third-generation squarks in $\sqrt{s} = 13\,\text{TeV}$ pp collisions," Eur. Phys. J. C 77, no. 5, 327 (2017).

Selected conference notes

- CMS Collaboration, "Machine learning-based identification of highly Lorentz-boosted hadronically decaying particles at the CMS experiment," CMS-PAS-JME-18-002.
- CMS Collaboration, "Search for the standard model Higgs boson decaying to charm quarks," CMS-PAS-HIG-18-031.
- CMS Collaboration, "First constraints on invisible Higgs boson decays using $t\bar{t}H$ production at $\sqrt{s} = 13\,\text{TeV}$," CMS-PAS-HIG-18-008.
- CMS Collaboration [CMS Collaboration], "Search for direct top squark pair production in the all-hadronic final state in proton-proton collisions at sqrt(s) = 13 TeV," CMS-PAS-SUS-16-049.
- CMS Collaboration, "Search for direct top squark pair production in the fully hadronic final state in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$ corresponding to an integrated luminosity of 12.9/fb," CMS-PAS-SUS-16-029.
- CMS Collaboration [CMS Collaboration], "Search for direct production of top squark pairs decaying to all-hadronic final states in pp collisions at sqrt(s) = 13 TeV," CMS-PAS-SUS-16-007.

Abstract

A search for Lorentz-boosted Higgs bosons decaying to charm quarks in the CMS experiment using deep neural networks

by

Huilin Qu

Measurement of the decay of the Higgs boson to charm quarks provides a direct probe of the Higgs coupling to second-generation quarks. Therefore, it is crucial for understanding the structure of Yukawa couplings. In this thesis, a search for the Higgs boson decaying to charm quarks with the CMS experiment is presented. The search is designed for Lorentz-boosted Higgs bosons produced in association with vector (V) bosons (W or Z bosons). A novel approach that reconstructs both quarks from the Higgs boson decay with a single large-radius jet is adopted. The charm quark pair is identified with an advanced deep learning–based algorithm. This approach leads to a highly competitive result: Using proton-proton collision data corresponding to an integrated luminosity of $35.9\,\mathrm{fb}^{-1}$, an observed (expected) upper limit on $\sigma(\mathrm{VH}) \times \mathcal{B}(\mathrm{H} \to \mathrm{c\bar{c}})$ of 71 (49) times the standard model expectation at 95% confidence level is obtained.

A detailed description of the deep learning–based boosted object identification algorithm is also presented in this thesis. It is a versatile algorithm designed to identify and classify hadronic decays of highly Lorentz-boosted top quarks and W, Z, Higgs bosons. Using deep neural networks to directly access and process the raw information of all constituent particle-flow candidates of a jet, this advanced algorithm has achieved significant performance improvements compared to traditional approaches.

Contents

Part I

Introduction to Experimental Particle Physics

The study of the properties and interactions of the elementary particles is a fascinating subject. It is a way to probe the finest structure of space-time and to uncover the fundamental laws of the universe. To investigate the most elusive microscopic phenomena of the elementary particles, large-scale and highly intricate macroscopic experimental apparatuses, such as the 27-km long Large Hadron Collider, the 15-m high Compact Muon Solenoid detector with hundreds of millions of electronic readout channels, are needed. This makes experimental particle physics a particular challenging but also exciting field.

This part of the thesis serves as a brief introduction to experimental particle physics. Chapter 1 provides a concise overview of the standard model of particle physics and the phenomenology of the Higgs boson at a particle collider. Chapter 2 describes the setup of the Compact Muon Solenoid experiment with which the studies presented in this thesis are performed.

Chapter 1

Theoretical Background

The quest for the fundamental laws of nature, particularly, the elementary building blocks of the universe, drives the development of particle physics. The discovery of the electron by J. J. Thomson in 1897 marks the beginning of the exploration of the subatomic world. Since then, a plethora of new particles has been discovered, as shown in Figure 1.1. Advancements in experiments are always intertwined with progress in theories. Quantum field theory, incorporating quantum mechanics and special relativity, was gradually developed in the 20th century and since then became the mathematical language of particle physics. Theories for the newly discovered strong and weak interactions were proposed, developed, and eventually evolved into the Standard Model (SM) of particle physics in the 1970s. The discovery of the W and Z bosons in 1983 marked a great triumph of the SM. Since then, the SM has undergone extensive precision tests at various experiments and has shown impressive agreement with observations. In 2012, the discovery of the Higgs boson marked the completion of the last missing piece of the SM, and particle physics has entered a brand new era.

Although the SM has achieved great success, various kinds of evidence, such as its incapability to account for the existence of dark matter, indicate that the SM is still

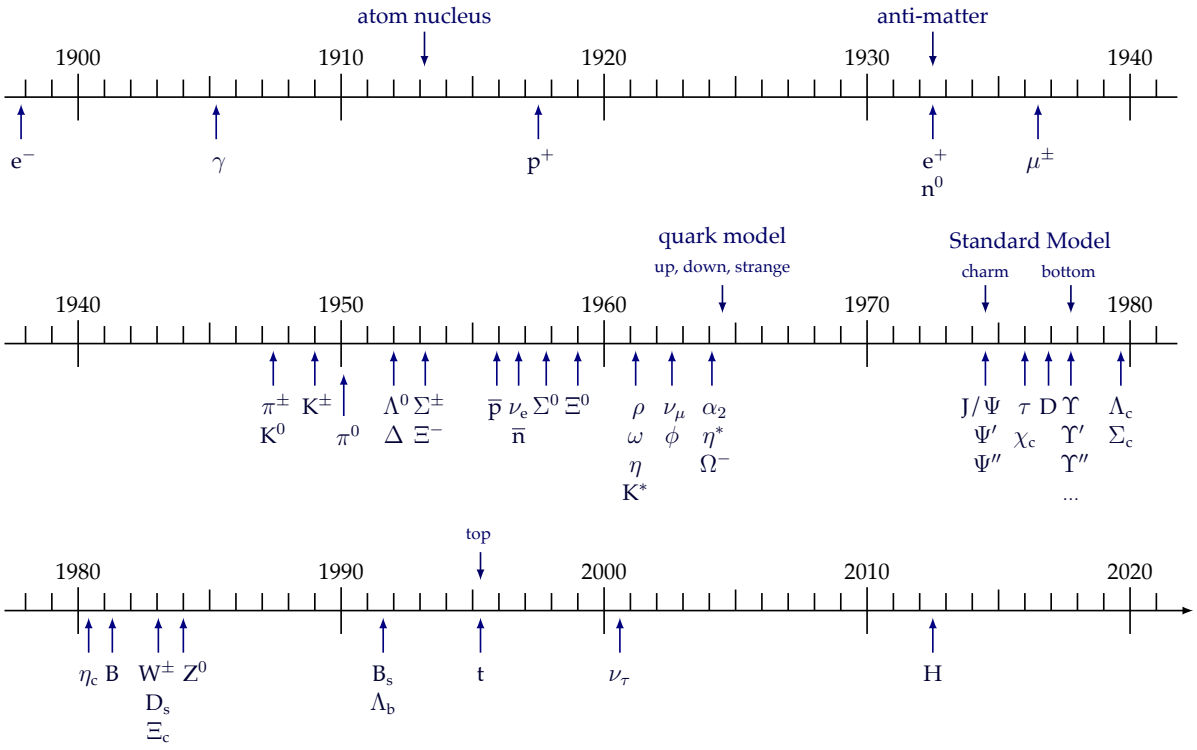


Figure 1.1: Timeline of particle discoveries. Figure adapted from [6].

not a complete theory. Furthermore, the newly discovered Higgs boson opens a whole new portal that remains yet to be fully scrutinized experimentally. Any deviation from the SM prediction, if observed, is a clear sign of new physics. Therefore, studying the properties of the Higgs boson is of paramount importance at current and future particle physics experiments.

This chapter provides a brief discussion of the theoretical background needed for studying the Higgs boson properties at a particle collider. Section 1.1 gives a concise overview of the SM[1], with an emphasis on the Brout-Englert-Higgs mechanism. Section 1.2 discusses the production mechanisms and the decay modes of the Higgs boson.

[1] A more thorough treatment of the SM can be found in many excellent textbooks on quantum field theory (e.g., Refs. [1, 2, 3]) and particle physics (e.g., Refs. [4, 5])

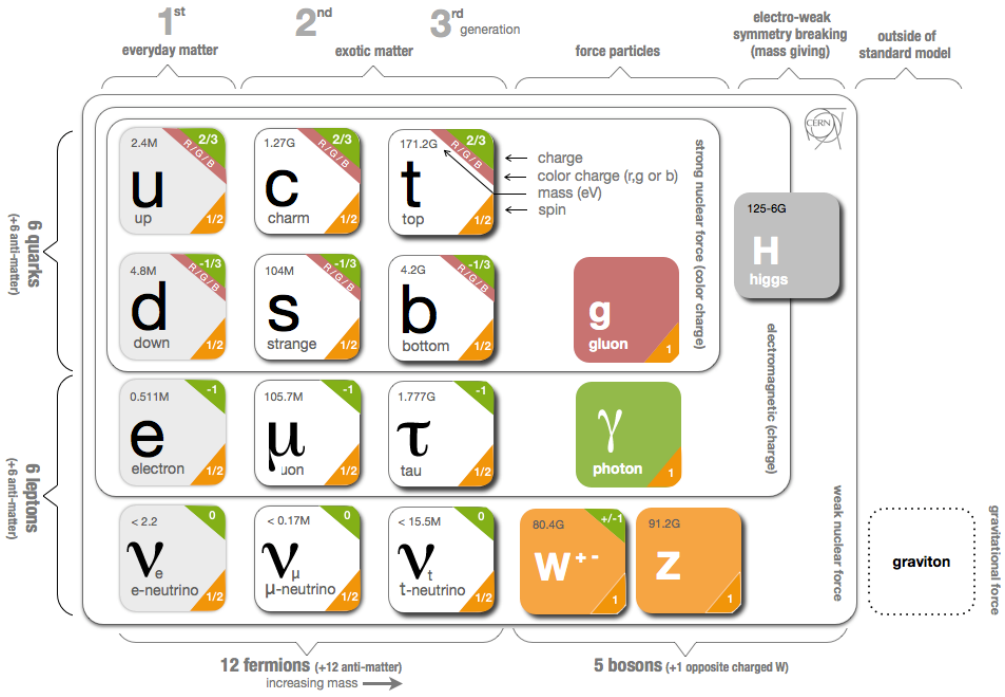


Figure 1.2: Summary of the particle content of the SM. Figure from [7].

1.1 The Standard Model of Particle Physics

The SM is a theoretical framework that describes all known elementary particles and their interactions[2]. A particle is considered *elementary* if it has no internal structures. In the SM, the elementary particles fall into 2(+1) categories: the spin-1/2 fermions and the spin-1(0) bosons. The spin-1/2 fermions are often referred to as "matter particles", as all macroscopic matter are composed of elementary particles in this category, specifically, the up and down quarks (constituting protons and neutrons and therefore the atomic nuclei) and electrons. The fermions can be further divided into quarks and leptons, with only the quarks interacting via the strong force. The leptons include the electrically charged electron, muon and tau, and their electrically neutral counterparts, the electron neutrino, muon neutrino, and tau neutrino. The spin-1 (vector) bosons are the so-called

[2]Gravity is not accounted for in the SM, as the gravitational force between elementary particles is extremely weak at energy scales probed so far.

"force carriers", i.e., they mediate the interactions between particles. The photon is responsible for the electromagnetic interaction, the $W^{\pm}$ and Z bosons are responsible for the weak interaction, and the gluons (with eight different "colors") are responsible for the strong interactions. As the only spin-0 elementary particle observed so far, the Higgs boson has a special role in the SM. It is an indispensable ingredient of the unified electroweak (EW) theory and is also the origin of the masses of the fermions and the $W^{\pm}$ and Z bosons. The various types of elementary particles of the SM are summarized in Figure 1.2, with the main quantum numbers, the mass, and the participating interactions of each particle also displayed.

The description of the strong, weak, and electromagnetic interactions in the SM is based on a quantum field theory with a SU(3) × SU(2) × U(1) gauge symmetry. The SU(3) group is the symmetry group of quantum chromodynamics (QCD) that describes the strong interaction, while SU(2)×U(1) corresponds to the electroweak theory unifying the weak and electromagnetic interactions. The Lagrangian of the SM can be written, in a rather formal manner without making explicit all the fields and indices, as

$$\mathcal{L} = -\frac{1}{4}F^a_{\mu\nu}F^{\mu\nu}_a + i\bar{\psi}\not{D}\psi + (y_{ij}\bar{\psi}_i\psi_j\phi + \text{h.c.}) + |D_\mu\phi|^2 - V(\phi), \tag{1.1}$$

where ψ denotes the fermion fields, and ϕ denotes the scalar field, $F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + gf^{abc}A^b_\mu A^c_\nu$ is the field strength tensor of a gauge filed A_μ, and $D_\mu = \partial_\mu - igA^a_\mu t_a$ is the gauge covariant derivative in which t_a is the group generator in the chosen representation. The definition of the field strength tensor and the covariant derivative ensures that the Lagrangian is gauge-invariant.

One feature of the SM Lagrangian in eq. (1.1) is that there are no explicit mass terms for the fermions or the gauge fields. In fact, a fermionic mass term like $m\bar{\psi}_L\psi_R + \text{h.c.}$ is not invariant under gauge transformation. Mass terms for the gauge bosons of the form

$m^2 A^a_\mu A^\mu_a$ are also prohibited by the gauge symmetry. This poses a severe problem as the fermions and the $W^\pm$ and Z bosons are massive. In the SM, this is solved by the Brout-Englert-Higgs mechanism [8, 9, 10, 11, 12, 13] by the introduction of a self-interacting complex scalar field ϕ which spontaneously breaks the SU(2) $\times$ U(1) symmetry. The scalar field is a SU(2) doublet with four degrees of freedom,

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}. \tag{1.2}$$

The scalar potential term in the SM Lagrangian is

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2, \tag{1.3}$$

where $\lambda > 0$ in order to have a finite minimum. When $\mu^2 > 0$, the potential has a unique minimum located at $\phi = (0\ 0)$. If $\mu^2 < 0$, the potential has an infinite set of minima defined by

$$\phi^\dagger \phi = \frac{-\mu^2}{2\lambda}. \tag{1.4}$$

The physical vacuum state will correspond to a particular configuration in this infinite set of minima. Without loss of generality, the vacuum state can be chosen to be

$$\langle \phi \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}, \quad \text{where } v = \sqrt{\frac{-\mu^2}{\lambda}}. \tag{1.5}$$

The choice of a specific vacuum state breaks the gauge symmetry, although the potential itself respects the symmetry. The scalar field can then be expanded around the vacuum expectation value (VEV) as

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ v + h + ia^0 \end{pmatrix}. \tag{1.6}$$

Substituting eq. (1.6) into the SM Lagrangian in eq. (1.1) yields three massless Goldstone degrees of freedom, corresponding to ϕ_1, ϕ_2 and a^0, which mix with the electroweak gauge fields and become the longitudinal components of the $W^{\pm}$ and Z physical gauge bosons. The $W^{\pm}$ and Z bosons acquire masses,

$$m_{\mathrm{W}}^2 = \frac{g^2 v^2}{4}, \quad m_{\mathrm{Z}}^2 = \frac{(g'^2 + g^2)v^2}{4}, \tag{1.7}$$

where g and g' are the SU(2) and SU(1) gauge couplings, respectively. The remaining degree of freedom of the scalar field, h, is the physical Higgs boson with a mass

$$m_{\mathrm{H}} = \sqrt{-2\mu^2} = \sqrt{2\lambda}v. \tag{1.8}$$

After the electroweak symmetry breaking, the fermions acquire mass through interactions with the Higgs field, i.e., the Yukawa interactions. The fermion mass m_f and the Yukawa coupling y_f is related by

$$m_f = \frac{1}{\sqrt{2}} y_f v. \tag{1.9}$$

In the SM, the VEV of the Higgs field is fixed by the Fermi coupling constant

$$G_F = \frac{\sqrt{2}g^2}{8m_{\mathrm{W}}^2} = 1.166\ 378\ 7(6) \times 10^{-5}\,\mathrm{GeV}^{-2}, \tag{1.10}$$

hence

$$v = \left(\sqrt{2}G_F\right)^{-1/2} \simeq 246\,\mathrm{GeV}. \tag{1.11}$$

The quartic coupling λ is a free parameter in the SM, so there is no a priori prediction for the mass of the Higgs boson. The experimentally measured Higgs mass, $m_{\mathrm{H}} \simeq 125\,\mathrm{GeV}$, implies that $\lambda \simeq 0.13$ and $|\mu| \simeq 88.8\,\mathrm{GeV}$. Once the Higgs mass is known, the SM Lagrangian in the Higgs sector is entirely fixed. The interaction terms involving the

Higgs field can be written as [14]:

$$L = -g_{\mathrm{H}ff}\bar{f}fh + \frac{g_{\mathrm{HHH}}}{6}h^3 + \frac{g_{\mathrm{HHHH}}}{24}h^4 + \delta_{\mathrm{V}}V_\mu V^\mu\left(g_{\mathrm{HVV}}h + \frac{g_{\mathrm{HHVV}}}{2}h^2\right) \tag{1.12}$$

with

$$g_{\mathrm{H}ff} = \frac{m_f}{v},\ g_{\mathrm{HVV}} = \frac{2m_{\mathrm{V}}^2}{v},\ g_{\mathrm{HHVV}} = \frac{2m_{\mathrm{V}}^2}{v^2},\ g_{\mathrm{HHH}} = \frac{3m_{\mathrm{H}}^2}{v},\ g_{\mathrm{HHHH}} = \frac{3m_{\mathrm{H}}^2}{v^2} \tag{1.13}$$

where $\mathrm{V} = \mathrm{W}^\pm$ or Z and $\delta_{\mathrm{W}} = 1$, $\delta_{\mathrm{Z}} = 1/2$. From eq. (1.13), we can see that the Higgs couplings to fermions are linearly proportional to their masses, while the couplings to bosons are proportional to the square of the boson masses. As a result, the dominant mechanisms for Higgs boson production and decay involve the coupling of H to $\mathrm{W}^\pm$, Z and/or the heavier third-generation fermions (top and bottom quarks and the τ leptons).

1.2 Higgs Boson Production and Decay Mechanisms

At a hadron collider such as the LHC, the Higgs boson can be produced mainly via the following processes:

- Gluon-gluon fusion (ggF), $\mathrm{gg} \to \mathrm{H}$;
- Vector boson fusion (VBF), $\mathrm{qq}' \to \mathrm{qq}'\mathrm{H}$;
- Associated production with a vector (W, Z) boson (VH), $\mathrm{q\overline{q}} \to \mathrm{WH/ZH}$ plus a small contribution ($\sim 14\%$ at $\sqrt{s} = 13\,\mathrm{TeV}$) from $\mathrm{gg} \to \mathrm{ZH}$ (ggZH);
- Associated production with a top quark-antiquark pair ($\mathrm{t\bar{t}H}$), $\mathrm{q\overline{q}/gg} \to \mathrm{t\bar{t}H}$;

Figure 1.3 shows the representative Feynman diagrams for these Higgs production processes. The cross sections of these production processes are shown in Figure 1.4a as

a function of the center of mass energy ($\sqrt{s}$) of the proton-proton collisions, and also summarized in Table 1.1 for $\sqrt{s} = 13\,\text{TeV}$.

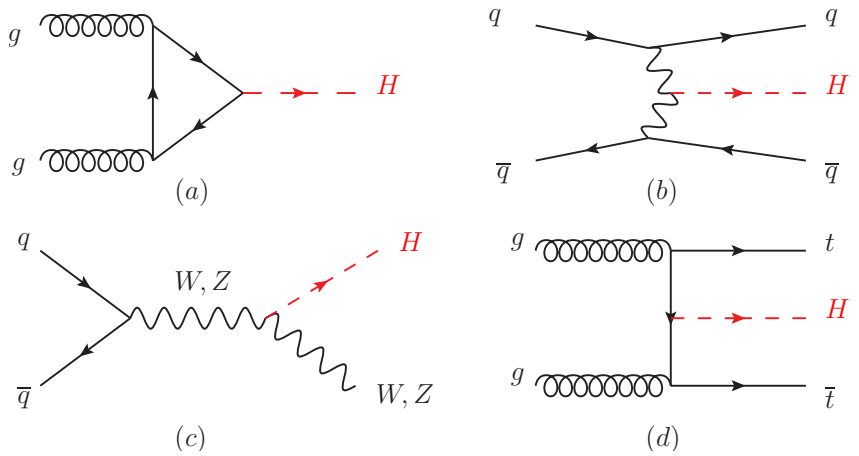
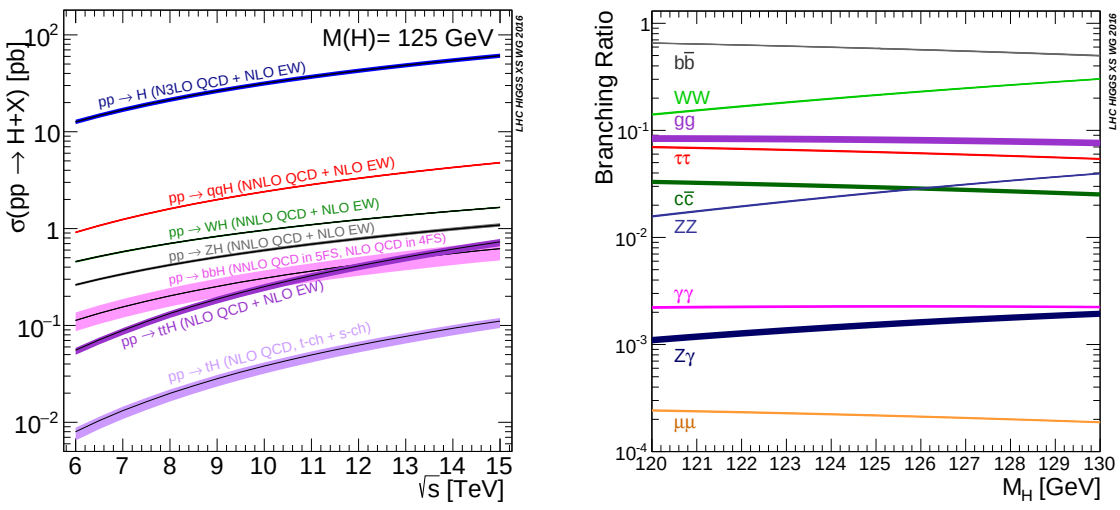


Figure 1.3: Main leading-order Feynman diagrams for Higgs boson production via the (a) ggF (b) VBF (c) VH and (d) $t\bar{t}H$ production processes. Figures from [14].

Production process	Cross section [pb]	TH uncertainty [%]	Accuracy
ggF	48.58	± 3.9	N^3LO QCD and NLO EW
VBF	3.782	± 2.1	(approx.) NNLO QCD and NLO EW
WH	1.373	± 2.0	NNLO QCD and NLO EW
ZH (incl. ggZH)	0.8839	$^{+4.1}_{-3.5}$	NNLO QCD and NLO EW
(ggZH)	0.1227	$^{+25.2}_{-19.1}$	NLO + NLL QCD
$t\bar{t}H$	0.5071	$^{+6.8}_{-9.9}$	NLO QCD and NLO EW

Table 1.1: Summary of the cross sections and the current theoretical (TH) uncertainties of the dominant SM Higgs boson production processes for proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$ [15]. The mass of the Higgs boson is assumed to be $m_H = 125\,\text{GeV}$.

The ggF process is the dominant Higgs production mechanism at the LHC, accounting for almost 90% of the total Higgs production cross section. As shown in Figure 1.3a, this process is loop-induced, with the most significant contribution coming from the top quark as it has the largest Yukawa coupling to the Higgs boson. Therefore, the measurement of the ggF cross section also provides an indirect probe of the Higgs coupling to the top quark.



(a) (b)

Figure 1.4: The SM Higgs boson (a) production cross sections as a function of the center of mass energy ($\sqrt{s}$) of the proton-proton collisions (b) branching fractions as a function of the Higgs boson mass. The bands in both figures correspond to the theoretical uncertainties. Figures from [15].

The VBF process is the sub-leading production mechanism, contributing about 7% to the total cross section. The Feynman diagram is shown in Figure 1.3b. The VBF process features the presence of two forward quark jets, which can be exploited to identify such events. The process can be used to probe the Higgs coupling to the W and Z bosons.

The Higgs coupling to the W and Z bosons can also be probed in the VH process, shown in Figure 1.3c. Though with a much smaller cross section, the VH process has an essential advantage over ggF and VBF for experimental searches, as the presence of a W or Z boson provides an additional handle for event selection. This is particularly valuable to searches for Higgs bosons decaying to quarks (e.g., $H \to b\bar{b}$, $H \to c\bar{c}$), as one can exploit the leptonic decays of the W or Z bosons to effectively suppress the overwhelming background consisting of quarks and gluons produced by QCD radiations. The event topology of the VH process is relatively simple, consisting of a vector boson and a Higgs boson recoiling against each other, with little additional activity in the event.

The $t\bar{t}H$ process has an even smaller cross section, yet it is considered highly important as it provides the only way to directly probe the Yukawa coupling to the top quark — the $H \to t\bar{t}$ decay is kinematically forbidden. The production of top quark-antiquark pair ($t\bar{t}$) constitutes the main background for the $t\bar{t}H$ process. Depending on the decay modes of the $t\bar{t}$ system (di-leptonic, semi-leptonic, fully hadronic) and the Higgs boson itself, the final state can be quite complicated for this process.

With a mass of 125 GeV, the Higgs boson can decay to almost all other SM particles (except the top quark which has a mass of 173 GeV). Therefore, it is possible to probe a large variety of the Higgs couplings by studying the decay of the Higgs boson to different particles. The branching fractions of the Higgs boson as a function of the Higgs boson mass are shown in Figure 1.4b, and the branching fractions for $m_{\mathrm{H}} = 125\,\mathrm{GeV}$ are summarized in Table 1.2. At $m_{\mathrm{H}} = 125\,\mathrm{GeV}$, the dominant decay mode is $H \to b\bar{b}$ and $H \to WW^*$, where W^* indicates an off-shell W boson. This is followed by $H \to gg$, $H \to \tau^+\tau^-$, $H \to c\bar{c}$ and $H \to ZZ^*$. Almost an order of magnitude smaller in branching fractions, the loop-induced $H \to \gamma\gamma$ and $H \to Z\gamma$ decays are much rarer. Due to the small mass of the muons, the branching fraction of $H \to \mu^+\mu^-$ is only 2.176×10^{-4}.

Decay channel	Branching fraction	Relative uncertainty [%]
$H \to b\bar{b}$	5.824×10^{-1}	± 2.2
$H \to \tau^+\tau^-$	6.272×10^{-2}	± 2.8
$H \to c\bar{c}$	2.891×10^{-2}	$^{+7.7}_{-3.4}$
$H \to \mu^+\mu^-$	2.176×10^{-4}	± 2.8
$H \to gg$	8.187×10^{-2}	± 8.2
$H \to \gamma\gamma$	2.270×10^{-3}	± 3.3
$H \to Z\gamma$	1.533×10^{-3}	± 7.3
$H \to WW^*$	2.137×10^{-1}	± 2.6
$H \to ZZ^*$	2.619×10^{-2}	± 2.6

Table 1.2: Summary of the branching fractions and the uncertainties for an SM Higgs boson with $m_H = 125\,\text{GeV}$ [15].

Chapter 2

Experimental Setup

This chapter provides an overview of the experimental apparatus on which all the studies in this thesis are based. A brief introduction of the Large Hadron Collider (LHC) is given in Section 2.1. The Compact Muon Solenoid (CMS) detector, one of the two general-purpose detectors at the LHC, is discussed in Section 2.2.

2.1 The Large Hadron Collider

Located on the border between Switzerland and France near Geneva, the LHC is the world's largest and highest-energy particle collider. The LHC is constructed and operated by the European Organization for Nuclear Research (CERN). It occupies a tunnel of 26.7 km in circumference and at a depth of between 50 to 175 meters underground, originally built for the Large Electron Positron (LEP) collider. After a 10-year construction from 1998 to 2008, the LHC was first turned on and commissioned in 2008. The first record-setting high-energy proton-proton collision at a center of mass energy of 7 TeV was established on 30 March 2010, marking the start of the main physics program. A detailed description of the working principles and the design choices of the LHC is presented in

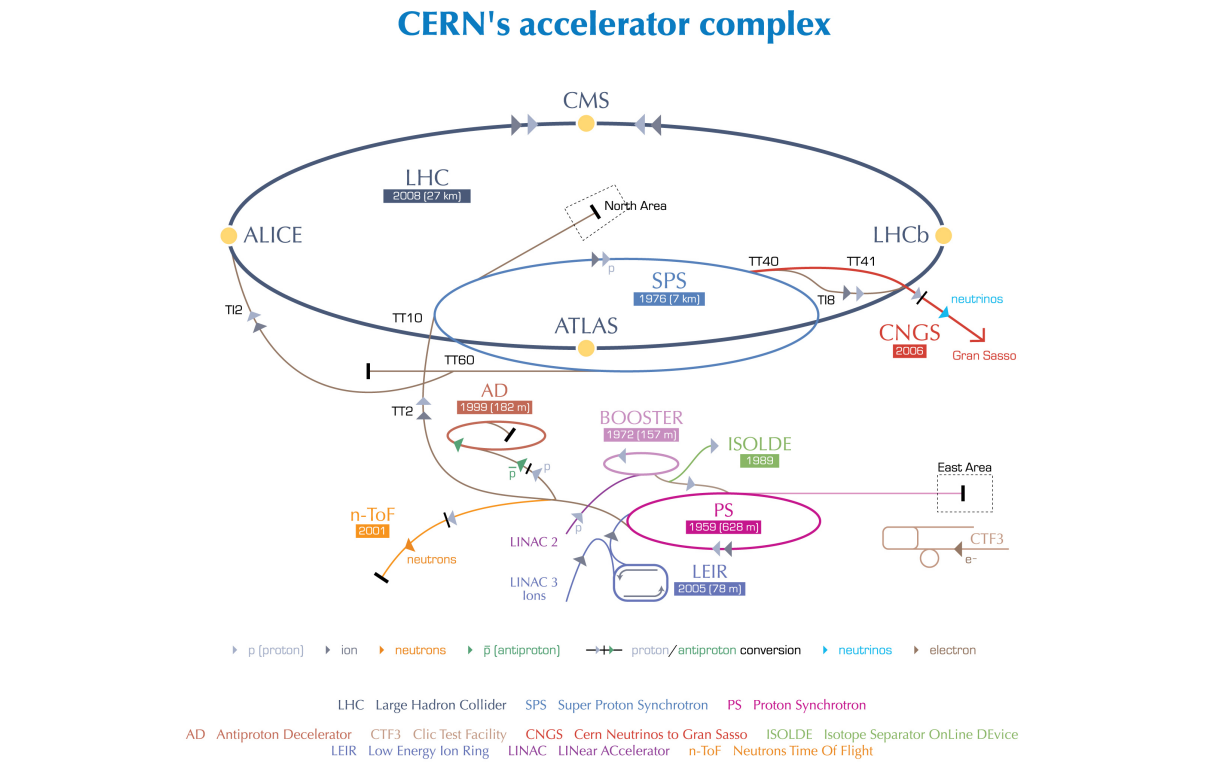


Figure 2.1: The CERN accelerator complex. Figure from [20].

the LHC Design Report [16, 17, 18], and an abridged version can be found in Ref. [19]. Below we provide a brief introduction based on Refs. [16, 17, 18, 19].

As a hadron collider, the LHC is designed to accelerate proton beams to an energy of up to 7 TeV, and lead ions to an energy of up to 2.76 TeV per nucleon. The acceleration of protons (or lead ions) to the highest energy at the LHC takes a series of steps and a complex chain of upstream systems. Figure 2.1 shows an illustration of the CERN accelerator complex. Using protons as an example, the acceleration process is briefly summarized in the following. The protons are first obtained by stripping hydrogen atoms of their electrons with an electric field. The extracted protons are sent to a linear collider (LINAC2) and accelerated to 50 MeV. This is followed by a series of three synchrotrons,

the Proton Synchrotron Booster, the Proton Synchrotron (PS), and the Super Proton Synchrotron (SPS). With increasing diameters, they further accelerate the protons to 1.4 GeV, 25 GeV, and 450 GeV, respectively. Protons leaving the SPS are eventually injected into the LHC main ring to be accelerated up to their maximum energy.

Proton beams orbit the LHC in two metal pipes with a very high vacuum – the beam pipes. A large number of superconducting magnets of various types are installed along the beam pipes to control the trajectories of the proton beams. A total number of 1232 dipole magnets are used to bend the path of proton beams to follow a circular trajectory. Made of NbTi, the superconducting dipole magnets operate at 1.9 K and can generate a magnetic field of up to 8.3 T. A large number of quadruple magnets are also used at the LHC to focus the beams before they collider in the detectors.

One of the key parameters of a collider is the instantaneous luminosity L. This is directly related to the observed rate of an interaction process, with

$$\frac{dN}{dt} = \sigma L, \tag{2.1}$$

where σ is the cross section of the process. The instantaneous luminosity L of a collider is given by

$$L = \frac{N_b^2 n_b f_{\text{rev}} \gamma_r}{4\pi \epsilon_n \beta^*} F, \tag{2.2}$$

where N_b is the number of particles per bunch, n_b is the number of bunches per beam, f_{rev} is the revolution frequency, $\gamma_r (\equiv E/m)$ is the relativistic gamma factor, ϵ_n is the normalized transverse beam emittance, β^* is the beta function at the collision point, and F is the geometric luminosity reduction factor due to the crossing angle at the interaction point. The design peak instantaneous luminosity of the LHC is $L = 10^{34}\,\text{cm}^{-2}\,\text{s}^{-1}$, while a record instantaneous luminosity of $L = 2.06 \times 10^{34}\,\text{cm}^{-2}\,\text{s}^{-1}$, twice the design value,

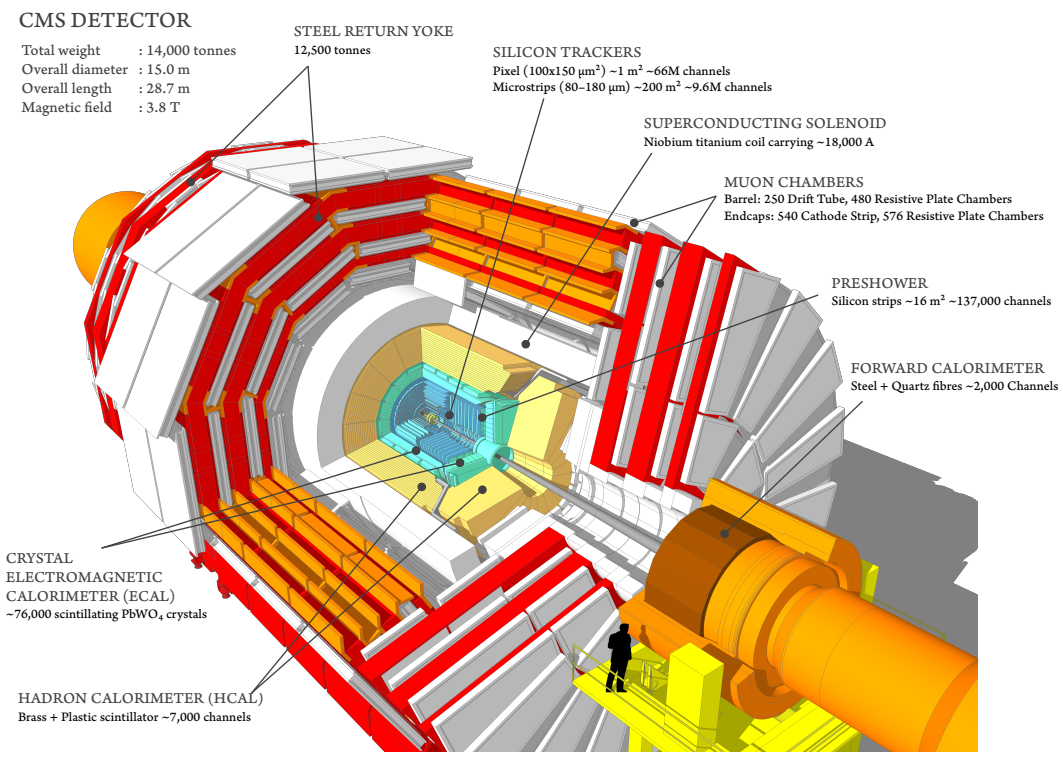


Figure 2.2: A cutaway view of the CMS detector. Figure from [21].

was achieved in 2017.

2.2 The Compact Muon Solenoid Detector

The CMS detector [22] is a general-purpose particle detector installed at interaction point 5 of the LHC. It features a large superconducting *solenoid*, capable of delivering an axial and uniform magnetic field of 3.8 T over a length of 12.5 m and a free-bore radius of 3.5 m. A *compact* design is made possible by the strong magnetic field: the tracking detector, the electromagnetic calorimeter (ECAL) and the hadron calorimeter (HCAL) are all installed inside the solenoid, therefore minimizing energy losses due to particle interactions with the coil and other supporting material in front of the calorimeters. Outside the solenoid is the iron return yoke of the magnet, interleaved with layers of the

muon detector. The detector is 15 m high and 21 m long and weighs a total of 14000 tons. A schematic view of the CMS detector and its components is shown in Figure 2.2.

The CMS detector uses a coordinate system as depicted in Figure 2.3. It is oriented such that the x axis points to the center of the LHC ring, the y axis points vertically upward, and the z axis is in the direction of the counterclockwise proton beam, when looking at the LHC from above. The origin is centered at the nominal collision point inside the experiment. The azimuthal angle ϕ is measured from the x axis in the (x, y) plane. The polar angle θ is measured with respect to the z axis in the (r, z) plane, where r is the radial coordinate in the (x, y) plane. The pseudorapidity, η, is defined as $\eta = -\ln \tan(\theta/2)$.

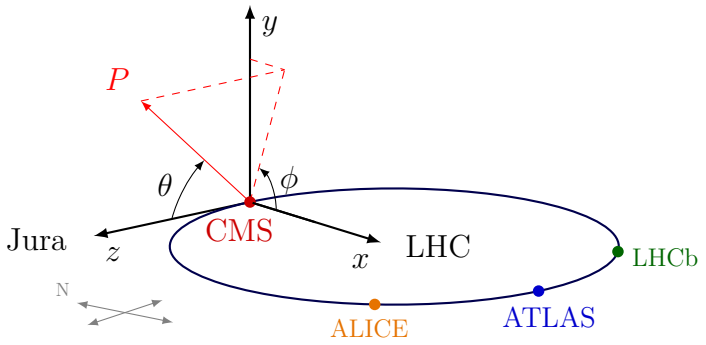


Figure 2.3: Illustration of the CMS coordinate system. Figure from [23].

2.2.1 Inner Tracking System

The innermost component of the CMS detector is a silicon-based tracking system (tracker) [24, 25]. It is a cylinder-shaped detector surrounding the beam pipe, with an outer radius of 1.2 m and a length of 5.6 m. It consists of a pixel detector with three (two) layers in the barrel (at each of the two endcaps), closer to the beam pipe, and a strip detector with ten (twelve) layers in the barrel (at each of the two endcaps) surrounding the pixel detector. A total of 16588 silicon sensor modules are finely segmented into 66

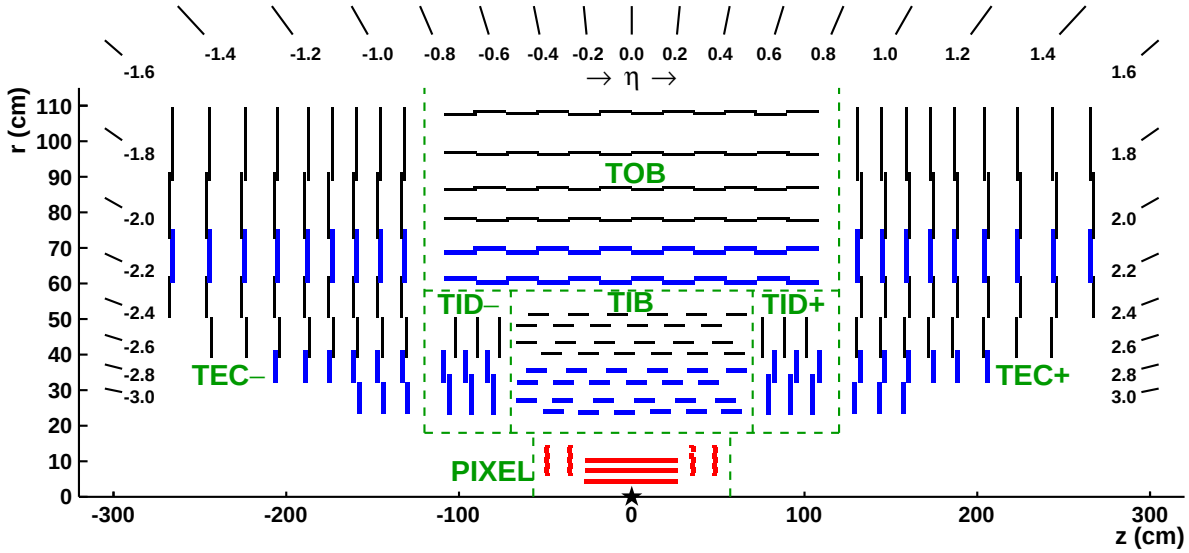


Figure 2.4: Schematic cross-sectional view of the CMS tracker in the (r, z) plane. The tracker is symmetric about the horizontal line $r = 0$, so only the top half is shown here. Figure from [26].

million $150 \times 100\,\mu$m pixels and 9.6 million 80-to-180 μm-wide strips. The layout of the tracker is shown in Figure 2.4. The pixel and strip detectors are combined to provide coverage in pseudorapidity of up to $|\eta| < 2.5$.

The performance of the CMS tracker was studied in detail in Run 1 [26]. The CMS tracker provides a precise and efficient measurement of the trajectories of the charged particles. Combined with the strong bending power of the 3.8 T magnetic field produced by the superconducting solenoid, the tracker achieves a high resolution in the measurement of the transverse momentum p_{T} of charged particles. The tracker measures the p_{T} of charged particles at normal incidence with a resolution of 1% for $p_{\mathrm{T}} < 20\,\mathrm{GeV}$, and $\sim 2\%$ at $p_{\mathrm{T}} \sim 100\,\mathrm{GeV}$. The tracker also provides a precise reconstruction of the primary and secondary vertices. For primary vertices with the sum of the track p_{T} above 100 GeV, the spatial resolution was measured to be better than 14 μm in the (x, y) plane, and better than 19 μm in z direction in the 2016 data [27].

2.2.2 Calorimeters

The electromagnetic calorimeter

The ECAL [28, 29] is a hermetic homogeneous calorimeter, consisting of a barrel covering $|\eta| < 1.479$ and two endcaps covering $1.479 < |\eta| < 3.0$. The lead tungstate ($PbWO_4$) crystal was chosen in the CMS ECAL design due to its short radiation length ($X_0 = 0.89\,\text{cm}$) and small Molière radius (2.2 cm). This allows the design of a compact ECAL inside the magnetic field, with fine granularity, excellent energy resolution, and radiation hardness. The length of the $PbWO_4$ crystals in the barrel (endcap) is 23 (22) cm. This corresponds to 25.8 (24.7) radiation lengths, making it sufficient for the $PbWO_4$ crystals to contain more than 98% of the energy of electrons and photons up to 1 TeV. The crystal material also amounts to about one interaction length, causing about two-thirds of the hadrons to start showering in the ECAL before entering the HCAL. The crystal transverse size is $2.2 \times 2.2\,\text{cm}^2$, corresponding to 0.0174×0.0174 in the (η, ϕ) plane, for the front face of the barrel crystals, and $2.9 \times 2.9\,\text{cm}^2$ for the endcap crystals. Such high transverse granularity makes the ECAL particularly powerful in separating, e.g., two collimated photons from the decay of a high momentum resonance (e.g., the Higgs boson), or photons and hadrons from a highly energetic hadronic jet. The intrinsic energy resolution of the ECAL barrel was measured with an ECAL supermodule directly exposed to an electron beam [30], with no attempt to account for the tracker material in front of the ECAL. The relative energy resolution is parameterized as a function of the electron energy as:

$$\frac{\sigma}{E} = \frac{2.8\%}{\sqrt{E}} \oplus \frac{12\%}{E} \oplus 0.3\%, \tag{2.3}$$

where the energy E is measured in GeV.

As a complement to the ECAL, the preshower detector is installed in front of each endcap of the ECAL. It has two layers, each consisting of a lead (Pb) radiator followed

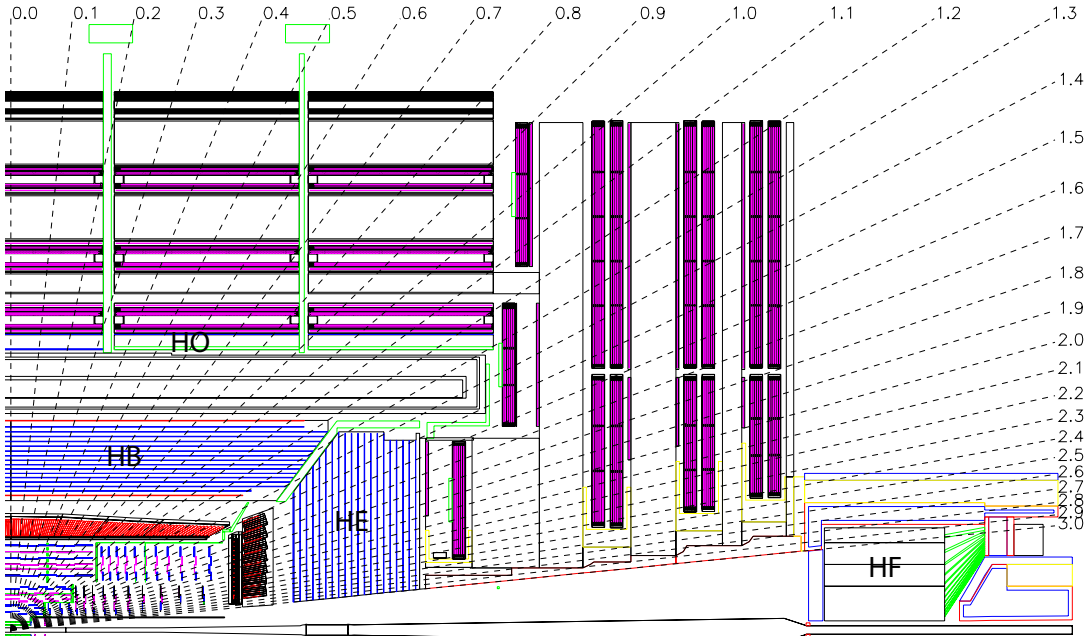


Figure 2.5: Cross-sectional view of a quadrant of the CMS detector in the (r, z) plane, showing various components of the HCAL systems. Figure from [22].

by a plane of silicon strip sensors. The depths of the two Pb radiators correspond to approximately two and one radiation lengths, respectively. The two planes of silicon sensors have orthogonal strips with a pitch of 1.9 mm. The high granularity of the preshower detector can provide an accurate measurement of the shower position, useful for e.g., discriminating the two collimated photons resulting from a π^0 decay from a single prompt photon.

The hadron calorimeter

The HCAL [31] is a hermetic sampling calorimeter. It is built around the ECAL inside the solenoid, with a barrel covering $|\eta| < 1.3$ and two endcaps covering $1.3 < |\eta| < 3.0$. The HCAL consists of alternating layers of brass absorber and plastic scintillator tiles. The brass absorber has a radiation length of 1.49 cm and a nuclear interaction length of 16.42 cm. The absorber thickness corresponds to about six interaction lengths in the

barrel and increases to over ten interaction lengths at larger pseudorapidities. The HCAL Outer (HO), installed outside the solenoid, serves as an additional absorber covering $|\eta| < 1.3$ and corresponding to 1.4 interaction lengths at normal incidence. This is further enhanced in the very central region by a 20 cm-thick layer of steel. Taking all of them into account (including also the ECAL), the total depth of the CMS calorimeter system reaches more than twelve (ten) interaction lengths in the barrel (endcaps). The HCAL has a much more coarse segmentation compared to the ECAL, 0.087×0.087 for $|\eta| < 1.6$ and 0.17×0.17 at larger pseudorapidities in the (η, ϕ) plane. The combined (ECAL+HCAL) calorimeter energy resolution was measured in a pion test beam [32] to be

$$\frac{\sigma}{E} = \frac{85\%}{\sqrt{E}} \oplus 7\%, \tag{2.4}$$

where the energy E is expressed in GeV.

The hadron forward (HF) calorimeter, installed at ± 11.2 m from the interaction point, further extends the angular coverage to $|\eta| \simeq 5$. The HF is a sampling calorimeter, consisting of grooved steel plates as the absorber and quartz fibers, inserted in the grooves along the beam direction, as the active material. Photomultipliers are used to read out the Cherenkov light signal, generated by charged shower particles above the Cherenkov threshold ($E \geq 190$ keV for electrons).

The layout of the CMS hadron calorimeter system is illustrated in Figure 2.5.

2.2.3 Muon System

The muon system [34, 33] is the outermost component of the CMS detector, also consisting of a barrel and two endcaps. As illustrated in Figure 2.6, the muon detection planes are interleaved with the steel yoke used to return the magnetic flux of the solenoid. Three types of gaseous particle detectors are used in the muon system. Drift tube (DT)

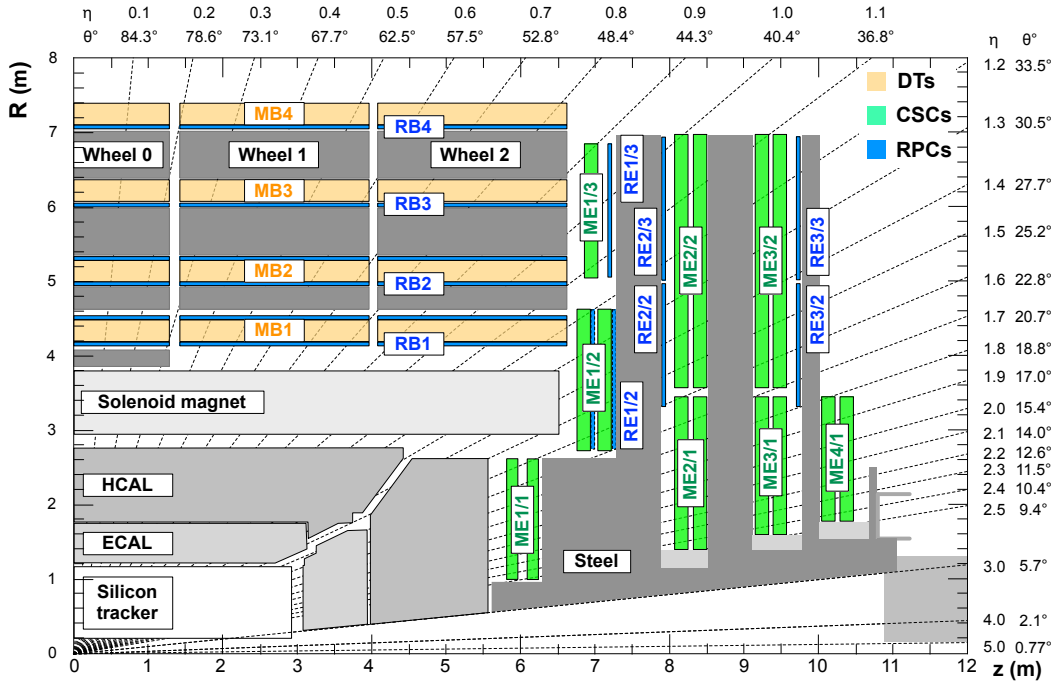


Figure 2.6: Cross-sectional view of a quadrant of the CMS detector in the (r, z) plane, showing the layout of the muon detector. Figure from [33].

chambers and cathode strip chambers (CSC) are used to detect muons in the barrel region ($|\eta| < 1.2$) and the endcap regions $0.9 < |\eta| < 2.4$, respectively. The DT and CSC subsystems not only provides muon identification and momentum measurement, but can also trigger on the p_{T} of muons with excellent efficiency and high background rejection, independent of the rest of the detector. The third type of gaseous detector, the resistive plate chambers (RPC), is installed in both the barrel and the endcaps as a redundant trigger system covering $|\eta| < 1.6$. The RPCs produce a fast response, with good time resolution in resolving muons from different bunch crossings, but their position resolution is coarser than the DTs or CSCs. The RPCs also help to resolve ambiguities when attempting to construct muon tracks from multiple hits in a chamber.

2.2.4 Triggering and Data Acquisition

At the LHC, proton beams cross at an interval of 25 ns, meaning that collision events are produced at an extremely high rate of 40 MHz. Clearly, it is impossible to store and process the large amount of data associated with all of these events. On the other hand, most events are from low-energy inelastic collisions and therefore are not of much interest. At CMS, the trigger system is designed to drastically reduce the event rate by selecting only events that are more likely to be important. A two-level trigger system is adopted by the CMS experiment, including a Level-1 (L1) Trigger [35] which first reduces the rate to less than 100 kHz, and a High-Level Trigger (HLT) [36] that further reduces the rate to ~ 1 kHz. The L1 Trigger analyzes each beam crossing using coarsely segmented data from the calorimeters and the muon system while holding the high-resolution data in pipelined memories in the front-end electronics. In less than 3.2 μs, the L1 Trigger must decide on whether to reject an event or to accept it for further evaluation, and transmit it to the detector front-end electronics. To meet such a strict requirement on the latency of the system, the L1 Trigger is solely based on specialized hardware. For reasons of flexibility, field-programmable gate arrays (FPGA) are most commonly used to implement the L1 Trigger, but application-specific integrated circuits (ASIC) and programmable memory lookup tables (LUT) are also widely used where the requirements on speed, density, and radiation resistance are higher. Events accepted by the L1 Trigger are sent to the HLT for further evaluation. At the HLT, more sophisticated algorithms, similar to those used in offline event reconstruction, are implemented to analyze each event. Readouts from the full detector can be incorporated, including information from the tracker. The HLT operates on a large computing farm, with about 22000 CPU cores in 2016.

2.3 Event Reconstruction

Data acquired by the CMS detector are processed by a set of advanced algorithms to identify the particle content of the event and to reconstruct the properties of each identified particle. Often, a particle leaves traces in more than one subdetectors, as illustrated in Figure 2.7. Therefore, simultaneously utilizing information from all subdetectors is likely to improve the particle identification and reconstruction performance significantly. In CMS, the particle-flow (PF) reconstruction algorithm [37] is designed to combine information from all subdetector systems to identify and reconstruct physics objects. The PF algorithm provides a global event description that leads to unprecedented performance for jet reconstruction and missing transverse momentum ($p_{\mathrm{T}}^{\mathrm{miss}}$) determination, as well as efficient mitigation of pileup contamination.

2.3.1 Reconstruction of Particle-Flow Elements

The PF algorithm starts with building charged-particle trajectories (tracks) and finding calorimeter clusters. Together, tracks and clusters form the basis of the PF algorithm and are combined later to reconstruct various kinds of particle candidates.

Track finding

The track reconstruction is based on an iterative tracking approach [26]. A combinatorial track finder based on Kalman Filtering (KF) [38] is used to build tracks from hits on the pixel and strip layers of the inner tracker. The combinatorial track finder runs in multiple iterations. In each iteration, initial seeds are first generated with only a few (2 or 3) hits compatible with a charged-particle trajectory, followed by a KF-based track building process to collect hits from all tracker layers compatible with the extrapolated charged-particle trajectory. A fitting procedure is subsequently performed to determine

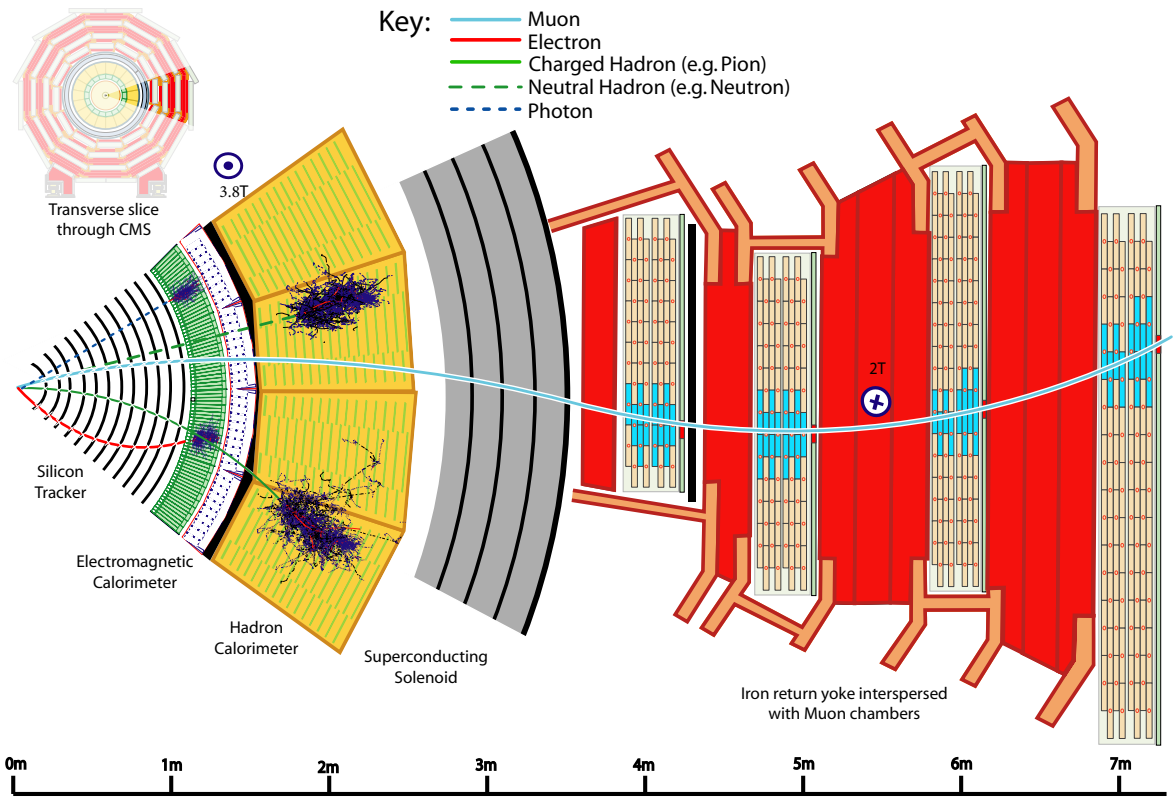


Figure 2.7: A sketch of the signatures of different types of particles in the CMS detector. Figure from [26].

all the track parameters and their uncertainties. Finally, track selection criteria are applied to set quality flags for the identified tracks and reject low-quality tracks. After each iteration, hits that are already associated with tracks are removed, and the next iteration, with a different seeding strategy and final selection criteria, is performed. The iterative tracking approach significantly increases the track finding efficiency while still preserving a low misreconstruction rate.

Muon tracks

Muon tracks are reconstructed using both the inner tracker and the outer muon detector. Depending on the different roles of the tracker and the muon detector, reconstructed

muons fall into three categories:

- *Standalone muon.* The standalone muons are reconstructed using only the muon detector. Muon tracks are seeded by hits within each DT or CSC detector, and then a track building process is performed to collect all DT, CSC, and RPC hits along the muon trajectory, followed by a fit to extract the muon track parameters.

- *Global muon.* If the track parameters of a standalone muon are compatible with that of a track in the inner tracker (when both are propagated onto a common surface), then the hits from both tracks are combined and fit together to form a global-muon track. Including hits on the muon detectors in the fit improves the momentum resolution for high-p_{T} muons.

- *Tracker muon.* Each inner track with $p_{\mathrm{T}} > 0.5\,\mathrm{GeV}$ and a total momentum $p > 2.5\,\mathrm{GeV}$ is extrapolated to the muon system. If at least one muon segment matches the extrapolated track, the inner track is considered as a tracker-muon track. This approach is particularly efficient for low-p_{T} muons.

Electron tracks

Electron tracks are reconstructed using information from both the inner tracker and the ECAL. Two approaches with different seeding strategies are used. The ECAL-based approach uses energetic ECAL clusters as initial seeds to infer the position of the expected hits in the innermost tracker layers and find compatible track seeds. Bremsstrahlung photons, due to interaction between the electron and the tracker material, must be taken into account in this process. Since the electron is bent in the magnetic field while the bremsstrahlung photons are not, the energy deposition of the photons is more spread in the ϕ direction. Therefore, superclusters that combine clusters in a small window in η

and an extended window in ϕ are used to collect the energy of the electron and possible bremsstrahlung photons. Superclusters with $E_{\mathrm{T}} > 4\,\mathrm{GeV}$ are used as seeds.

While the ECAL-based approach works well for high p_{T} and well-isolated electrons, for less isolated electrons such as those in jets, or electrons with low p_{T}, it shows large inefficiencies. A tracker-based electron seeding method performs better in such a scenario. It considers all tracks from the iterative tracking with $p_{\mathrm{T}} > 2\,\mathrm{GeV}$ as potential seeds for electrons. The large probability for electrons to radiate in the tracker material can be exploited to disentangle electron tracks from charged hadron tracks. With increasing energy loss due to radiation, the track may contain fewer hits or have a higher fit χ^2. Therefore, a preselection based on the number of hits, and the χ^2 is applied and the selected tracks are fit again with a Gaussian-sum filter (GSF) [39], which is more suitable than the KF for electrons as it allows for sudden and substantial energy losses along the trajectory. The final selection for the tracker-based electrons is based on a boosted-decision-tree (BDT) classifier that combines several variables, including the number of hits, the χ^2 of the GSF track fit and its ratio to that of the KF track fit, the energy loss along the GSF track, and the distance between the extrapolation of the track to the ECAL inner surface and the closest ECAL cluster.

In the PF algorithm, electron seeds obtained with both the ECAL-based and the tracker-based methods are merged, and a full GSF tracking with twelve components is performed on this collection of seeds.

Calorimeter clusters

The clustering algorithm is performed separately in each component of the calorimeter system: ECAL barrel and endcaps, HCAL barrel and endcaps, and the two preshower layers. The clustering algorithm is seeded by calorimeter cells with an energy above a threshold and larger than the energy of the neighboring cells. Then, topological clusters

are grown from the seeds by adding neighboring cells with an energy above twice the noise level. An expectation-maximization (EM) algorithm based on a Gaussian-mixture model is then used to disentangle the clusters within a topological cluster. The energy deposition of a topological cluster is modeled as a sum of N Gaussian energy deposits, where N corresponds to the number of seeds. The parameters of the Gaussian, i.e., the amplitude and the center (mean) coordinates in the (η, ϕ) plane, are extracted with the EM algorithm. Dedicated methods are developed to calibrate the calorimeter clusters, as an accurate measurement of their energy is crucial for a consistent global event description.

2.3.2 Particle Identification and Reconstruction

The linking algorithm

For a typical particle, multiple PF elements (tracks and clusters) are often produced in different subdetectors, and these elements need to be linked together to fully reconstruct the particle and avoid double-counting in different subdetectors. In the PF algorithm, a link between a track and a calorimeter cluster is established in the following way. The track is extrapolated from its last hit in the tracker, first to the two layers of the preshower, then to the ECAL at a depth corresponding to the expected maximum of a typical longitudinal electron shower profile, and finally to the HCAL at a depth corresponding to one interaction length. If the extrapolated position of the track is within the boundary of a cluster, then they are considered as linked. The link distance, defined as the distance between the extrapolated position and the cluster position in the (η, ϕ) plane, provides a metric to quantify the quality of the link.

In order to collect the energy of Bremsstrahlung photons emitted by electrons, tangents to the tracks are extrapolated to the ECAL from the intersection points between

the track and each of the tracker layers. A cluster is linked to the track as a potential Bremsstrahlung photon if the extrapolated tangent position is within the boundaries of the cluster.

Linking between ECAL clusters and HCAL clusters, and between ECAL clusters and preshower clusters is also performed. A link is established if the cluster position in the more granular calorimeter (ECAL or preshower) is within the boundary of a cluster in the other calorimeter (HCAL or ECAL). The link distance is defined similarly as that of the track-cluster link.

Finally, linking is also performed between tracks reconstructed in the inner tracker and segments reconstructed in the muon detector, as described in Section 2.3.1.

The output of the link algorithm is a collection of PF blocks, with each block consisting of elements associated either by a direct link or by an indirect link through common elements. Owing to the high granularity of the CMS detector, most of the PF blocks contain only a handful of elements originating from one or few particles. From each PF block, particle candidates are identified and reconstructed in the following order: Muon candidates have a very clear signature and are reconstructed first. This is followed by electrons and isolated photons, identified in the same step. The remaining elements are then identified and reconstructed as charged hadrons, neutral hadrons, and (nonisolated) photons. A list of mutually exclusive PF candidates is produced at the end.

Muons

In the PF algorithms, muons are reconstructed with the inner tracker and the outer muon detector as described in Section 2.3.1. Additional identification criteria, based on various quality parameters from the muon reconstruction, as well as the isolation of the muon, are applied to the standalone, global, or tracker muon candidates in order to effectively suppress misidentified muons (e.g., punch-through charged hadrons) while

preserving a high efficiency for both the isolated mouns and muons inside jets. The muon momentum is chosen to be that of the inner track if $p_{\mathrm{T}} < 200\,\mathrm{GeV}$. Above this threshold, the momentum is determined by the fit with the smallest χ^2 of the following: tracker only, tracker and first muon detector plane, global, and global without the muon detector planes featuring a high occupancy.

Electrons and isolated photons

An electron candidate is seeded from a GSF track if the corresponding ECAL cluster is not linked to three or more additional tracks. A photon candidate is seeded from an ECAL supercluster with $E_{\mathrm{T}} > 10\,\mathrm{GeV}$ and not linked to any GSF track. To account for the energy loss due to bremsstrahlung photons and photon conversion to $\mathrm{e}^+\mathrm{e}^-$ pairs, all ECAL clusters in the PF block linked either to the supercluster or to one of the GSF track tangents are associated with the candidate. Tracks linked to these ECAL clusters are associated if the track momentum and the energy of the corresponding ECAL cluster are compatible with the electron hypothesis. The tracks and ECAL clusters belonging to identified photon conversions linked to the GSF track tangents are associated as well. The total energy of the collected ECAL clusters is corrected for the energy missed in the association process. For the photon, this corrected energy is assigned as its energy, while its direction is defined by the primary vertex and the centroid of the supercluster. Additional identification criteria based on the isolation, the energy distribution on ECAL cells, and the ratio between the HCAL and ECAL energies are applied. For the electron, the final energy is obtained from a combination of the corrected ECAL energy with the momentum of the GSF track, and its direction is chosen to be that of the GSF track. Additional identification criteria based on a BDT combining properties of the GSF track, the ECAL cluster, the track-cluster association, etc., are applied.

Hadrons and nonisolated photons

Once muons, electrons, and isolated photons are identified and removed from the PF blocks, the remaining are mostly hadrons and nonisolated photons from quark fragmentation and hadronization. They are identified broadly into a few categories, namely charged hadrons, neutral hadrons, and photons in the PF algorithm, without any attempt to uncover the specific species of the particle (e.g., pions, Koans, protons). These PF candidates are essential ingredients for jet reconstruction, as will be discussed in Section 2.3.3.

The ECAL and HCAL clusters not linked to any track are reconstructed as photons and neutral hadrons. Within the tracker acceptance ($|\eta| < 2.5$), all these ECAL clusters are turned into photons, and all these HCAL clusters are turned into neutral hadrons. The precedence is given to photons for the ECAL clusters, because in hadronic jets, 25% of the jet energy is carried by photons, while neutral hadrons leave only 3% of the jet energy in the ECAL. Outside the tracker acceptance, such an assignment can no longer be justified as charged and neutral hadrons cannot be separated anymore, which together deposit $\sim$25% of the jet energy in ECAL, comparable to photons. Therefore, ECAL clusters linked to a given HCAL cluster are assumed to arise from the same (charged- or neutral-) hadron shower, while ECAL clusters without such a link are reconstructed as photons.

Remaining HCAL clusters of the PF block are linked to tracks, and these tracks may, in turn, be linked to some of the remaining ECAL clusters. If the calibrated calorimetric energy is above the sum of the track momenta by a certain amount, the excess may be interpreted as the presence of photons and neutral hadrons. Each track is reconstructed as a charged hadron, whose momentum and energy are directly taken from the corresponding track momentum assuming a mass of the charged-pion. If the

calibrated calorimetric energy is compatible with the sum of the track momenta, no neutral particle is identified. The charged-hadron momenta are redefined by a χ^2 fit of the measurements in the tracker and the calorimeters, taking advantage of the better resolution of the tracker at low energy and the better resolution of the calorimeters at high energy.

2.3.3 Jets and Missing Transverse Momentum

Jets

Jets are collimated sprays of particles reconstructed with a clustering algorithm. They are typically used to reconstruct the showers of quarks or gluons and to infer the properties of the quarks or gluons in the hard-scattering process. In CMS, jets are typically reconstructed by clustering the PF candidates with the anti-k_{T} algorithm [40] implemented in the FastJet package [41]. The anti-k_{T} algorithm works as follows:

1. For each pair of particles, i, j, compute the anti-k_{T} distance

$$d_{ij} = \min(p_{\mathrm{T},i}^{-2}, p_{\mathrm{T},j}^{-2}) \frac{\Delta R_{ij}^2}{R^2} \tag{2.5}$$

   with $\Delta R_{ij}^2 \equiv (y_i - y_j)^2 + (\phi_i - \phi_j)^2$, where y_i, ϕ_i are the rapidity and the azimuth of particle i, respectively. R is a distance parameter related to the radius of the jet cone. For each particle i, compute also the beam distance

$$d_{iB} = p_{\mathrm{T},i}^{-2} \tag{2.6}$$

2. Find the minimum, d_{min}, of all the particle distances d_{ij} and beam distance d_{iB}. If d_{min} is a d_{ij}, then merge particles i and j into a single particle and sum up their

four-momenta; otherwise, i.e., if d_{min} is a d_{iB}, then particle i is declared to be a final jet and removed from the list.

3. The two steps are repeated until no particles are left.

The distance parameter R is chosen based on the use cases of the resulting jet collection. The jet collection clustered using the anti-k_{T} algorithm with a small distance parameter $R = 0.4$, referred to as the "AK4" jet collection, is the default jet collection in CMS. Due to the small R, each AK4 jet typically corresponds to a single quark or gluon from the hard-scattering process. Jet collections produced with larger R values, such as $R = 0.8$ and $R = 1.5$ (referred to as "AK8" and "AK15" jets, respectively), are widely used to reconstruct Lorentz-boosted hadronically decaying massive particles (top quarks and W, Z and Higgs bosons). When these particles are produced with a high boost, their decay products become collimated and can be contained in a single large-R jet. The identification of these hadronically decaying boosted particles with large-R jets is the subject of Part II.

Particles originating from pileup (additional proton-proton collisions) can cause significant contamination to the reconstructed jets. For the AK4 jets, effects due to pileup are mitigated with the charged-hadron subtraction (CHS) method [42], which removes the charged PF candidates identified as originating from pileup interactions before the jet clustering. An event-by-event jet area–based correction is applied to the jet four-momentum to remove the remaining (neutral-particle) energy from pileup interactions. For the large-R jets, the pileup contamination is mitigated with the "PileUp Per Particle Identification (PUPPI)" method [43], which estimates the probability for each particle to originate from the primary interaction vertex based on local energy distribution around the particle, event pileup properties and tracking information (for charged particles) and uses this estimation to scale the four-momentum of each PF candidate, before clustering

them into jets. A detailed description of the implementation of the PUPPI algorithm in CMS can be found in Ref. [44]. For large-R jets, the PUPPI method shows improved performance and robustness compared to the CHS method, particularly for high pileup.

Reconstructed jets are calibrated to have the correct energy scale. The jet energy scale calibration [42] is performed in a series of steps. The first step corrects for the offset energy due to pileup. The offset corrections are derived from the simulation of dijet events processed with and without pileup overlaid. Residual differences between data and simulation are determined using the random cone method in zero-bias events. This pileup offset correction step is applied only for CHS jets, while PUPPI jets do not need such a correction. Then, effects due to detector response are corrected using correction factors derived from simulated samples as a function of the jet p_{T} and η. Last, residual differences between data and simulation are measured in Z+jets ($\mathrm{Z} \to \mathrm{ee}$ or $\mathrm{Z} \to \mu\mu$) events, photon+jets events, and dijet events, and corrections are applied to data to account for the difference. Uncertainties on the jet energy correction are also derived in the last step.

Missing transverse momentum

The missing transverse momentum vector $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ is defined as the negative of the vector sum of the p_{T} of all the PF candidates in an event:

$$\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}} = -\sum_{i} \vec{p}_{\mathrm{T},i}. \tag{2.7}$$

Its magnitude is denoted as $p_{\mathrm{T}}^{\mathrm{miss}}$. The $p_{\mathrm{T}}^{\mathrm{miss}}$ measures the undetected transverse momentum in the event, which can arise from neutrinos in the SM, or new particles that do not interact with any SM particles as predicted in some theories beyond the standard model. As $p_{\mathrm{T}}^{\mathrm{miss}}$ is reconstructed using all the PF candidates in an event, it is sensitive to

the responses of all parts of the detector. To improve the estimation of $p_{\mathrm{T}}^{\mathrm{miss}}$, corrections derived for the (AK4) jets can be propagated to $p_{\mathrm{T}}^{\mathrm{miss}}$ in the following way [45]:

$$\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}} = \vec{p}_{\mathrm{T}}^{\,\mathrm{miss,\,raw}} - \sum_{\mathrm{jets}} \left(\vec{p}_{\mathrm{T,\,jet}}^{\,\mathrm{corr}} - \vec{p}_{\mathrm{T,\,jet}}^{\,\mathrm{raw}} \right), \tag{2.8}$$

where $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss,\,raw}}$ refers to the uncorrected $p_{\mathrm{T}}^{\mathrm{miss}}$ as defined in eq. (2.7), and the sum is over jets with $p_{\mathrm{T}} > 15\,\mathrm{GeV}$.

Part II

Identification of Highly Lorentz-Boosted Hadronically Decaying Particles

The identification ("tagging") of highly Lorentz-boosted hadronically decaying massive particles (top quarks and W, Z, and Higgs bosons) has become an increasingly important topic at the LHC, as it provides powerful handles for both searches for new physics and probes of the SM in the high-momentum regime. This part of the thesis discusses some new developments in this area. We begin with a brief introduction of jet substructure techniques and machine learning methods in Chapter 3. Then, top quark and W boson tagging algorithms based on boosted decision trees (BDT), developed for a search for supersymmetric partners of the top quark, is presented in Chapter 4. A continuation of this development resulted in the DeepAK8 algorithm, which is an advanced multi-class identification algorithm for boosted top quarks and W, Z, and Higgs bosons using particle-level inputs and deep neural networks (DNN). The DeepAK8 algorithm is described in detail in Chapter 5. A more novel machine learning approach for such identification, ParticleNet, which treats a jet as an unordered set of particles and exploits the permutation symmetry explicitly in the neural network, is introduced in Chapter 6.

Chapter 3

Introduction

A jet is a collimated spray of particles. It is often associated with the production of a colored particle, i.e., quark or gluon, whose evolution is governed by the strong interaction. The parton (i.e., quark or gluon), produced in the hard-scattering process of the proton-proton collision, keeps radiating as it travels, and the emitted partons also radiate, therefore resulting in a "parton shower", i.e., a cascade of quarks and gluons, at the end. These partons eventually hadronize and become color-neutral particles that constitute a jet. Therefore, a jet serves as a handle to probe the elementary particle from the hard-scattering process that initiates the jet.

Jets are ubiquitously produced at the LHC due to the proton-proton collision nature of the machine. The vast majority of the jets, though, are initiated by a single quark or gluon, and therefore are not of particular interest here. On the other hand, a high-momentum electroweak scale particle, such as the top quark and the W, Z, and Higgs boson, with a subsequent hadronic decay, results in a triplet or pair of highly collimated quarks, which then can lead to a single large-radius jet, instead of several well-separated jets that correspond to each individual quark. Such large-radius jets initiated by highly Lorentz-boosted hadronically decaying massive particles have characteristic ra-

diation patterns. Therefore, they can be distinguished from the ubiquitous jets initiated by a single quark or gluon, by studying the internal structure, or "substructure", of the jets.

The study of jet substructure has become a very active research area on both the theoretical and experimental sides. For LHC experiments, jet substructure techniques are now commonly used in searches for new heavy particles that subsequently decay to highly boosted top quarks or W, Z, and Higgs bosons (e.g., [46, 47, 48, 49]), and have been increasingly adopted in measurements of the SM processes (e.g., [50, 51]) as well, particularly in the high-momentum regime. On the theoretical side, many new substructure observables and techniques have been proposed in recent years, significantly improving the discrimination power and the robustness of the jet substructure–based tagging algorithms. The quest for better jet substructure techniques has triggered a renewed interest in Quantum Chromodynamics (QCD), thus deepening our understanding of the strong interaction. Most recently, advance machine learning techniques, particularly deep learning, have brought further progress and deeper insights into jet substructure. A comprehensive review of the current status of jet substructure is provided in Ref. [52] and [53] on the theoretical and experimental sides, respectively. A more pedagogical overview of jet substructure can be found in Ref. [54].

In this chapter, we first discuss some substructure techniques that are widely used in the CMS experiment. They serve as the basis for the new developments and provide a baseline to evaluate the performance of the new algorithms. Then, a brief introduction to machine learning is presented in Section 3.2, with an emphasis on its application to jet tagging.

3.1 Jet Substructure

For the identification of boosted massive particles, AK8 jets, introduced in Section 2.3.3, are typically used in CMS. These AK8 jets are obtained by clustering the PF candidates using the anti-k_{T} algorithm with a distance parameter $R = 0.8$. As the angular separation between the decay products of a Lorentz boosted particle with mass m scales like $\Delta R \sim 2m/p_{\mathrm{T}}$, jets clustered with $R = 0.8$ can in principle contain all decay products from W and Z bosons with $p_{\mathrm{T}} \gtrsim 200\,\mathrm{GeV}$, Higgs bosons with $p_{\mathrm{T}} \gtrsim 300\,\mathrm{GeV}$ and top quarks with $p_{\mathrm{T}} \gtrsim 400\,\mathrm{GeV}$.

3.1.1 Groomed Mass

The jet mass, defined as the invariant mass of the sum of the four-momenta of the jet constituents, is one of the most powerful observables to separate signal jets (originating from a top quark or a W, Z or Higgs boson) and background jets (originating from a gluon or a light quark). At leading order in the perturbation theory, the quark and gluon jets should have a mass closer to zero, while the mass of the signal jets should be much higher, close to the intrinsic mass of the top quark or the W, Z or Higgs boson. In reality, the picture is much more complicated, as effects due to initial state radiation, the underlying event (interactions between partons other than the hard-scattering ones) and pileup (additional proton-proton interactions) can contribute significantly to the reconstructed jet mass, especially for large-radius jets.

Methods have been developed to remove soft or uncorrelated radiation from jets, called "grooming" methods. The grooming method most commonly used in CMS is the "modified mass drop tagger" algorithm (mMDT) [55], which is a special case of the "soft drop" (SD) algorithm [56]. The SD algorithm operates by first reclustering the jet constituents with the Cambridge-Aachen (CA) [57, 58] algorithm, and then declustering

the CA jet to remove wide-angle soft radiation. The declustering starts from breaking the CA jet into two subjets by undoing the last stage of the CA clustering and then checks the SD condition:

$$\frac{\min(p_{\mathrm{T1}}, p_{\mathrm{T2}})}{p_{\mathrm{T1}} + p_{\mathrm{T2}}} > z_{\mathrm{cut}} \left(\frac{\Delta R_{12}}{R_0} \right)^{\beta}, \tag{3.1}$$

where p_{T1} (p_{T2}) is the p_{T} of the leading (sub-leading) subjet and ΔR_{12} is their angular separation. The parameters z_{cut} and β define what the algorithm considers "soft" and "collinear," respectively. The values used in CMS are $z_{\mathrm{cut}} = 0.1$ and $\beta = 0$ (making this identical to the mMDT algorithm). If the SD condition is not met, the lower-p_{T} subjet is removed, and the same procedure is repeated until eq. (3.1) is satisfied, or no further declustering can be performed.

The two subjets returned by the SD algorithm are used to calculate the jet mass. Figure 3.1 shows the distribution of the jet mass after the SD grooming (m_{SD}) in signal and background jets in simulation. The m_{SD} of background jets peaks at low values and falls rapidly, whereas for signal jets a peak around the mass of the corresponding signal particle (top quark or W, Z, Higgs boson) is observed.

3.1.2 N-subjettiness

An additional handle to separate signal from background jets is to exploit the energy distribution inside the jet. For a boosted massive particle, the partons from the decay typically all carry a sizeable fraction of the initial particle's momentum. Therefore, the resulting jet tends to have multiple hard cores ("prongs"). In contrast, quark and gluon jets, which are dominated by the radiation of soft gluons, typically have only one hard

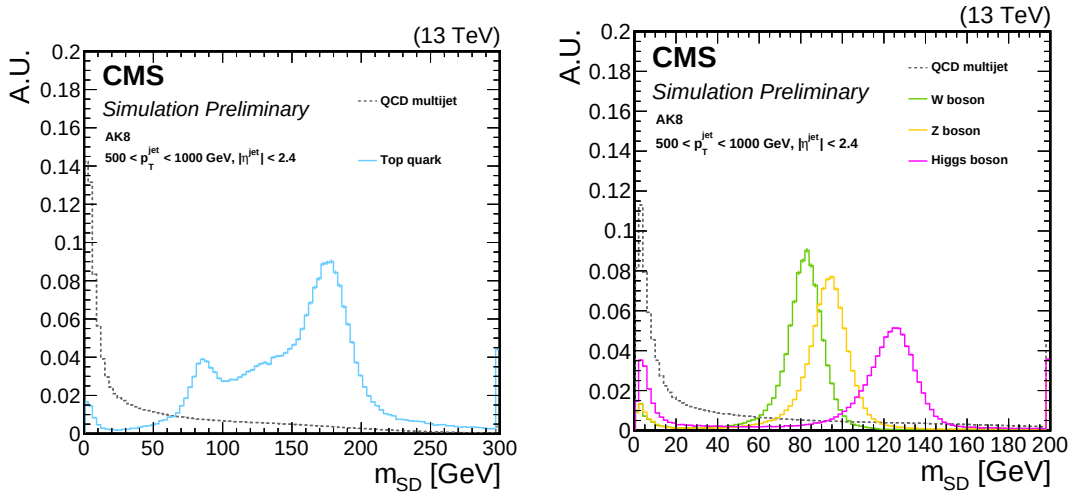


Figure 3.1: Comparison of the m_{SD} shape between signal and background AK8 jets in simulation.

core in each jet. The N-subjettiness variables [59, 60],

$$\tau_N = \frac{1}{d_0} \sum_i p_{\mathrm{T},i} \min \left[\Delta R_{1,i}, \Delta R_{2,i}, \ldots, \Delta R_{N,i} \right], \tag{3.2}$$

provide a measure of how compatible the jet is with having N "subjets", i.e., hard radiation cores. The index i refers to the jet constituents, while the ΔR terms represent the spatial distance between a given jet constituent and "subjets". The quantity d_0 is a normalization constant. The "subjets" are found by performing the exclusive k_T algorithm [61, 62] on the jet constituents before the application of any grooming techniques. A smaller τ_N value implies that the jet is more compatible with having N or more "subjets". Better discrimination power can be achieved by the ratio of different τ_N variables. For example, the ratio $\tau_{21} \equiv \tau_2/\tau_1$ can be used to identify 2-prong jets, e.g., those arising from W, Z or Higgs bosons, and $\tau_{32} \equiv \tau_3/\tau_2$ can be used for tagging the 3-prong top quark jets. The distributions of τ_{21} and τ_{32} for signal and background AK8 jets are shown in Figure 3.2.

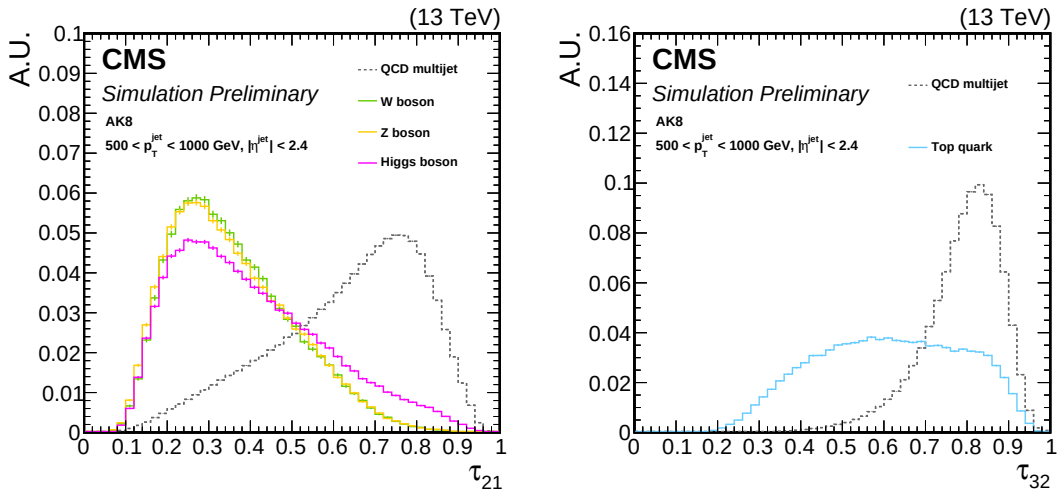


Figure 3.2: Comparison of the τ_{21} (left) and τ_{32} (right) shapes between signal and background AK8 jets in simulation.

3.2 Machine Learning Techniques

Broadly speaking, machine learning (ML) involves algorithms and statistical methods that allow a computer system to perform specific tasks without explicit instructions. A typical scheme is that the ML algorithm is first presented with a "training" dataset to build a mathematical model for the task, and then the "learned" model can be used to make predictions or decisions for previously unseen data. The current machine learning algorithms roughly fall into three categories, as illustrated in Figure 3.3. For supervised learning, a target output is always provided for each input object during the training phase, and the goal of the algorithm is to produce a model that can predict the output as closely as possible. Depending on whether the target output is discrete or continuous, supervised learning is further subdivided into classification (for discrete output) and regression (for continuous output). Within the context of ML, jet tagging is an ideal task for classification algorithms. Another example in particle physics is to distinguish between signal and background events, for which boosted decision trees (BDT) have been extensively used for many years. Another category is the so-called unsupervised

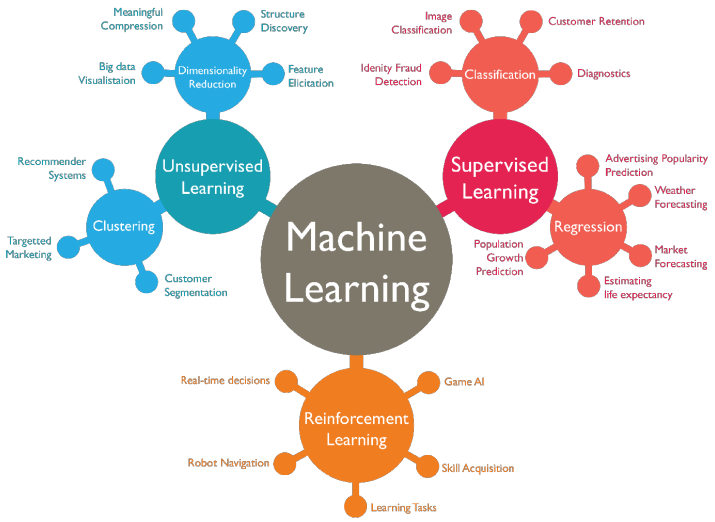


Figure 3.3: Three main approaches in machine learning and examples of the corresponding applications. Image from [63].

learning, for which no target output is provided (or can even be unambiguously defined) for each input object, and the algorithms are supposed to find emergent patterns from the inputs. An example of unsupervised learning in jet physics can be found in Ref. [64, 65]. Reinforcement learning, on the other hand, involves how the system should take a series of actions in response to an environment to maximize some notion of reward. This has been applied to study jet grooming recently in Ref. [66].

In recent years, deep learning [67], characterized by the use of big training datasets, often containing millions of samples, and deep neural networks (DNN) with millions of parameters, has become the prevailing ML approach due to the tremendous improvements it brought. This has certainly inspired research in high energy physics, especially jet physics where tools and algorithms are constantly being improved by both theorists and experimentalists. Indeed, the application of deep learning techniques brings significant performance improvements, as will be presented in Chapter 5 and Chapter 6. Below we give a very brief introduction to the basics of deep learning, with a focus on its usage in a classification task such as jet tagging. An accessible introduction and hands-on tutorial

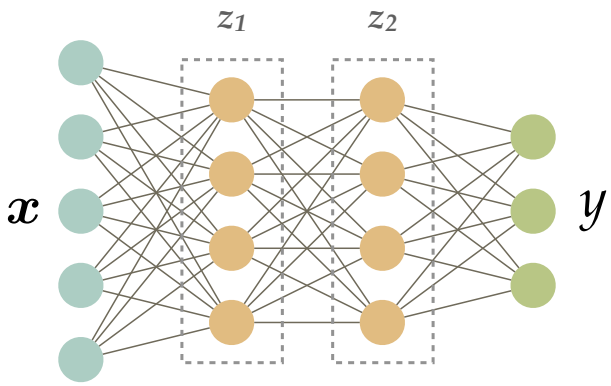


Figure 3.4: Illustration of a simple neural network with two hidden layers.

to deep learning is available in Ref. [68], and a more in-depth treatment of this subject can be found in Ref. [69].

Mathematically, a neural network is a parameterized function,

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x}; \boldsymbol{\theta}), \tag{3.3}$$

where $\boldsymbol{x}$ is the (high-dimensional) input vector, $\boldsymbol{y}$ is the output (vector), and $\boldsymbol{\theta}$ represents the set of parameters that will be learned from the training data. The functional form of f depends on the architecture of the neural network. One of the simplest architectures, called a "fully-connected" (or "dense") network, is illustrated in Figure 3.4. It consists of four "layers": one input layer, one output layer, and two intermediate ("hidden") layers. Except for the input layer, each (hidden or output) layer is a mapping[1], typically consisting of a linear transformation followed by a point-wise (nonlinear) function ("activation"), in the form

$$\boldsymbol{z} = \sigma(\boldsymbol{W}\boldsymbol{x} + \boldsymbol{b}), \tag{3.4}$$

where $\boldsymbol{W}\boldsymbol{x} + \boldsymbol{b}$ is a linear transformation with learnable parameters $\boldsymbol{W}$ ("weights") and $\boldsymbol{b}$ ("biases"), and $\sigma(\cdot)$ denotes the activation function. A nonlinear activation function is

[1]The input layer can be considered as an identity map.

necessary for hidden layers, as otherwise, a stack of linear transformation is equivalent to one linear transformation with a different set of parameters. For hidden layers, the most commonly used activation function nowadays is the Rectified Linear Unit (ReLU) [70]:

$$\mathrm{ReLU}(x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{3.5}$$

while other variations, such as ELU [71], SELU [72], LeakyReLU [73], and PReLU [74] are also widely used. The choice of the activation function for the output layer depends on the task. For a binary classification task (i.e., discriminating between "signal" and "background"), the sigmoid activation function

$$\mathrm{sigmoid}(z) = \frac{1}{1 + e^{-z}} \tag{3.6}$$

is often used, producing an output within $[0, 1]$ which can be interpreted as a probability. For multi-class classification tasks, the softmax activation function

$$\mathrm{softmax}(z)_i = \frac{e^{z_i}}{\sum_j e^{z_j}} \tag{3.7}$$

is typically used, where i denotes the ith class. The outputs are all within $[0, 1]$ and therefore can also be considered as probabilities. A *deep* neural network consists of many layers, and can be used to approximate a very complex function.

The training of a DNN can be viewed as an optimization problem, or a "fit" to the training dataset, which is a more familiar concept to particle physicists. The goal is to find an optimal set of parameters that minimize an objective function, often referred to as a "loss function" in ML. For classification tasks, the cross-entropy loss function is most

widely used:

$$\mathrm{CE}(\boldsymbol{\theta}) = -\sum_i \hat{y}_i \log y_i(\boldsymbol{\theta}), \tag{3.8}$$

where $\hat{y}_i$ is the truth label for the ith class, which equals 1 if the sample[2] is from the ith class and 0 otherwise, and y_i is the DNN output for the ith class, after applying the softmax activation function. A lower loss function indicates a more accurate prediction. The minimization of the loss function is typically carried out via a mini-batch stochastic gradient descent (SGD) method. The whole training dataset is divided into mini-batches with a typical mini-batch size of $\mathcal{O}(10)$ to $\mathcal{O}(1000)$. For each mini-batch, the total loss function $L(\boldsymbol{\theta})$ is evaluated as the sum of the loss function of each sample, and the gradient of the total loss function with respect to the parameters, $\nabla L(\boldsymbol{\theta})$, is computed and used to update the parameters. In the simplest form, the parameter update is formulated as:

$$\boldsymbol{\theta}' = \boldsymbol{\theta} - \frac{\eta}{n}\nabla L(\boldsymbol{\theta}), \tag{3.9}$$

where n is the size of the mini-batch, and η is the learning rate that controls the magnitude of the update. In practice, more sophisticated algorithms, such as Adam [75] or SGD with momentum [76], are used to achieve better convergence and reduce the minimization time. In addition, the choice of the parameter initialization method, the hyperparameters such as the learning rate, the mini-batch size, the regularization method, and particularly the neural network architecture, such as the depth of the network and the number of units in each layer, have significant impact on the final performance, and often need to be carefully tuned to achieve the best performance. Some helpful guidelines on the hyperparameter tuning for DNNs are provided in Ref. [77].

[2]Here a *sample* refers to a single item in the dataset.

Chapter 4

BDT-Based Top Quark and W Boson Tagging Algorithms

To improve the performance for identifying highly boosted top quarks and W bosons, BDT-based algorithms using jet-level observables were developed in the search for supersymmetric (SUSY) partners of the top quark in the all-jet final state [46]. The algorithms target hadronically decaying top quarks (W bosons) with $p_{\mathrm{T}} > 400$ (200) GeV, whose decay products are expected to be contained in a cone with $\Delta R < 0.8$. Therefore, AK8 jets are used to identify and reconstruct the boosted top quarks and W bosons.

An AK8 jet is required to pass a preselection before being considered as a candidate for a top quark or a W boson. The preselection requires the AK8 jet to satisfy $p_{\mathrm{T}} > 200\,\mathrm{GeV}$, $|\eta| < 2.4$, $m_{\mathrm{SD}} > 50\,\mathrm{GeV}$, and have exactly two subjets identified by the soft drop algorithm, both with $p_{\mathrm{T}} > 20\,\mathrm{GeV}$. For the identification of top quarks, the p_{T} threshold on the AK8 jet is raised to 400 GeV. To uniquely identify each AK8 jet, a requirement of $m_{\mathrm{SD}} > 110$ ($50 < m_{\mathrm{SD}} < 110$) GeV is further imposed for top quark (W boson) candidate. After preselection, two separate BDTs are developed for tagging boosted top quarks and W bosons. The input variables, training procedure, and performance of the BDTs are

described in the following sections.

4.1 Input Variables

A total of 14 (12) input variables are used in the BDT for top quark (W boson) tagging. The input variables include:

m_{SD}: The groomed mass of an AK8 jet is one of the most discriminative variables for separating signal jets (originating from top quarks or W bosons) from background jets (originating from QCD radiation), as the mass distribution of the signal jets peaks around the particle mass (m_{t} or m_{W}).

N-subjettiness ratios: The N-subjettiness ratios τ_{32} (τ_{21}) are powerful for identifying jets with 3 (2) hard radiation centers, which are characteristics for top quark (W boson) jets.

Kinematics of the subjets: Signal and background jets show differences in the kinematics of the subjets, such as the p_{T} balance, the angular separation, etc. A few variables are constructed to exploit such differences. The product of the p_{T} of the AK8 jet and the ΔR between the two subjets, $p_{\mathrm{T}} \cdot \Delta R$, is centered around $2m_{\mathrm{t}}$ or $2m_{\mathrm{W}}$ for signal jets, while for background the distribution is more widely spread. The p_{T} of the two subjets are also more balanced for signal jets. An observable inspired by the soft drop condition defined in eq. (3.1) with $\beta = -2$, $\mathrm{SD}^{\beta=-2}$, is found to be quite powerful for W boson tagging.

Quark and gluon jet discrimination variables: Three variables related to quark and gluon jet discrimination [78], i.e., the constituent multiplicity, the major axis σ_1, and the jet fragmentation distribution $p_{\mathrm{T}}D$, are used for each subjet. Subjets of

the signal jets are always quark-induced, while they can be gluon-induced for background jets. Therefore, the use of these variables brings additional improvement in performance.

Heavy flavor tagging discriminant: The identification of top quarks can benefit greatly from tagging the b quark coming from the top quark decay. Therefore, the CSVv2 [79] b-tagging discriminants of the subjets are also used in the BDTs. For W boson tagging, the use of b-tagging discriminants can help reject background jets originating from b quarks.

Note that variables strongly correlated with the p_{T} of the candidate, such as the ΔR of the two subjets, are not used in the training. Instead, $p_{\mathrm{T}} \cdot \Delta R$, which is largely uncorrelated with p_{T}, is used. This is to avoid introducing any bias on the candidate p_{T} distribution in the BDTs. As signal jets typically have a harder p_{T} spectrum than background jets in the training samples, the BDTs can learn to exploit such difference and favor high-p_{T} candidates as signals, when the input variables contain p_{T}-related information. However, as the mass scale of the SUSY particles is unknown, so is the p_{T} spectrum of the top quarks or W bosons from the decay of the SUSY particles. Therefore an unbiased tagging algorithm is much more desirable.

The full lists of input variables and their definitions are summarized in Table 4.1 and 4.2 for the top quark and W boson tagging BDTs, respectively. In the case of top quark tagging, the two subjets are ranked by the b-tagging discriminants and denoted "CSV_1" and "CSV_2" in Table 4.1. The reason for this choice is that "CSV_1", with a higher b-tagging discriminant, is more likely to correspond to the b quark from the top decay when the jet is originated by a real top quark. Therefore, ranking the two subjets in this way provides a more consistent distinction between the two subjets, and was found to provide better performance than other orderings. On the other hand, no such clear

distinction exists in the case of W boson tagging[1]. Therefore, the two subjets are simply ranked by p_{T} and denoted by "sj_1" and "sj_2" in Table 4.2.

Variable	Description
m_{SD}	soft drop mass of the AK8 jet
$p_{\mathrm{T}} \cdot \Delta R$	the product of the p_{T} of the AK8 jet and the ΔR between the two subjets
p_{T} asymmetry	normalized p_{T} asymmetry between the two subjets, $\lvert p_{\mathrm{T}}^{\mathrm{CSV}_1} - p_{\mathrm{T}}^{\mathrm{CSV}_2}\rvert / p_{\mathrm{T}}$
τ_{21}	N-subjettiness ratio, τ_2/τ_1
τ_{32}	N-subjettiness ratio, τ_3/τ_2
m_{CSV_1}	mass of the subjet leading in b-tagging discriminant
m_{CSV_2}	mass of the subjet subleading in b-tagging discriminant
$p_{\mathrm{T}}D(\mathrm{CSV}_1)$	jet fragmentation distribution, $p_{\mathrm{T}}D$, of the subjet leading in b-tagging discriminant
$p_{\mathrm{T}}D(\mathrm{CSV}_2)$	jet fragmentation distribution, $p_{\mathrm{T}}D$, of the subjet subleading in b-tagging discriminant
$\sigma_1(\mathrm{CSV}_1)$	major axis, σ_1, of the subjet leading in b-tagging discriminant
$\sigma_1(\mathrm{CSV}_2)$	major axis, σ_1, of the subjet subleading in b-tagging discriminant
multiplicity(CSV_1)	constituent multiplicity of the subjet leading in b-tagging discriminant
multiplicity(CSV_2)	constituent multiplicity of the subjet subleading in b-tagging discriminant
$\mathrm{CSV}_{\mathrm{max}}$	maximum b-tagging discriminant of the two subjets

Table 4.1: Variables used in the training of the top quark tagging BDT.

Variable	Description
m_{SD}	soft drop mass of the AK8 jet
$p_{\mathrm{T}} \cdot \Delta R$	the product of the p_{T} of the AK8 jet and the ΔR between the two subjets
p_{T} asymmetry	normalized p_{T} asymmetry between the two subjets, defined as $\lvert p_{\mathrm{T}}^{\mathrm{sj}_1} - p_{\mathrm{T}}^{\mathrm{sj}_2}\rvert / p_{\mathrm{T}}$
τ_{21}	N-subjettiness ratio, τ_2/τ_1
$\mathrm{SD}^{\beta=-2}$	soft drop condition with $\beta = -2$, defined as $\frac{\min(p_{\mathrm{T}}^{\mathrm{sj}_1}, p_{\mathrm{T}}^{\mathrm{sj}_2})}{p_{\mathrm{T}}^{\mathrm{sj}_1} + p_{\mathrm{T}}^{\mathrm{sj}_2}} \cdot \frac{1}{\Delta R^{-2}(\mathrm{sj}_1, \mathrm{sj}_2)}$
$p_{\mathrm{T}}D(\mathrm{sj}_1)$	jet fragmentation distribution, $p_{\mathrm{T}}D$, of the subjet leading in p_{T}
$p_{\mathrm{T}}D(\mathrm{sj}_2)$	jet fragmentation distribution, $p_{\mathrm{T}}D$, of the subjet subleading in p_{T}
$\sigma_1(\mathrm{sj}_1)$	major axis, σ_1, of the subjet leading in p_{T}
$\sigma_1(\mathrm{sj}_2)$	major axis, σ_1, of the subjet subleading in p_{T}
multiplicity(sj_1)	constituent multiplicity of the subjet leading in p_{T}
multiplicity(sj_2)	constituent multiplicity of the subjet subleading in p_{T}
$\mathrm{CSV}_{\mathrm{max}}$	maximum b-tagging discriminant of the two subjets

Table 4.2: Variables used in the training of the W boson tagging BDT.

[1] Actually, one possible handle to exploit here is the c quark, which is present in around 50% of the hadronic decays of the W boson. However, the existing flavor tagging algorithms for subjets at the time when this analysis was carried out were not powerful enough to bring any noticeable improvement.

4.2 Training Procedure

The top quark and W boson tagging BDTs are trained with simulated jets using the Toolkit for Multivariate Data Analysis (TMVA) [80]. For the top quark tagging BDT, the signal jets are those matched to the hadronic decays of the generated top quarks in a simulated SUSY sample enriched in highly boosted top quarks. The matching requires that a signal AK8 jet is close to a hadronically decaying top quark, with $\Delta R(\mathrm{AK8}, \mathrm{t}) < 0.8$, and all three quarks from the decay of the top quark are contained in the AK8 jet, satisfying $\Delta R(\mathrm{AK8}, \mathrm{q}) < 0.8$. The background jets are those that cannot be matched to a hadronically decaying top quark, selected from a simulated $\mathrm{t\bar{t}}$ sample instead, as the simulated SUSY sample contains very few un-matched AK8 jets.

A similar definition for signal and background jets are used for the training of the W boson tagging BDT. In this case, both signal and background jets are selected from a simulated $\mathrm{t\bar{t}}$ sample. Each signal jet is required to be matched to the hadronic decays of a generated W boson, with $\Delta R(\mathrm{AK8}, \mathrm{W}) < 0.8$, and both quarks from the W boson decay must be contained in the AK8 jet, i.e., $\Delta R(\mathrm{AK8}, \mathrm{q}) < 0.8$. The background jets are the remaining ones that cannot be matched.

4.3 Performance

A single working point is defined for each of the top quark and W boson tagging BDTs and used in the search for SUSY partners of the top quark [46]. The chosen working points correspond to a typical tagging efficiency of 30–50%, depending on p_{T}. The efficiencies in identifying matched top quarks and W bosons as a function of the p_{T} of the generated top quarks and W bosons are shown in Figure 4.1. The corresponding misidentification rate for jets initiated by gluons or light quarks, determined from sim-

ulated Z+jets events where the Z boson decays to neutrinos, are shown in Figure 4.2. The typical misidentification rate ranges from 1 to 4% (2 to 10%) for the top quark (W boson) tagging BDT.

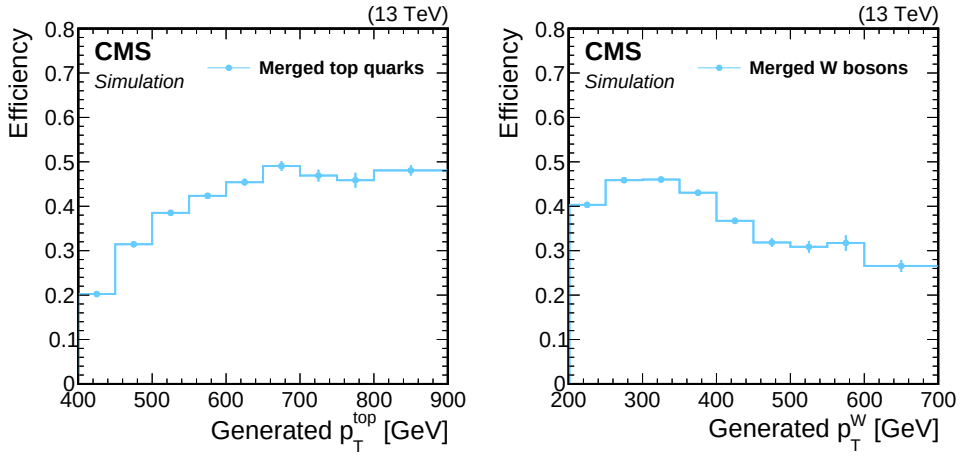


Figure 4.1: Efficiencies in simulation for identifying the hadronic decays of top quarks (left), and W bosons (right), as a function of the p_T of the generated top quarks or W bosons to which the jets are matched. Figures from [46].

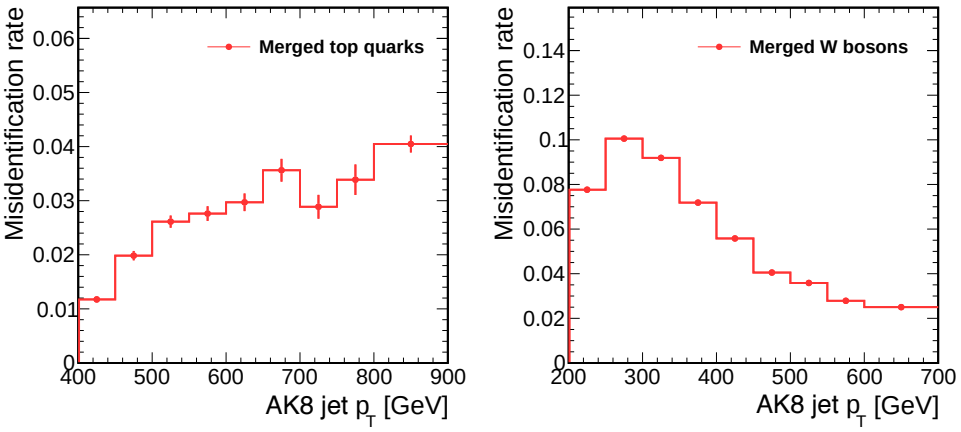


Figure 4.2: Misidentification rate for jets initiated by gluons or light quarks for the top quark (left) and W boson (right) tagging BDTs, as a function of the p_T of the AK8 jets.

Chapter 5

DeepAK8: DNN-Based Identification Algorithms Using Particle-Level Inputs

In CMS, jets are reconstructed by clustering the PF candidates with a jet finding algorithm. The PF reconstruction combines measurements from all subdetectors, therefore providing an excellent description with a rich set of properties for each individual particle. The most basic properties are the energy and momentum of each particle, which are used in all jet substructure observables. The PF reconstruction also identifies the type of each particle candidate, i.e., electron, muon, photon, charged hadron, or neutral hadron. This provides additional information for each jet that can be exploited. In addition, for charged particles, the trajectories are reconstructed by the tracker, therefore an even larger set of properties are available, such as the association to the reconstructed vertices, displacement from the primary vertex, the quality of the reconstructed track, a more precise momentum and energy in most cases, etc. These properties are extensively used in the identification of heavy-flavor jets, but have not been used so widely for tag-

ging boosted massive particles (top quarks and W, Z, and Higgs bosons), though they are expected to provide significant help here too, especially in the case in which one or more b or c quarks are present in the decays of these boosted particles.

Traditional tagging algorithms typically rely on a few jet-level observables constructed from the constituent particles, or additionally a small subset of properties of a limited number of selected constituent particles. An example is the top quark and W boson tagging BDTs described in Chapter 4, which use only $\mathcal{O}(10)$ jet-level observables and no particle-level properties. To achieve as high performance as possible for tagging boosted massive particles, all properties of all constituent particles of a jet should be exploited. However, this leads to a very high-dimensional input, with $\mathcal{O}(1000)$ variables, for each jet. Moreover, the correlation between these variables is likely to be rather complicated. On the other hand, this is particularly suitable for ML, especially using a DNN, which is capable of handling high-dimensional inputs and complex correlations. Motivated by this, the DeepAK8 algorithm has been developed, featuring the use of a DNN to directly process the constituents of a jet for the identification of boosted massive particles. AK8 jets reconstructed from PUPPI-weighted [43] PF candidates are used as the basis for this algorithm. The DeepAK8 algorithm shows significant improvement in performance compared to traditional multivariate approaches such as the BDT-based algorithm in Chapter 4.

5.1 Jet Labelling

The DeepAK8 algorithm is designed to be a multi-class tagging algorithm that uses a single DNN to separate jets originating from top quarks, W, Z and Higgs bosons, or from QCD radiations. The multi-class approach has significant advantages over the binary approach in simplicity and flexibility, as the training needs to be performed only

once for all classes, and the output scores can actually be aggregated and transformed for different tagging purposes. When proper output scores are used, the multi-class approach typically shows no loss of performance compared to a dedicated binary tagger.

Well-defined truth labels are of vital importance in the development of any ML classifiers, especially a multi-class classifier. To design a comprehensive algorithm for the identification of boosted massive particles, we implement a set of fine-grained labels for the AK8 jets. Five main categories are first defined, depending on if an AK8 jet is matched to the hadronic decay of a top quark, W, Z or Higgs boson, or not matched to any of them (labeled as "QCD" jets). The matching requires that all quarks from the decay of the massive particle be contained in the AK8 jet with $\Delta R(\mathrm{AK8}, \mathrm{q}) < 0.8$. In the case of top quarks, two scenarios are considered. The fully merged scenario requires that all three quarks be contained in the jet, while the partially merged scenario requires only the b-quark and one of the quarks from the W boson to be contained in the jet. Within each category, jets are further divided based on the decay modes of the matched particle (e.g., $\mathrm{Z} \to \mathrm{b\overline{b}}$, $\mathrm{Z} \to \mathrm{c\overline{c}}$ and $\mathrm{Z} \to \mathrm{q\overline{q}}$ in the Z boson category), or based on the number of b-hadrons or c-hadrons associated to a jet in the case of an un-matched QCD jet.

The full list of 17 classes is summarized in Table 5.1. The 17 classes are mutually exclusive to each other by construction. In the case where a jet can be matched to more than one truth particle (e.g., a top quark and the W boson from its decay), priority is always given to the Higgs boson first, followed by the top quark, and then the W and Z bosons, as shown in the table. The orthogonality of the labels allows for easy aggregation of different classes to target the needs of different analyses, therefore making DeepAK8 a highly versatile algorithm.

Category	Class
Higgs	H(bb)
	H(cc)
	H($WW^*/ZZ^* \to qqqq$)
Top (fully merged)	Top(bcq)
	Top(bqq)
Top (partially merged)	Top(bc)
	Top(bq)
W	W(cq)
	W(qq)
Z	Z(bb)
	Z(cc)
	Z(qq)
QCD	QCD(bb)
	QCD(cc)
	QCD(b)
	QCD(c)
	QCD(others)

Table 5.1: Summary of the truth label definition for the DeepAK8 algorithm.

5.2 Simulated Samples

The DeepAK8 algorithm is developed using simulated samples. The signal jets, i.e., those matched to hadronically decaying top quarks or W, Z, or Higgs bosons, are selected from simulations of BSM processes that are enriched in boosted massive particles. The signal jets are obtained from either heavy spin-1 Z' resonances or spin-2 graviton resonances, with subsequent decays to a pair of top quarks or a pair of W, Z, or Higgs bosons. These Z' or Graviton resonances are generated with masses ranging from 500 to 5000 GeV to produce sufficient signal jets over a wide range of p_T. They also have narrow intrinsic widths, equal to 1% of the resonance mass. The Z' and Graviton samples are generated with MadGraph [81] and interfaced with pythia 8.212 [82] for parton showering and hadronization. The leading-order (LO) NNPDF3.0 [83] parton distribution function

(PDF) set is used to produce these samples. For the background jets, a simulated QCD multijet production sample generated with PYTHIA 8.212 using the LO NNPDF2.3 [84] PDF set is used to acquire a large number of jets originating from gluons or light quarks. For all these samples, the CUETP8M1 tune [85] for the underlying-event modeling of PYTHIA is used. The response of the CMS detector is simulated with GEANT4 [86]. Additional pp interactions in the same or neighboring bunch crossings (referred to as "pileup") are simulated and used to overlay the hard-scattering process.

5.3 Input Variables

The DeepAK8 algorithm features the direct use of jet constituent particles as inputs. Basic kinematic properties of each particle, such as the momentum, the energy deposit, the angular separation between the particle and the jet axis or the subjet axes, etc., are included to help the algorithm learn about the radiation pattern of a jet. Additionally, the electric charge, the particle identification information, and the PUPPI weight of each particle are also included. For charged particles, a large set of extra properties measured by the tracking detector are included, such as the impact parameters and the covariance matrix of the track, the quality flags of the trajectory fit, etc. These inputs are particularly useful for the algorithm to identify the existence of heavy-flavor (b or c) quarks. In total, 42 variables are included for each constituent particle. The full list of the input variables, along with their definitions, is shown in Table 5.2.

To further strengthen the capability of identifying heavy-flavor quarks, properties of the secondary vertices reconstructed with the inclusive vertex finding algorithm [87, 79] are added to the inputs of the DeepAK8 algorithm. Only secondary vertices contained in the AK8 jet, i.e., $\Delta R(\mathrm{SV}, \mathrm{AK8}) < 0.8$, are considered. A total of 15 variables are included for each SV, such as the kinematics, the impact parameters, and the quality of

Variable	Definition
For both charged and neutral particles.	
$\log p_{\mathrm{T}}$	logarithm of the particle's p_{T}
$\log E$	logarithm of the particle's energy
$\log(p_{\mathrm{T}}/p_{\mathrm{T}}(\mathrm{jet}))$	logarithm of the particle's p_{T} relative to the jet p_{T}
$\lvert\eta\rvert$	absolute value of the particle's pseudorapidity
$\Delta\eta(\mathrm{jet})$	difference in pseudorapidity between the particle and the jet axis
$\Delta\phi(\mathrm{jet})$	difference in azimuthal angle between the particle and the jet axis
$\Delta R(\mathrm{jet})$	angular separation between the particle and the jet axis
$\Delta R(\mathrm{subjet\ 1})$	angular separation between the particle and the subjet leading in p_{T}
$\Delta R(\mathrm{subjet\ 2})$	angular separation between the particle and the subjet subleading in p_{T}
$\min \Delta R(\mathrm{SV})$	angular separation between the particle and closest secondary vertex
w_{PUPPI}	PUPPI weight of the particle
q	electric charge of the particle
`isMuon`	if the particle is identified as a muon
`isElectron`	if the particle is identified as an electron
`isPhoton`	if the particle is identified as a photon
`isChargedHadron`	if the particle is identified as a charged hadron
`isNeutralHadron`	if the particle is identified as a neutral hadron
f_{HCAL}	fraction of energy deposited in HCAL
For charged particles only. A default value of 0 is assigned for neural particle.	
`pvAssociationQuality`	flag related to the association of the track to the primary vertices
`lostInnerHits`	quality flag of the track related to missing hits on the pixel layers
d_{xy}	transverse impact parameter of the track
d_z	longitudinal impact parameter of the track
$d_{xy}/\sigma_{d_{xy}}$	significance of the transverse impact parameter
d_z/σ_{d_z}	significance of the longitudinal impact parameter
χ^2/dof	χ^2 value of the trajectory fit normalized to the number of degrees of freedom
`qualityMask`	quality flag of the track
$\mathrm{cov}(q/p, q/p)$	variance of the track parameter q/p
$\mathrm{cov}(\lambda, \lambda)$	variance of the track parameter λ
$\mathrm{cov}(\phi, \phi)$	variance of the track parameter ϕ
$\mathrm{cov}(d_{xy}, d_{xy})$	variance of the track parameter d_{xy}
$\mathrm{cov}(d_z, d_z)$	variance of the track parameter d_z
$\mathrm{cov}(d_{xy}, d_z)$	covariance of the track parameter d_{xy} and d_z
$\mathrm{cov}(\phi, d_{xy})$	covariance of the track parameter ϕ and d_{xy}
$\mathrm{cov}(\lambda, d_z)$	covariance of the track parameter λ and d_z
η_{rel}	pseudorapidity of the track relative to the jet axis
$p_{\mathrm{T,rel}}$ ratio	track momentum perpendicular to the jet axis, divided by the magnitude of the track momentum
$p_{\mathrm{par,rel}}$ ratio	track momentum parallel to the jet axis divided by the magnitude of the track momentum
d_{2D}	signed 2D impact parameter (i.e., in the transverse plane) of the track
$d_{\mathrm{2D}}/\sigma_{\mathrm{2D}}$	signed 2D impact parameter significance of the track
d_{3D}	signed 3D impact parameter of the track
$d_{\mathrm{3D}}/\sigma_{\mathrm{3D}}$	signed 3D impact parameter significance of the track
`trackDistance`	distance between the track and the jet axis at their point of closest approach

Table 5.2: Input variables of each jet constituent particle.

Variable	Definition
$\log p_{\mathrm{T}}$	logarithm of the SV's p_{T}
$\log E$	logarithm of the SV's energy
$\log(p_{\mathrm{T}}/p_{\mathrm{T}}(\mathrm{jet}))$	logarithm of the SV's p_{T} relative to the jet p_{T}
$\lvert\eta\rvert$	absolute value of the SV's pseudorapidity
$\Delta\eta(\mathrm{jet})$	difference in pseudorapidity between the SV and the jet axis
$\Delta\phi(\mathrm{jet})$	difference in azimuthal angle between the SV and the jet axis
$\Delta R(\mathrm{jet})$	angular separation between the SV and the jet axis
m_{SV}	mass of the SV
N_{tracks}	number of tracks associated with the SV
χ^2/dof	χ^2 value of the SV fit normalized to the number of degrees of freedom
d_{2D}	signed 2D impact parameter (i.e., in the transverse plane) of the SV
$d_{\mathrm{2D}}/\sigma_{\mathrm{2D}}$	signed 2D impact parameter significance of the SV
d_{3D}	signed 3D impact parameter of the SV
$d_{\mathrm{3D}}/\sigma_{\mathrm{3D}}$	signed 3D impact parameter significance of the SV
$\cos(\vec{p}_{\mathrm{SV}}, \overrightarrow{(\mathrm{PV},\mathrm{SV})})$	cosine of the angle between the SV momentum and the vector pointing from the primary vertex to the SV

Table 5.3: Input variables for each secondary vertex (SV) inside the jet.

the SV reconstruction, etc. The full list of input variables for the secondary vertices is shown in Table 5.3.

5.4 DNN Training

The DeepAK8 algorithm uses a DNN classifier to process the high-dimensional constituent particle and secondary vertex inputs for the identification of boosted massive particles. The development of such a DNN, including the preprocessing of the inputs and the training samples, the architecture of the DNN, the training procedure, etc., is presented in this section.

5.4.1 Preprocessing

Sample reweighting

As described in Section 5.2, signal jets for the training of the DNN are selected from heavy resonance samples, while background jets are selected from a QCD multijet sample. Therefore, signal and background jets have very different distributions in p_{T}.

To avoid the DNN exploiting such artificial differences in p_{T} as a handle to separate signal and background jets, the jets in the training sample are "reweighted" to have a flat distribution in p_{T}, and the contributions from each category (top quark, W, Z, Higgs boson, and QCD jets) are equalized. This eliminates any difference in the p_{T} distribution seen by the DNN during the training, therefore preventing the DNN responses from showing any bias on p_{T}.

Such a "reweighting" can be implemented in a number of ways. The most straightforward way is to assign a weight to each jet and take the weight into account when computing the loss function during the training. However, from our studies, we found that this approach could lead to instability or even failure in convergence, especially when the weights vary over a very large range (e.g., orders of magnitude). Alternatively, events can be removed from the training sample to achieve the reweighting, but at the cost of a potentially significant reduction on the size of the training dataset, therefore limiting the effectiveness of the training. As a result, we implement the reweighting in a different way. It is performed "on-the-fly" by randomly sampling the training dataset to form each mini-batch for the training, with the probability for each jet to be selected being proportional to the weight. This approach is more efficient than removing events completely, as in principle, all events can still be utilized, therefore preserving the diversity of the training sample. Meanwhile, the training does not suffer from any instability as events are "unweighted" in each mini-batch.

Input transformation

A common practice in the training of DNNs is to "standardize" the inputs such that all inputs fall in similar ranges, e.g., in $[0, 1]$, after the transformation. This makes the training more stable and can potentially improve performance. Such standardization is particularly important for DeepAK8, as different input variables have different charac-

teristic scales, and they can range over several orders of magnitude for different variables. The most commonly used standardization method is to subtract the mean and then divided by the standard deviation for each variable. This works well if the distribution of the variable is close to a normal distribution. However, some of the variables used in DeepAK8 have long tails in the distribution and, in a few cases, extreme outliers. The mean and standard deviation are sensitive to tails and outliers, which can cause the bulk of a distribution to be squeezed into a very small range and become unusable. Thus, we use percentiles for standardization. Each input variable x is transformed as follows:

$$\tilde{x} = \frac{x - p_{50\%}}{p_{84\%} - p_{50\%}}, \tag{5.1}$$

where $p_{50\%}$ and $p_{84\%}$ are the 50th and 84th percentiles of the variable x. In the case when $p_{84\%} = p_{50\%}$ (which happens for some discrete variables), the denominator is simply taken to be 1, as those discrete variables are typically $\mathcal{O}(1)$. The use of percentiles instead of the mean and the standard deviation tends to be less sensitive to outliers and long tails of the distributions, leading to more unified scales for different variables. The transformed values are further clipped to be in the range of $[-5, 5]$ to prevent any extreme outliers from being fed into the neural networks and demolishing the training.

Jets typically have varying numbers of particles and secondary vertices associated with them. Figure 5.1 shows the distribution of the number of particles for AK8 jets of different categories. The number of particles in an AK8 jet ranges from around 20 to almost 100 depending on the category and p_{T} of the jet. The particle multiplicity is generally higher as the jet p_{T} increases, and top quark jets tend to have more constituent particles than other types of jets. Across all categories and p_{T} ranges, very few ($<\sim 6\%$) jets have more than 100 particles. The distribution of the number of secondary vertices associated with AK8 jets is shown in Figure 5.2. As expected, secondary vertex multiplicity is higher

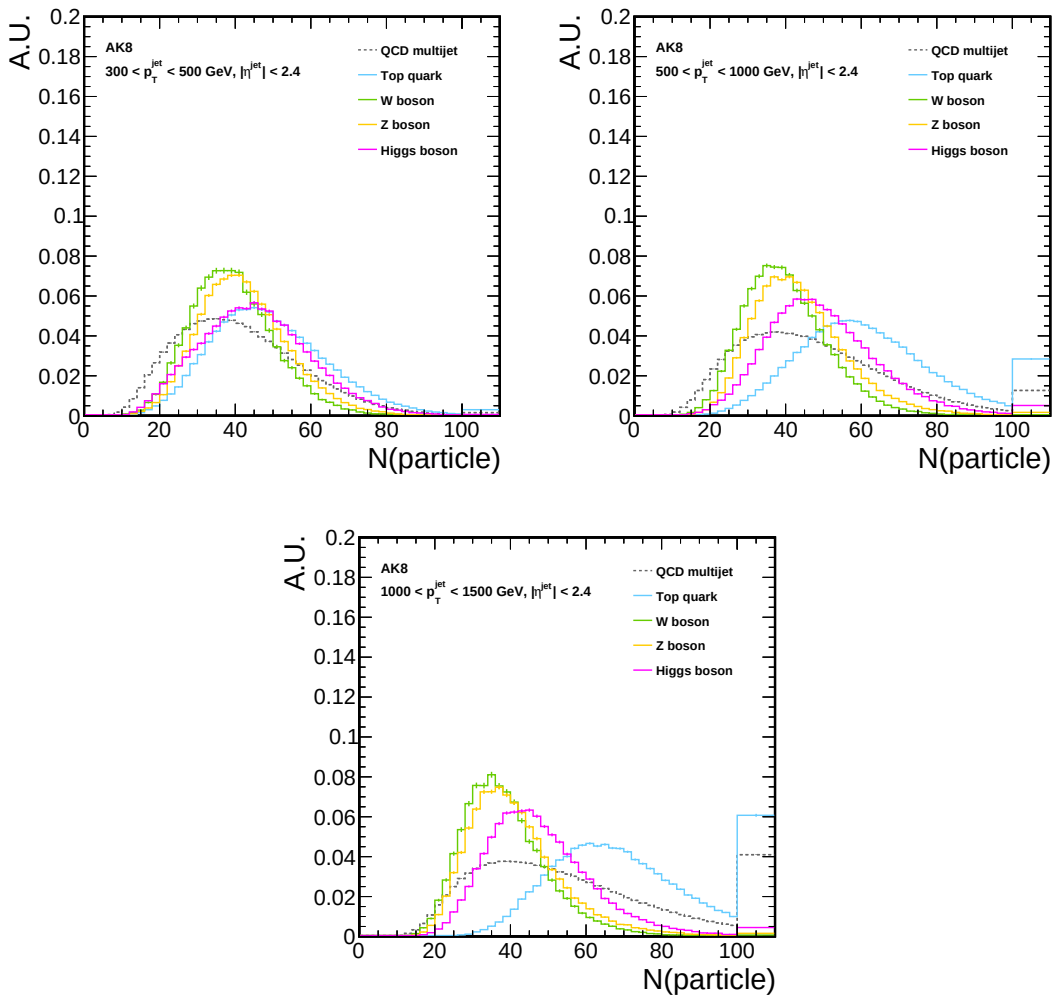


Figure 5.1: The distribution of the number of particles for AK8 jets of different categories in three p_{T} ranges. Top left: $300 < p_{\mathrm{T}} < 500\,\mathrm{GeV}$; Top right: $500 < p_{\mathrm{T}} < 1000\,\mathrm{GeV}$; Bottom: $1000 < p_{\mathrm{T}} < 1500\,\mathrm{GeV}$.

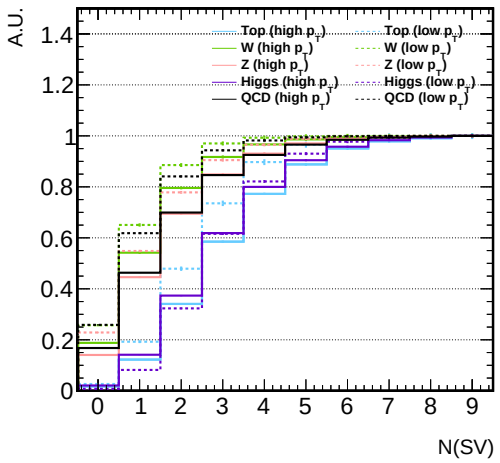


Figure 5.2: The integrated distribution of the number of secondary vertices associated with AK8 jets of different categories in two p_{T} ranges (low p_{T}: $400 < p_{\mathrm{T}} < 500\,\mathrm{GeV}$; high p_{T}: $1000 < p_{\mathrm{T}} < 1400\,\mathrm{GeV}$). The vertical axis shows the fraction of jets with no more than N secondary vertices, where N is the label of each bin.

for top quark and Higgs boson jets due to the presence of b quarks in their decays. Overall, less than $\sim 2\%$ of jets contain more than 7 secondary vertices. Based on these observations, only up to 100 particles and up to 7 secondary vertices are used as inputs to the DNN. The particles of each AK8 jet are sorted in descending p_{T} order, while the secondary vertices are ranked by their signed 2D impact parameter significance, also in descending order. Standardized input variables of the leading 100 particles and the leading 7 secondary vertices are assembled into two fixed-size arrays, respectively. In the case that a jet has less than 100 particles or less than 7 secondary vertices, zero values are used to fill the remaining entries of the two arrays. These two arrays, referred to as the "particle" array and "SV" array hereafter, are the inputs to the DNN.

5.4.2 Network Architecture

A significant challenge posed by the direct use of particle-level information is a substantial increase in the number of inputs. Meanwhile, the correlations between these

inputs are of vital importance. Thus an algorithm that can both process the inputs efficiently and exploit the correlations effectively is needed. A customized DNN architecture was developed in DeepAK8 to fulfill this requirement. As illustrated in Figure 5.3, the architecture consists of two steps. In the first step, two one-dimensional convolutional neural networks (CNN) are applied to the particle array and the SV array separately to transform the inputs and extract useful features. Then, in the second step, the outputs of the two CNNs are combined and processed by a simple fully-connected network to perform the jet classification. The CNN structure in the first step is based on the ResNet model [88], but adapted from two-dimensional images to one-dimensional particle lists, in which each variable corresponds to a "color" channel. The CNN for the particle array has 14 layers[1], and the one for the SV array has 10 layers[2]. A convolution window of length 3 is used, and the number of output channels in each convolutional layer ranges between 32 and 128. The ResNet architecture allows for an efficient training of deep CNNs, thus leading to better exploitation of the correlations between the large inputs and improving the performance. With the CNNs in the first step being already very powerful, the fully-connected network in the second step can be fairly simple: it consists of only one layer with 512 units, followed by a ReLU [89] activation function and a Dropout [90] layer of 20% drop rate. The detailed configuration of the network is illustrated in Figure 5.4(a). The neural network is implemented using the MXNet package [91] and trained with the Adam [75] optimizer to minimize the cross-entropy loss. A mini-batch size of 1024 is used. The initial learning rate is set to 0.001 and then reduced by a factor of 10 at the 10th and 20th epochs to improve convergence. The training is stopped after 35 epochs. A sample of 50 million jets is used, of which 80% are used for training, and 20% are used for development and validation.

[1]Including 13 convolutional layers and 1 global pooling layer.
[2]Including 9 convolutional layers and 1 global pooling layer.

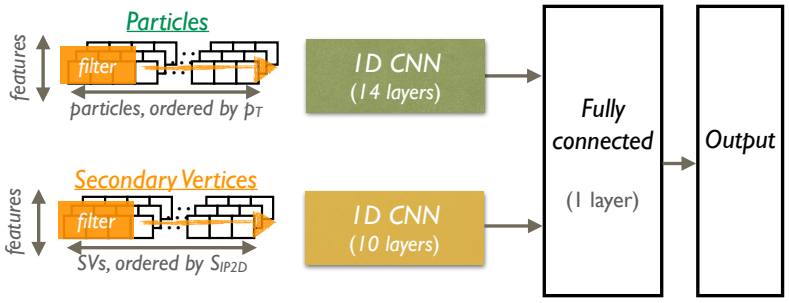


Figure 5.3: A simplified schematic of the DeepAK8 network.

5.4.3 Decorrelation with the Jet Mass

As will be discussed in Section 5.5.3, background jets selected by the DeepAK8 algorithm exhibit a modified mass distribution which is similar to that of the signal. This is because the mass of a jet is one of the most discriminating variables, and although it is not directly used as an input to the algorithm, the CNNs can extract features that are correlated to the mass to improve the discrimination power. However, such modification of the mass distribution may be undesirable if the mass variable itself is to be used for separating signal and background processes. Thus, an alternative DeepAK8 algorithm, "DeepAK8-MD", is developed to be largely decorrelated with the mass of a jet while preserving the discrimination power as much as possible. The adversarial training approach [92] is used to achieve this goal. In addition, jets from various signal and background samples are reweighted to yield flat distributions in both p_{T} and m_{SD} to aid the training.

The training procedure of DeepAK8-MD is illustrated in Figure 5.5, and the detailed configuration of the network is shown in Figure 5.4(b). Compared to the nominal version of DeepAK8, a mass prediction network is added with the goal of predicting the mass of a jet from the features extracted by the CNNs. When properly trained, the mass prediction network becomes a good indicator of how strongly the features extracted by the CNNs are correlated with the mass of a jet, as the stronger the correlation is, the

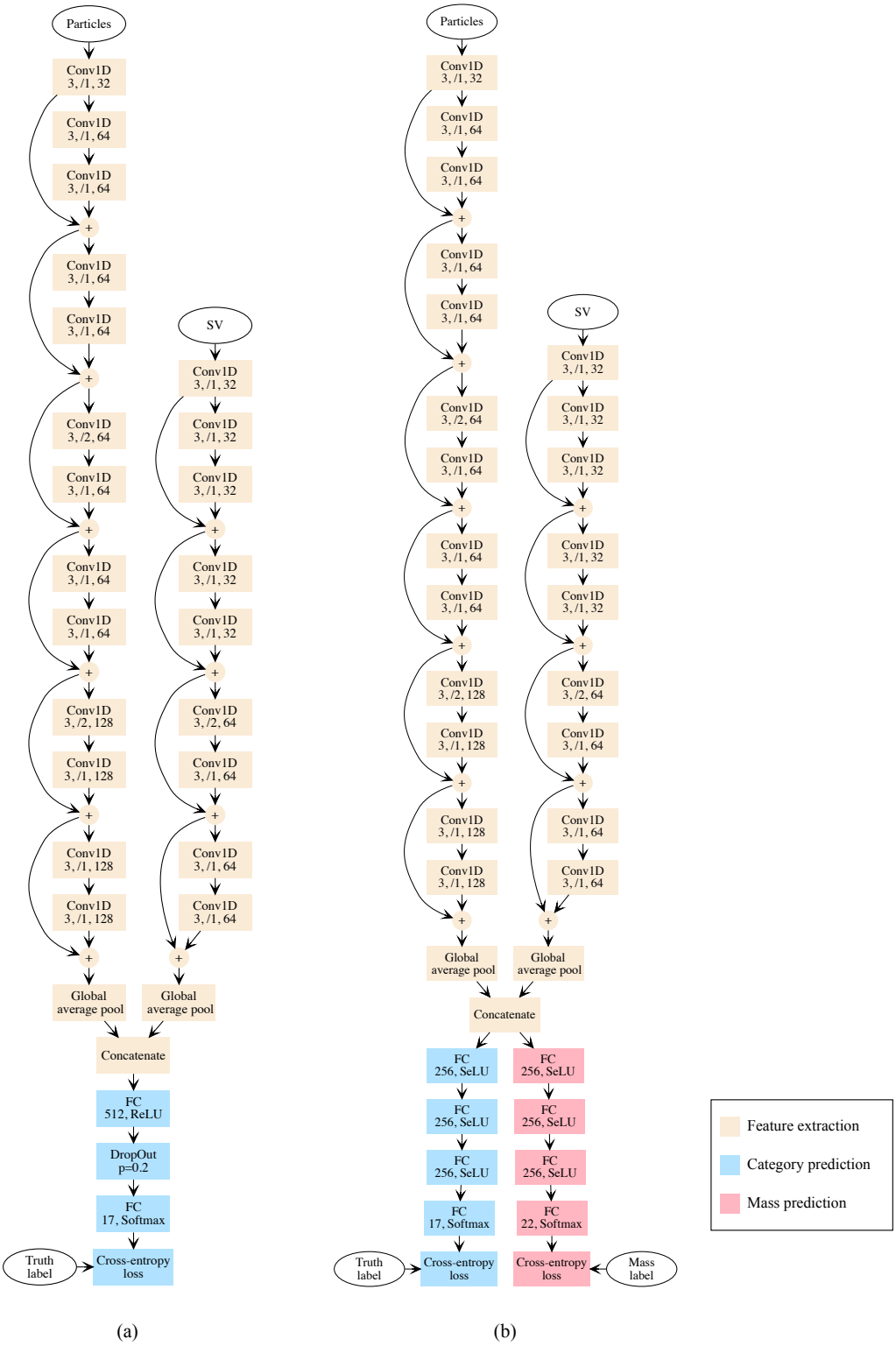


Figure 5.4: The network architecture of (a) DeepAK8 and (b) DeepAK8-MD.

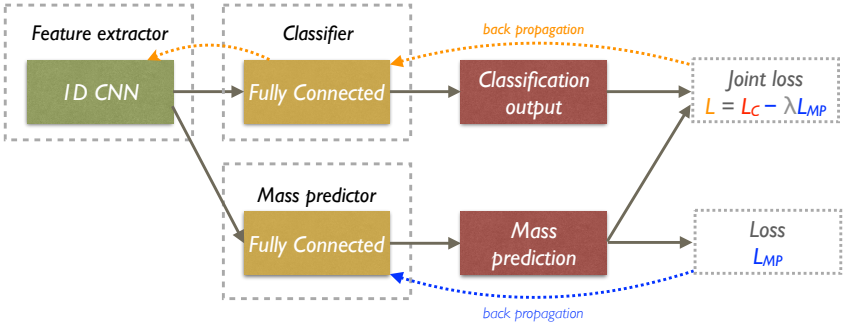


Figure 5.5: A simplified schematic of the training procedure of the DeepAK8-MD network.

more accurate the mass prediction will be. With the introduction of the mass prediction network, the training target of the algorithm can be modified to include the accuracy of the mass prediction as a penalty, therefore preventing the CNNs from extracting features that are correlated with the mass. In this way, the final prediction of the algorithm also becomes largely independent of the mass. As the features extracted by the CNNs evolve during the training process, the mass prediction network itself needs to be updated regularly to adapt to the changes of its inputs and remain as an effective indicator of mass correlation. Forcing the algorithm to be decorrelated with the jet mass inevitably leads to a loss of discrimination power, and the resultant algorithm is a balance between performance and mass-independence. As the training of DeepAK8-MD is carried out only on jets with $30 < m_{\mathrm{SD}} < 250\,\mathrm{GeV}$, jets with m_{SD} outside this range should be removed when using DeepAK8-MD. Typically, to achieve the best performance, a suitable mass window requirement should be applied together with the DeepAK8-MD algorithm.

5.5 Performance in Simulation

The performance of the DeepAK8 algorithm is first studied in simulation, using the signal and background samples described in Section 5.2. Only events not used in the

training are considered when evaluating the performance. To avoid any bias from the difference in the p_{T} distribution of different samples, events in the signal samples are reweighted to have the same distribution in p_{T} as the background sample.

5.5.1 Signal and Background Efficiency

The performance of the algorithms is evaluated using the signal and background efficiency, ϵ_S and ϵ_B, respectively, as a figure of merit. The ϵ_S and ϵ_B are defined as:

$$\epsilon_S = \frac{N_S^{\mathrm{tagged}}}{N_S^{\mathrm{total}}} \quad \text{and} \quad \epsilon_B = \frac{N_B^{\mathrm{tagged}}}{N_B^{\mathrm{total}}}, \tag{5.2}$$

where $N_{S(B)}^{\mathrm{tagged}}$ is the number of signal (background) jets satisfying the identification criteria of each algorithm, and $N_{S(B)}^{\mathrm{total}}$ is the total number of generated particles considered to be signal (background). Hadronically decaying top quarks and W, Z, and Higgs bosons are considered as the signal, while light quarks and gluons from the QCD multijet process are considered to be the background. For the Higgs boson, only the $\mathrm{H} \to \mathrm{b\overline{b}}$ decay is considered.

The discrimination power, i.e., ϵ_B as a function of ϵ_S, is evaluated in terms of receiver operating characteristic (ROC) curves. Figures 5.6–5.9 show the ROC curves of DeepAK8 and DeepAK8-MD for the identification of top quarks and W, Z, and Higgs bosons, respectively. Baseline algorithms widely used in CMS are also included for comparison. The baseline algorithm for top quark (W, Z boson) tagging is based on the soft drop mass m_{SD} and the N-subjettiness ratio τ_{32} (τ_{21}), described in Chapter 3. For top quark tagging, both versions with and without the subjet b-tagging requirement are considered, and the version requiring at least one of the two subjets to be b-tagged is labeled with $m_{\mathrm{SD}} + \tau_{32} + \mathrm{b}$. For Higgs bosons tagging, the double-b algorithm [79] designed to specifically target Higgs decays to pairs of b quarks in the boosted regime

is used as a baseline. The comparisons are performed at low and high values of truth particle p_{T}. The fiducial selection criteria applied to the truth particles are displayed on the plots. For all algorithms except the nominal DeepAK8, a selection on the soft drop mass is always applied and indicated on the plots, while for the nominal DeepAK8, the mass information is learned by the algorithm based on the particle kinematics inputs and no explicit cut is needed. On the contrary, allowing the DNN to apply the mass selection implicitly results in better performance than imposing a mass window by hand. Overall, both the nominal DeepAK8 and mass-decorrelated DeepAK8-MD algorithms achieve significant improvement in performance, lowering ϵ_B by an order of magnitude compared to the baseline algorithms in some cases.

To gain a deeper understanding of the large performance improvement in DeepAK8, alternative versions of DeepAK8 were trained using a subset of the input features. Three sets of input features were studied and compared. The "Particle (kinematics)" set consists of only the kinematic information of the PF candidates, e.g., the four momenta, the distances to the jet and subjet axes. This set serves as a baseline to evaluate the performance using only the substructure of the jets. The "Particle (w/o Flavour)" set includes additional experimental information of each PF candidate, such as the electric charge, particle identification, and track quality information. Compared to the nominal DeepAK8 algorithm, input features that contribute to the identification of heavy-flavor quarks, such as the displacement of the tracks, the association of tracks to the reconstructed vertices, as well as the SV features, are not included in the "Particle (w/o Flavour)" set. The performance of the three versions of DeepAK8 is compared in Figure 5.10 for top quark and Z boson identification. In both cases, the addition of experimental information brings a sizable improvement in performance. While the additional features contributing to heavy flavor identification yield no improvement for the identification of Z bosons decaying to a pair of light quarks, a significant improvement is observed for Z

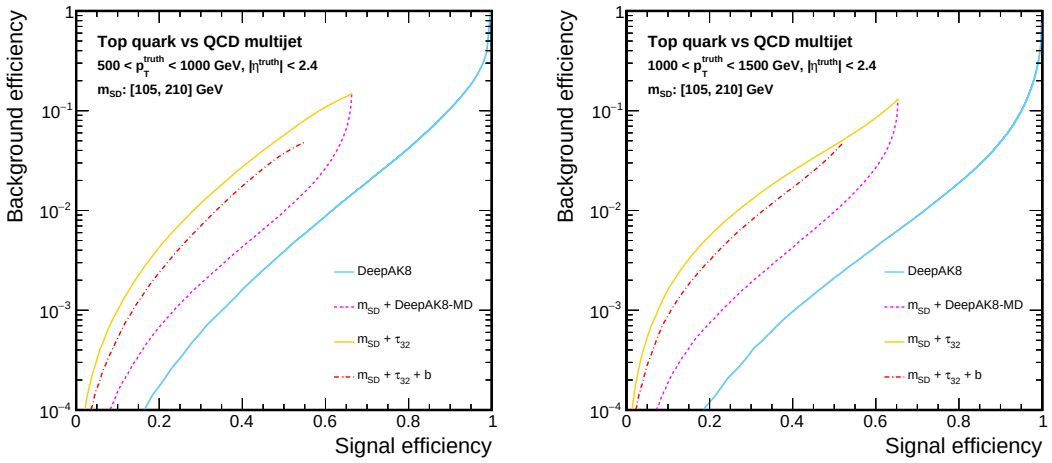


Figure 5.6: Performance comparison of the hadronically decaying top quark tagging algorithms in terms of ROC curves in two regions of the truth particle's p_{T}. Left: $500 < p_{\mathrm{T}} < 1000\,\mathrm{GeV}$; Right: $1000 < p_{\mathrm{T}} < 1500\,\mathrm{GeV}$. Additional fiducial selection criteria applied to the truth particles are displayed on the plots. For all algorithms except the nominal DeepAK8, a selection on the soft drop mass is applied and indicated on the plots.

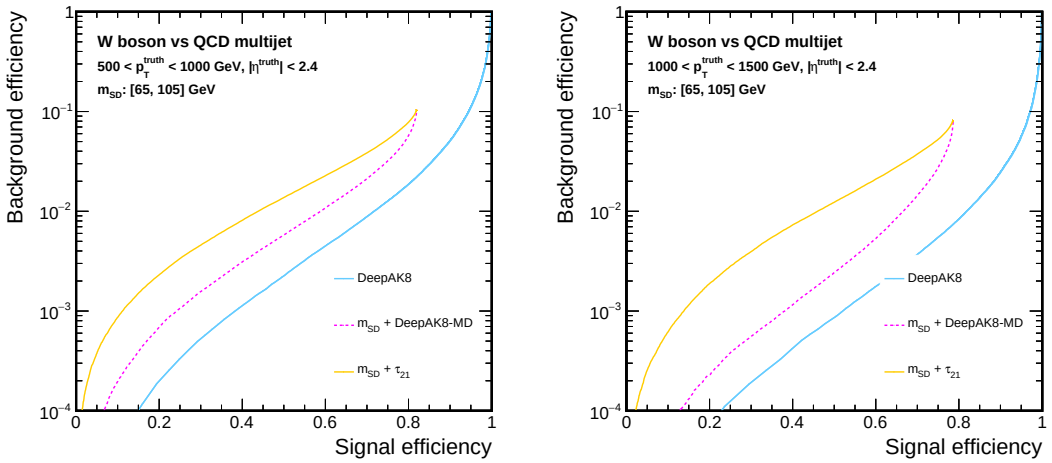


Figure 5.7: Performance comparison of the hadronically decaying W tagging algorithms in terms of ROC curves in two regions of the truth particle's p_{T}. Left: $500 < p_{\mathrm{T}} < 1000\,\mathrm{GeV}$; Right: $1000 < p_{\mathrm{T}} < 1500\,\mathrm{GeV}$. Additional fiducial selection criteria applied to the truth particles are displayed on the plots. For all algorithms except the nominal DeepAK8, a selection on the soft drop mass is applied and indicated on the plots.

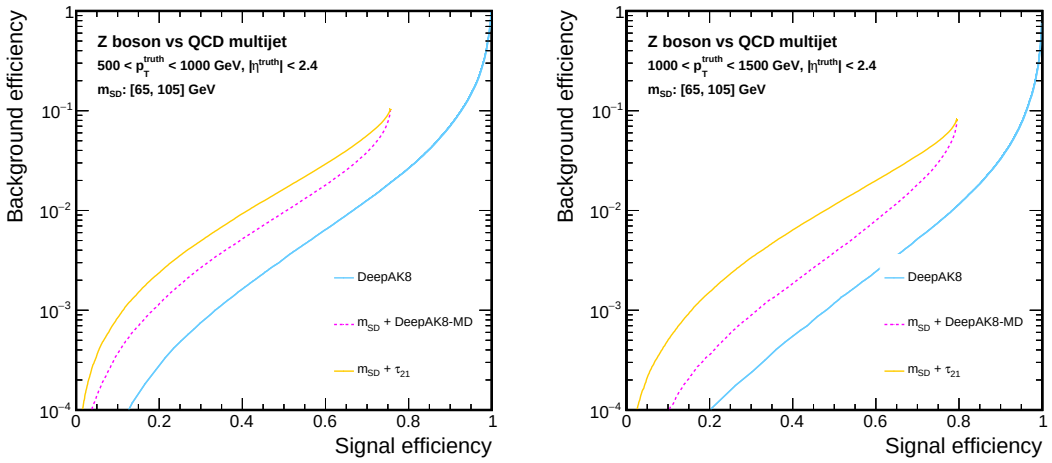


Figure 5.8: Performance comparison of the hadronically decaying Z boson tagging algorithms in terms of ROC curves in two regions of the truth particle's p_{T}. Left: $500 < p_{\mathrm{T}} < 1000\,\mathrm{GeV}$; Right: $1000 < p_{\mathrm{T}} < 1500\,\mathrm{GeV}$. Additional fiducial selection criteria applied to the truth particles are displayed on the plots. For all algorithms except the nominal DeepAK8, a selection on the soft drop mass is applied and indicated on the plots.

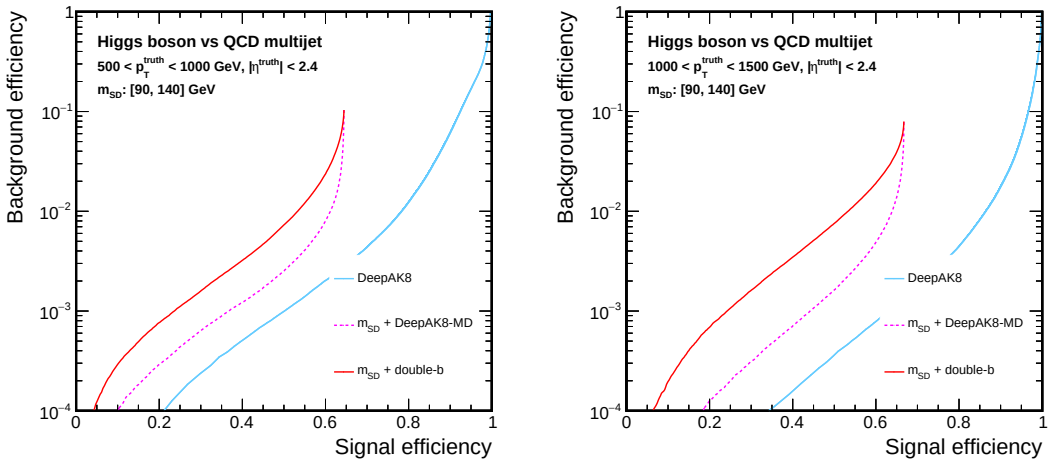


Figure 5.9: Performance comparison of the hadronically decaying Higgs boson tagging algorithms in terms of ROC curves in two regions of the truth particle's p_{T}. Left: $500 < p_{\mathrm{T}} < 1000\,\mathrm{GeV}$; Right: $1000 < p_{\mathrm{T}} < 1500\,\mathrm{GeV}$. Additional fiducial selection criteria applied to the truth particles are displayed on the plots. For all algorithms except the nominal DeepAK8, a selection on the soft drop mass is applied and indicated on the plots.

decaying to a pair of b quarks, as well as the identification of top quarks, demonstrating the strong complementarity between heavy flavor identification and jet substructure for heavy resonance identification when heavy flavor quarks are involved in the decay.

5.5.2 Robustness of Tagging Algorithms

An important aspect of a tagging algorithm is its stability with respect to jet kinematics (e.g., p_{T}) and data-taking conditions (e.g., pileup). To quantify this, we study the ϵ_S and ϵ_B of the algorithms as a function of the p_{T} of the truth particle and the number of reconstructed vertices ($\mathrm{N_{vtx}}$). For the sake of these studies, a common working point is defined, corresponding to $\epsilon_S = 30\%$ (50%) for top quarks (W, Z, and Higgs bosons) with $500 < p_{\mathrm{T}}(\text{truth particle}) < 600\,\mathrm{GeV}$.

The distributions of the ϵ_S and ϵ_B as a function of the p_{T} of the truth particle for different tagging scenarios are displayed in Figures 5.11 and 5.12, respectively. Overall, the ϵ_S and ϵ_B of the DeepAK8 and DeepAK8-MD algorithms are fairly stable as a function of p_{T}. In the top quark tagging case, the ϵ_S for all the algorithms drops rapidly as p_{T} decreases below $p_{\mathrm{T}} \sim 600\,\mathrm{GeV}$, due to the fact that the decay products of a top quark are not fully merged to be contained in an AK8 jet at such low p_{T}. Similar behavior is observed for the top quark misidentification rate of the DeepAK8 and DeepAK8-MD algorithms.

The dependence on $\mathrm{N_{vtx}}$ is also examined using simulated events. Figure 5.13 (5.14) displays the distribution of ϵ_S (ϵ_B) as a function of $\mathrm{N_{vtx}}$ for truth particles with $500 < p_{\mathrm{T}}(\text{truth particle}) < 1000\,\mathrm{GeV}$, operating at a working point with $\epsilon_S = 30\%$ (50%) for top quark (W, Z, and Higgs boson) identification as defined above. Since all the algorithms make use of jets that exploit PUPPI for pileup mitigation, the resulting algorithms exhibit pretty good stability, showing a roughly constant ϵ_S and ϵ_B distribution across

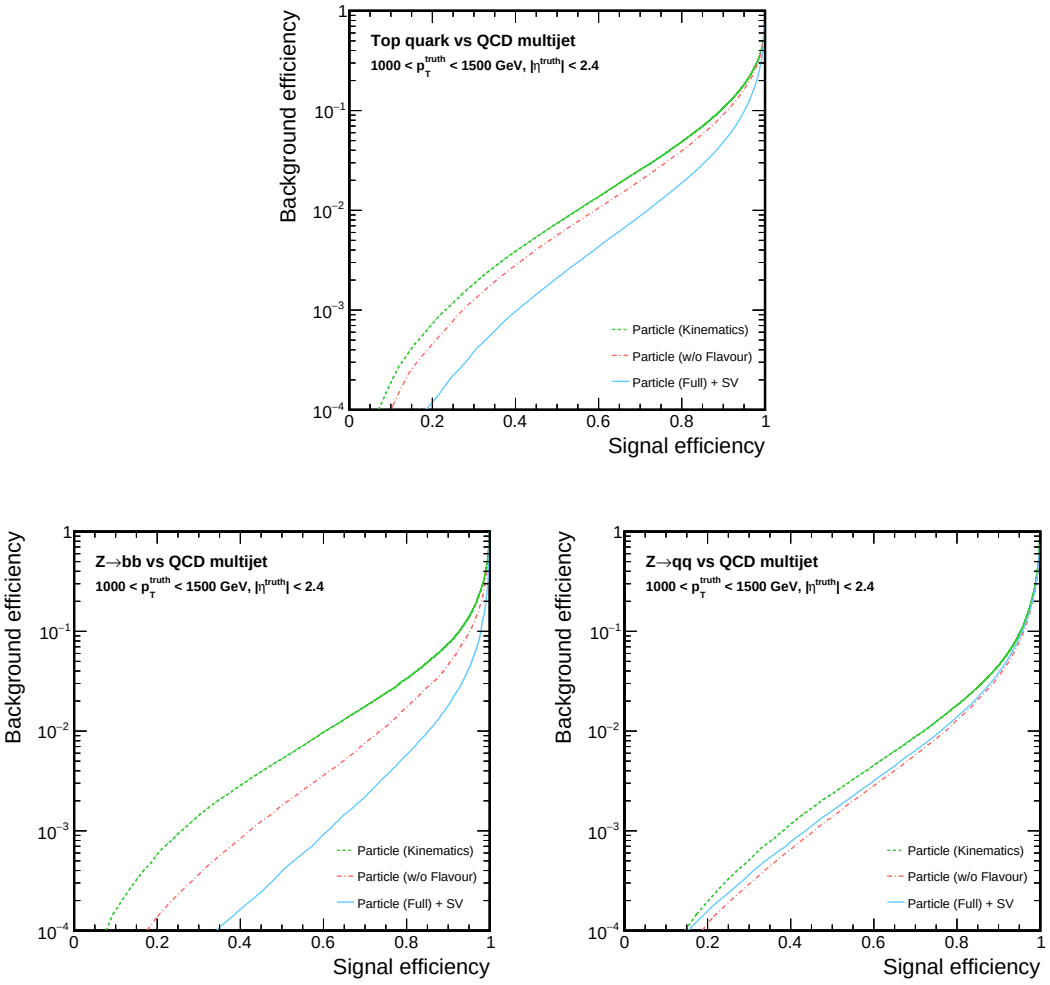


Figure 5.10: Alternative versions of DeepAK8 trained using a subset of the input features. The details about each version are discussed in the text. The performances of the three versions of DeepAK8 are compared for top quark (upper) and Z boson (lower) identification. For the latter, the left plot corresponds Z bosons decaying to a pair of b quarks, and the right to a pair of light quarks (excluding b and c quarks).

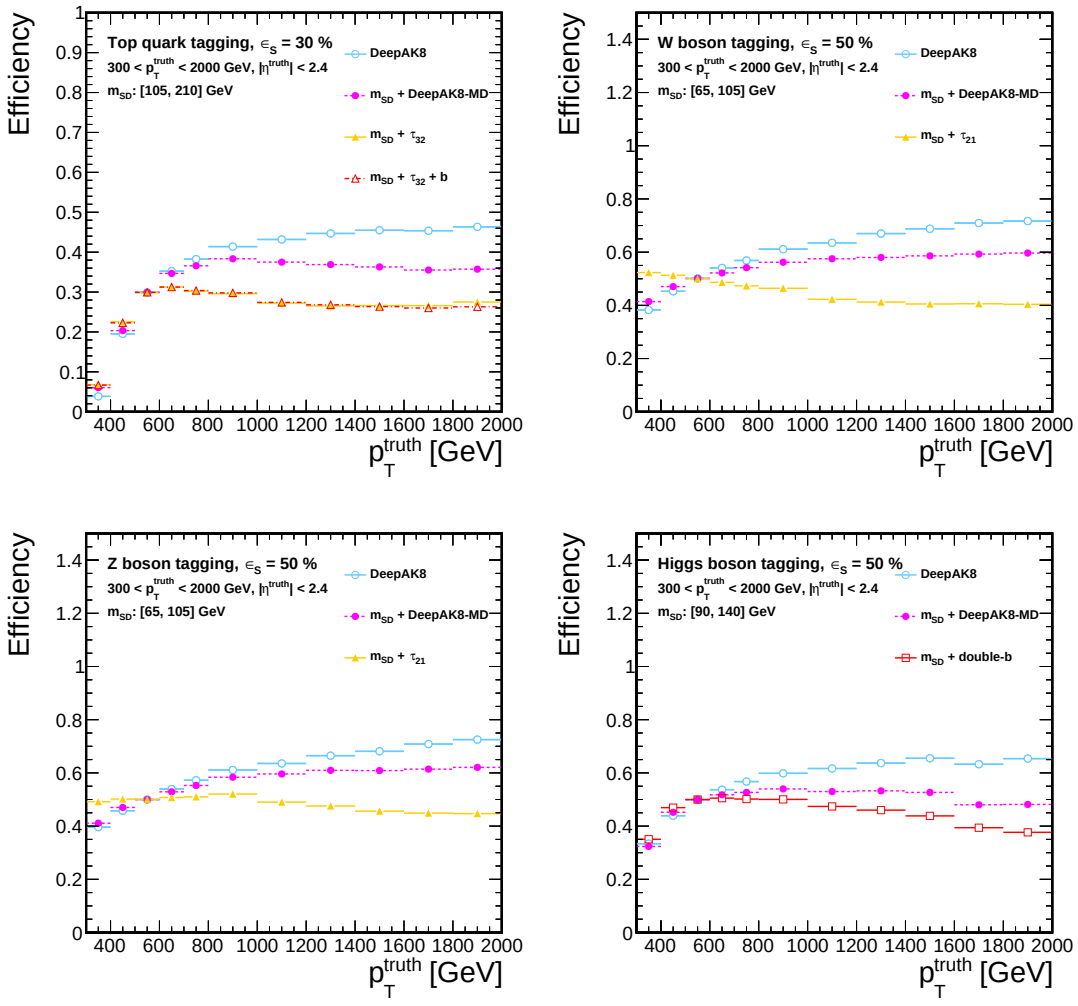


Figure 5.11: The distribution of ϵ_S as a function of the p_{T} of the truth particle for a working point corresponding to $\epsilon_S = 30\%$ (50%) for top quark (W, Z, and Higgs boson) identification. Upper left: top quark, upper right: W boson, lower left: Z boson, lower right: Higgs boson. Additional fiducial selection criteria applied to the truth particles are displayed on the plots. For all algorithms except the nominal DeepAK8, a selection on the soft drop mass is applied and indicated on the plots.

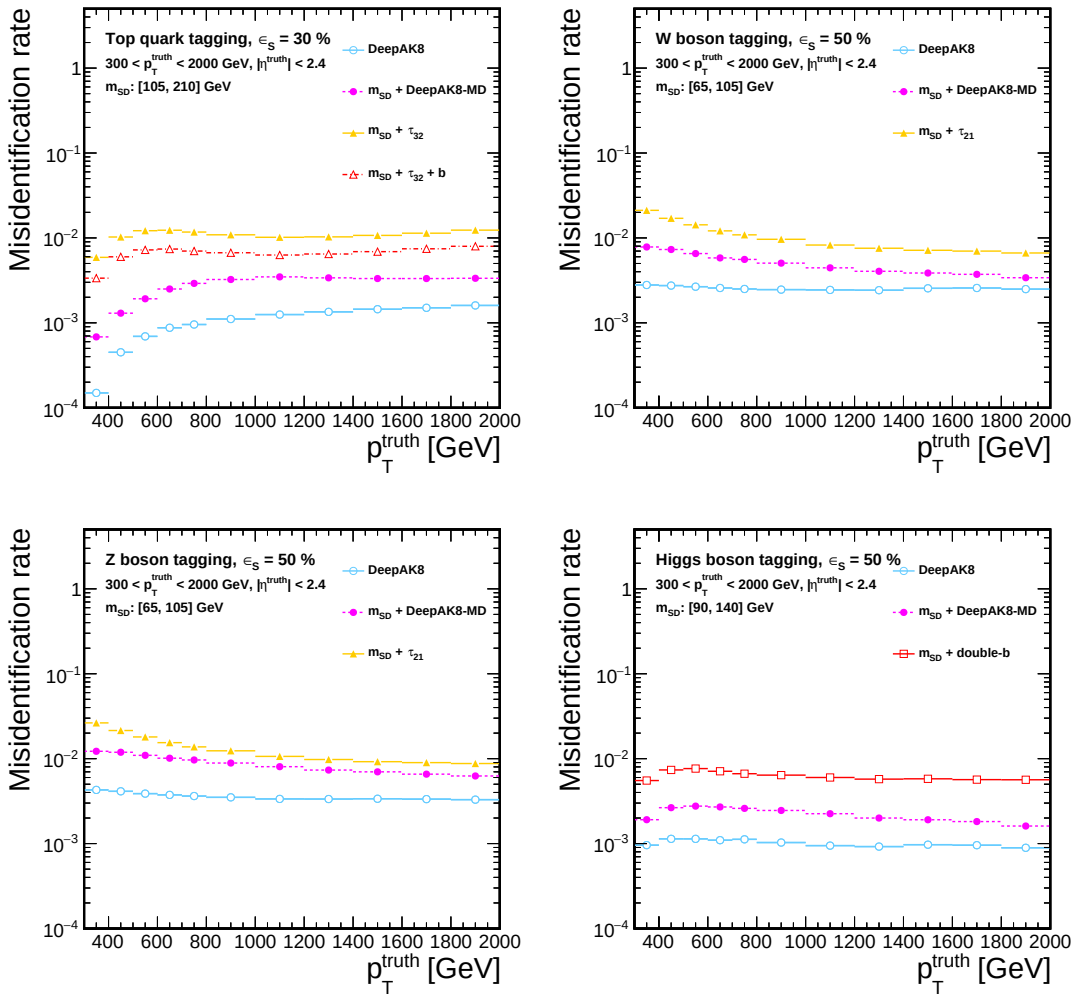


Figure 5.12: The distribution of ϵ_B as a function of the p_{T} of the truth particle for a working point corresponding to $\epsilon_S = 30\%$ (50%) for top quark (W, Z, and Higgs boson) identification. Upper left: top quark, upper right: W boson, lower left: Z boson, lower right: Higgs boson. Additional fiducial selection criteria applied to the truth particles are displayed on the plots. For all algorithms except the nominal DeepAK8, a selection on the soft drop mass is applied and indicated on the plots.

the different pileup scenarios.

5.5.3 Correlation with Jet Mass

Finally, a set of studies was performed to understand the correlation of the DeepAK8 algorithm with the jet mass. Figure 5.15 displays the m_{SD} distribution of jets obtained from the QCD multijet sample, inclusively and after applying a selection on each tagging algorithm. The working point chosen corresponds to $\epsilon_S = 30\%$ (50%) for top quark (W, Z, and Higgs boson) tagging, as defined in Section 5.5.1. The nominal version of DeepAK8 exhibits significant "mass sculpting", i.e., after applying a selection on the DeepAK8 tagging algorithm, the m_{SD} distribution of the background jets is modified significantly and resembles that of the signal jets, showing a peak around the mass of the signal particle. Such mass sculpting itself is not necessarily a problem, but it can limit the applicability of the DeepAK8 algorithm, as many analyses require a smoothly falling background jet mass spectrum under a signal peak (for instance, in Ref. [93]). On the other hand, DeepAK8-MD trained with dedicated mass-decorrelation techniques show very little modification of the background jet mass spectrum.

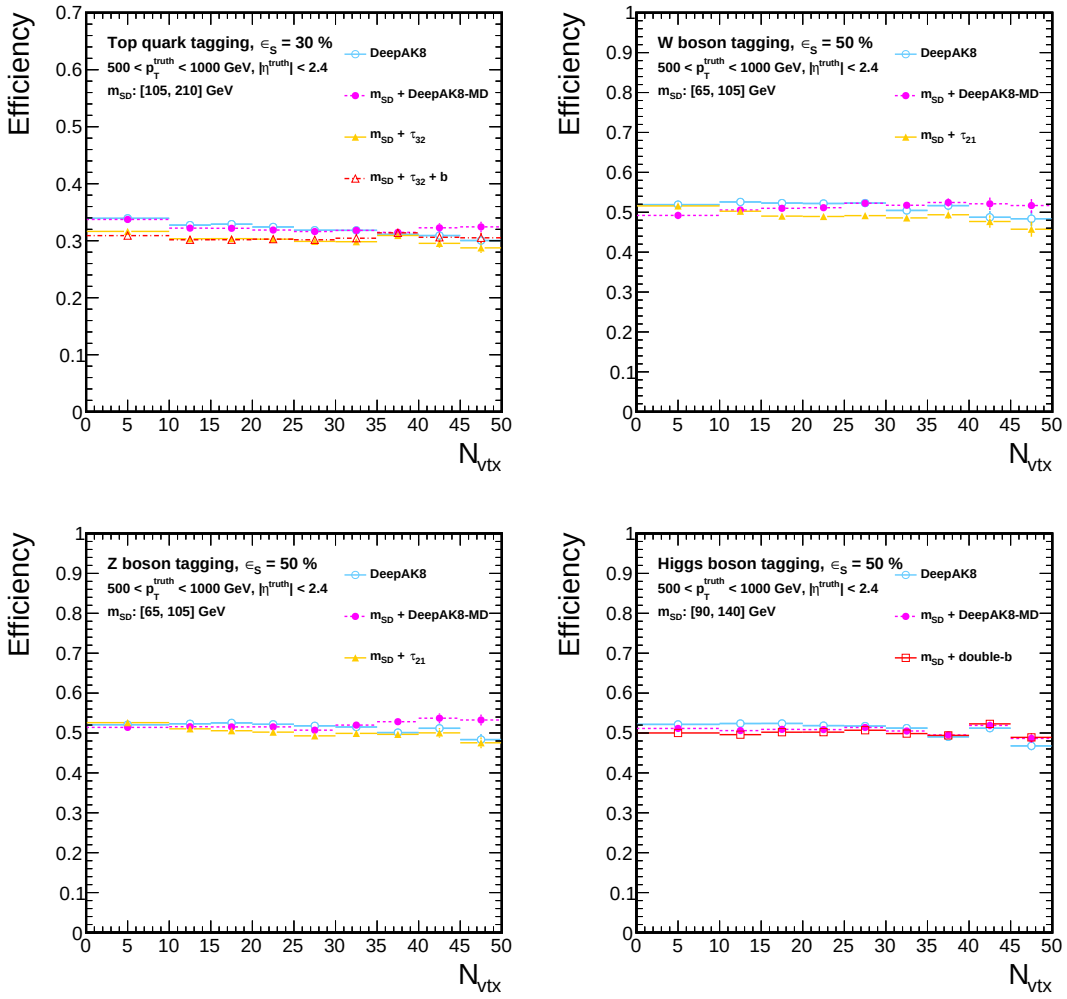


Figure 5.13: The distribution of ϵ_S as a function of N_{vtx} for a working point corresponding to $\epsilon_S = 30\%$ (50%) for top quark (W, Z, and Higgs boson) identification. Upper left: top quark, upper right: W boson, lower left: Z boson, lower right: Higgs boson. Additional fiducial selection criteria applied to the truth particles are displayed on the plots. For all algorithms except the nominal DeepAK8, a selection on the soft drop mass is applied and indicated on the plots.

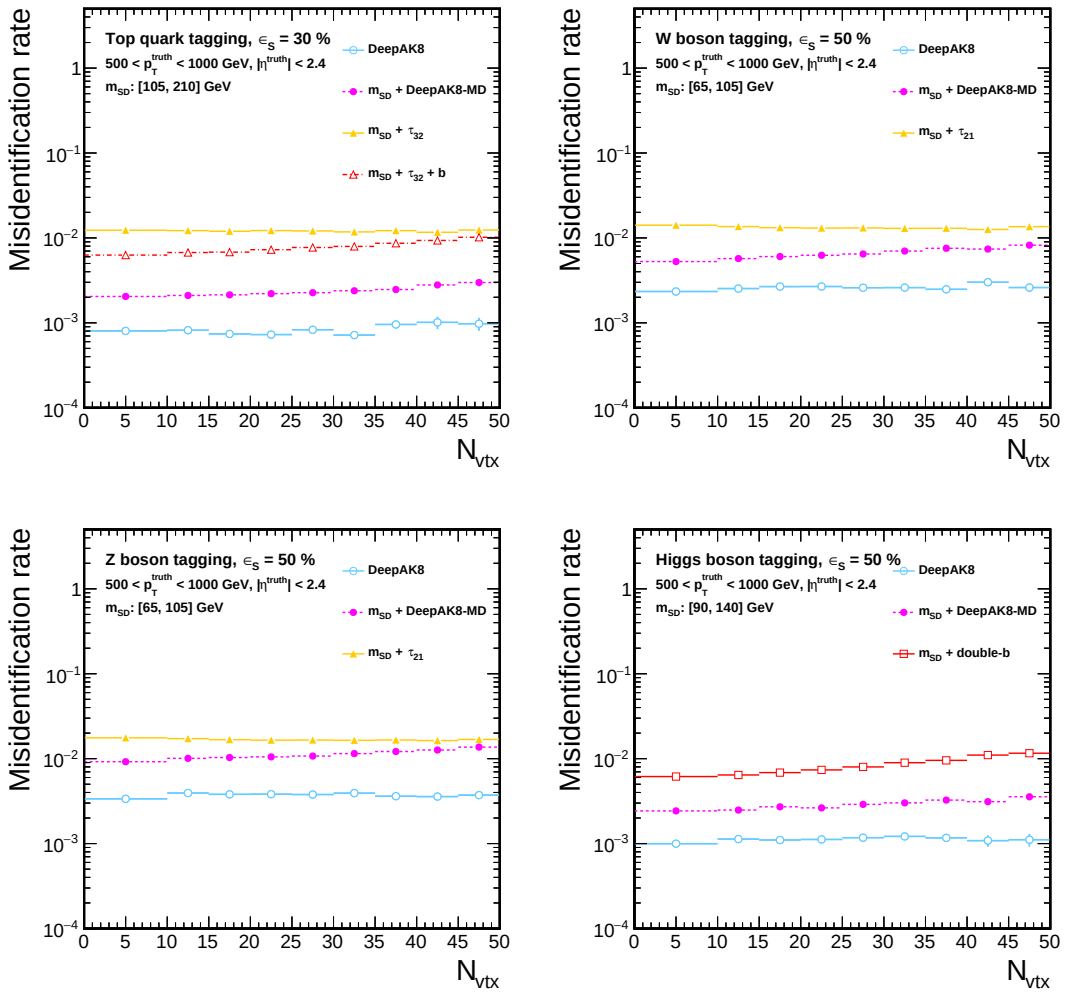


Figure 5.14: The distribution of ϵ_B as a function of N_{vtx} for a working point corresponding to $\epsilon_S = 30\%$ (50%) for top quark (W, Z, and Higgs boson) identification. Upper left: top quark, upper right: W boson, lower left: Z boson, lower right: Higgs boson. Additional fiducial selection criteria applied to the truth particles are displayed on the plots. For all algorithms except the nominal DeepAK8, a selection on the soft drop mass is applied and indicated on the plots.

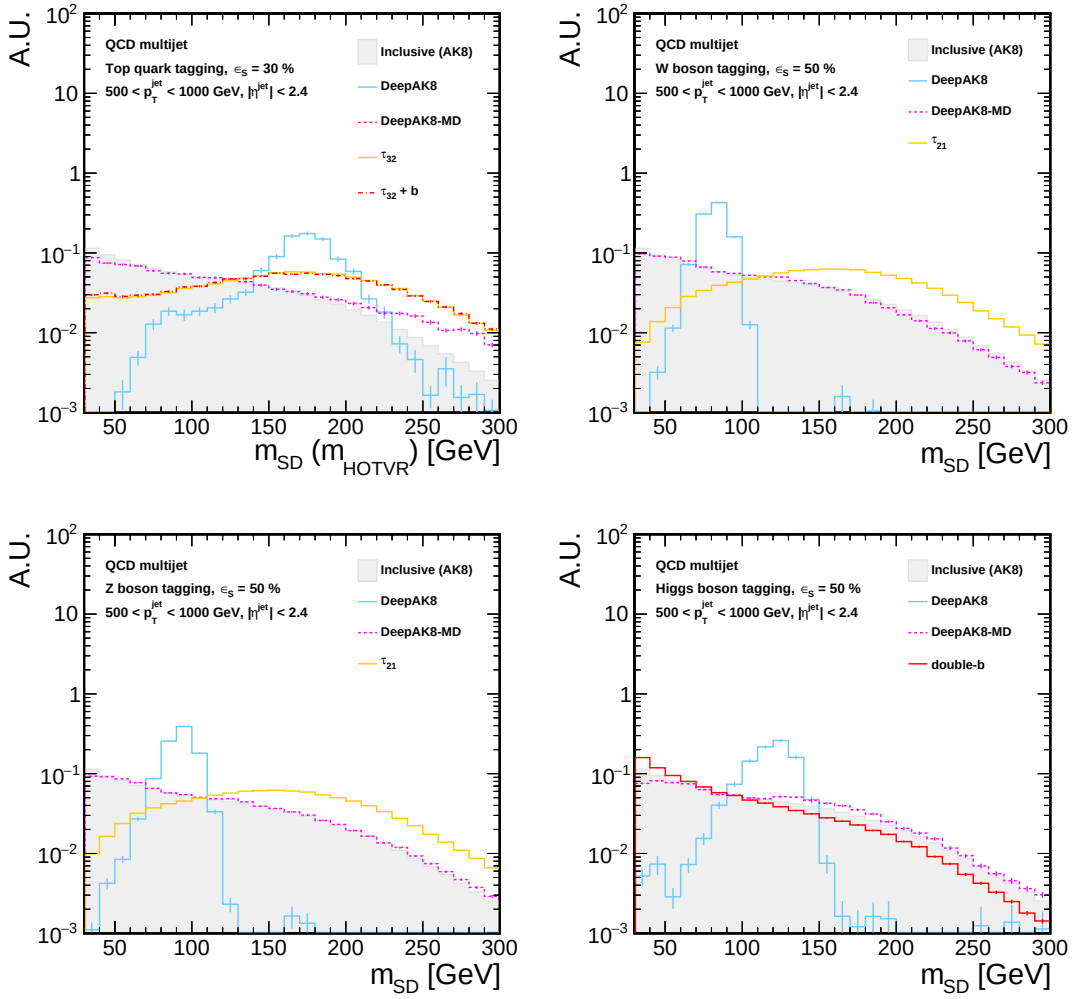


Figure 5.15: The shape of the soft drop mass distribution for background jets with $500 < p_{\mathrm{T}}(\mathrm{jet}) < 1000\,\mathrm{GeV}$, inclusively and after selection by each algorithm. The working point chosen corresponds to $\epsilon_S = 30\%$ (50%) for top quark (W, Z, and Higgs boson) tagging. Upper left: top quark, upper right: W boson, lower left:Z boson, lower right: Higgs boson. The error bars represent the statistical uncertainty in each specific bin, due to the limited number of simulated events. Additional fiducial selection criteria applied to the jets are displayed on the plots.

5.6 Top Quark and W Boson Identification Performance in Data

The performance of the DeepAK8 tagging algorithm is validated in data using several samples. In this section, we focus on the validation of the top quark and W boson identification. The signal efficiency for the top quark and W boson identification is calibrated in the single-μ sample, which is enriched in semi-leptonic $t\bar{t}$ events. The multijet sample, dominated by light-flavor quarks and gluons, is used to study the misidentification rate of background jets in a wide range of p_T. The misidentification rate depends on the flavor of the parton that initiated the jet. Therefore, in addition to the multijet sample, the single-γ sample is further utilized. The multijet and single-γ samples differ in the light-quark and gluon fractions, with the former having a larger fraction of gluon jets than the latter. A detailed description of the data and simulated samples, the event selections, and the systematic uncertainties is presented in Ref. [94]. In this section, we briefly discuss the event selection and summarize the results.

5.6.1 Event Selection

The single-μ sample

The single-μ sample is collected using a single-muon trigger that selects events with a high p_T muon. Candidate events are required to have exactly one muon with $p_T > 55\,\text{GeV}$ and $|\eta| < 2.4$ satisfying the tight identification requirement [95]. A custom isolation criterion is applied, by requiring a minimum distance between the muon and the nearest "AK4" jet, $\Delta R(\mu, \text{AK4}) > 0.4$, or the perpendicular component of the muon p_T with respect to the nearest AK4 jet, $p_{T,\text{rel}} > 25\,\text{GeV}$. The AK4 jets used in this selection are reconstructed by clustering the PF candidates using the anti-k_T algorithm with a

distance parameter of $R = 0.4$, after removing muons with $p_{\mathrm{T}} > 55\,\mathrm{GeV}$ from the event. The custom isolation requirement has better efficiency than the standard muon isolation variable computed with an isolation cone of $\Delta R = 0.4$ [95], as the lepton from the W boson decay often overlaps with the b jet from the top quark decay in boosted semi-leptonic $\mathrm{t\bar{t}}$ events.

To suppress the contribution from QCD multijet processes, we require $p_{\mathrm{T}}^{\mathrm{miss}} > 50\,\mathrm{GeV}$. To enhance the purity of $\mathrm{t\bar{t}}$ events, we require the presence of at least one b-tagged AK4 jet. In addition, to enrich boosted $\mathrm{t\bar{t}}$ events, we require the p_{T} of the leptonically decaying W boson, defined as $\vec{p}_{\mathrm{T}}(\mathrm{W}) = \vec{p}_{\mathrm{T}}(\mu) + \vec{p}_{\mathrm{T}}^{\mathrm{miss}}$, and the scalar sum p_{T} of the AK4 jets, denoted H_{T}, to be greater than 250 GeV. The highest p_{T} AK8 jet in the event with $p_{\mathrm{T}} > 200\,\mathrm{GeV}$ is selected to probe the hadronically decaying top quark or W boson. To further improve the purity, we require the azimuthal angle $\Delta\phi$ between the AK8 jet and the muon to be greater than 2 radians. The purity of the sample in semi-leptonic $\mathrm{t\bar{t}}$ events is $\sim 70\%$. Other contributions arise from QCD multijet ($\sim 15\%$) and W+jets ($\sim 10\%$) processes.

The multijet sample

The multijet sample is recorded using a trigger that requires high H_{T}. Events with $H_{\mathrm{T}} > 1000\,\mathrm{GeV}$ are selected to ensure 100% trigger efficiency. Events are required to have at least one AK8 jet and the absence of electrons or muons, leading to a sample dominated by jets from the QCD multijet process.

The single-γ sample

The single-γ sample is collected using a single-photon trigger that requires the presence of a high p_{T}, isolated photon. Events with a photon with $p_{\mathrm{T}} > 200\,\mathrm{GeV}$ and $|\eta| < 2.5$ are selected to ensure 100% trigger efficiency. The photon is further required to satisfy

the standard CMS identification criteria [96]. In addition, the events are required to have at least one AK8 jet. The sample consists of $\sim 80\%$ γ+jets events, with a smaller ($\sim 15\%$) contribution from QCD multijet events.

5.6.2 Signal Efficiency in the Single-μ Sample

The signal efficiency for the top quark and W boson identification can be determined using semi-leptonic $t\bar{t}$ events. The leptonically decaying W boson from one of the top quarks provides a clear signature and can be exploited for identifying such events. The other top quark, which decays hadronically, provides a boosted top quark or W boson that is needed to probe the tagging efficiency.

The data-to-simulation comparison of fundamental jet substructure variables, such as m_{SD}, the p_{T}(jet), the N-subjettiness ratios, τ_{32} and τ_{21}, are shown in Figure 5.16. Figure 5.17 displays the distributions of the top quark and W boson tagging discriminants for the DeepAK8 and DeepAK8-MD algorithms. To focus the comparison on the shape of the distribution, the total background yield is normalized to the observed number of data events to account for a small difference in the overall normalization between data and simulation. The systematic uncertainties discussed in Ref. [94] are taken into account and shown via the shaded blue band in the figures. Overall, the shapes in data are compatible with the expectation from simulation within uncertainties.

To measure the top quark and W boson tagging efficiency, the total SM sample is decomposed into three categories based on if the AK8 jet being probed is matched to a hadronically decaying top quark or a W boson. The "merged top quark" category requires that all three quarks from the top decay are fully contained in an AK8 jet, with $\Delta R(\mathrm{AK8}, \mathrm{q}) < 0.6$. The "merged W boson" category requires that only the two quarks from the W boson decay are within $\Delta R < 0.6$ of the AK8 jet, and the b quark

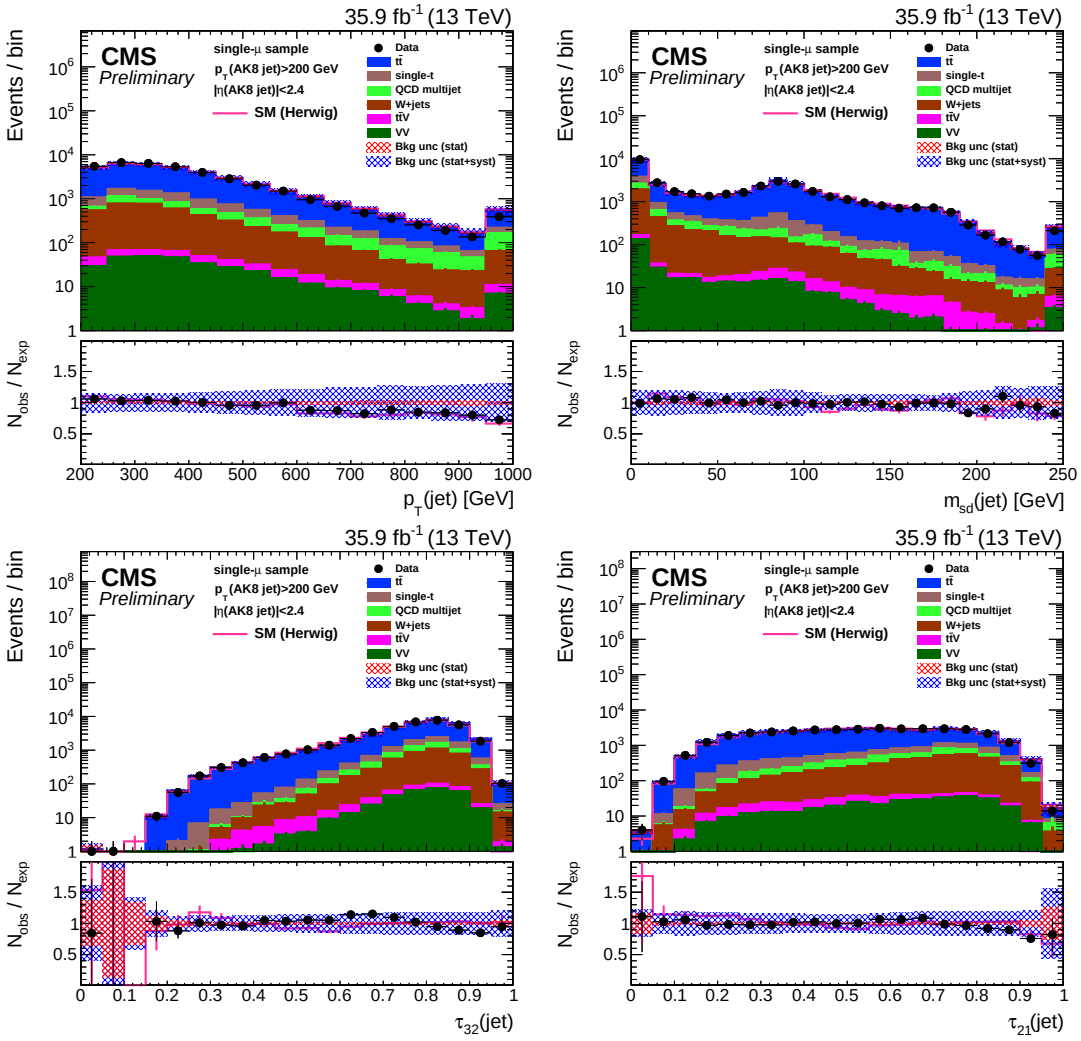


Figure 5.16: Distribution of the jet p_{T} (upper-left), the soft drop mass m_{SD} (upper-right), the N-subjettiness ratios, τ_{32} (lower-left) and τ_{21} (lower-right) in data and simulation in the single-μ sample. The nominal $\mathrm{t\bar{t}}$ sample is generated with POWHEG and showered with PYTHIA8, while the solid pink line corresponds to the distribution obtained from the $\mathrm{t\bar{t}}$ sample showered with HERWIG++. The background event yield is normalized to the total observed data yield. The lower panel shows the data to simulation ratio. The shaded blue (red) band corresponds to the total uncertainty (statistical uncertainty of the simulated samples), the pink line corresponds to the data to simulation ratio using the alternative $\mathrm{t\bar{t}}$ sample, and the vertical lines correspond to the statistical uncertainty of the data. Figures taken from [94].

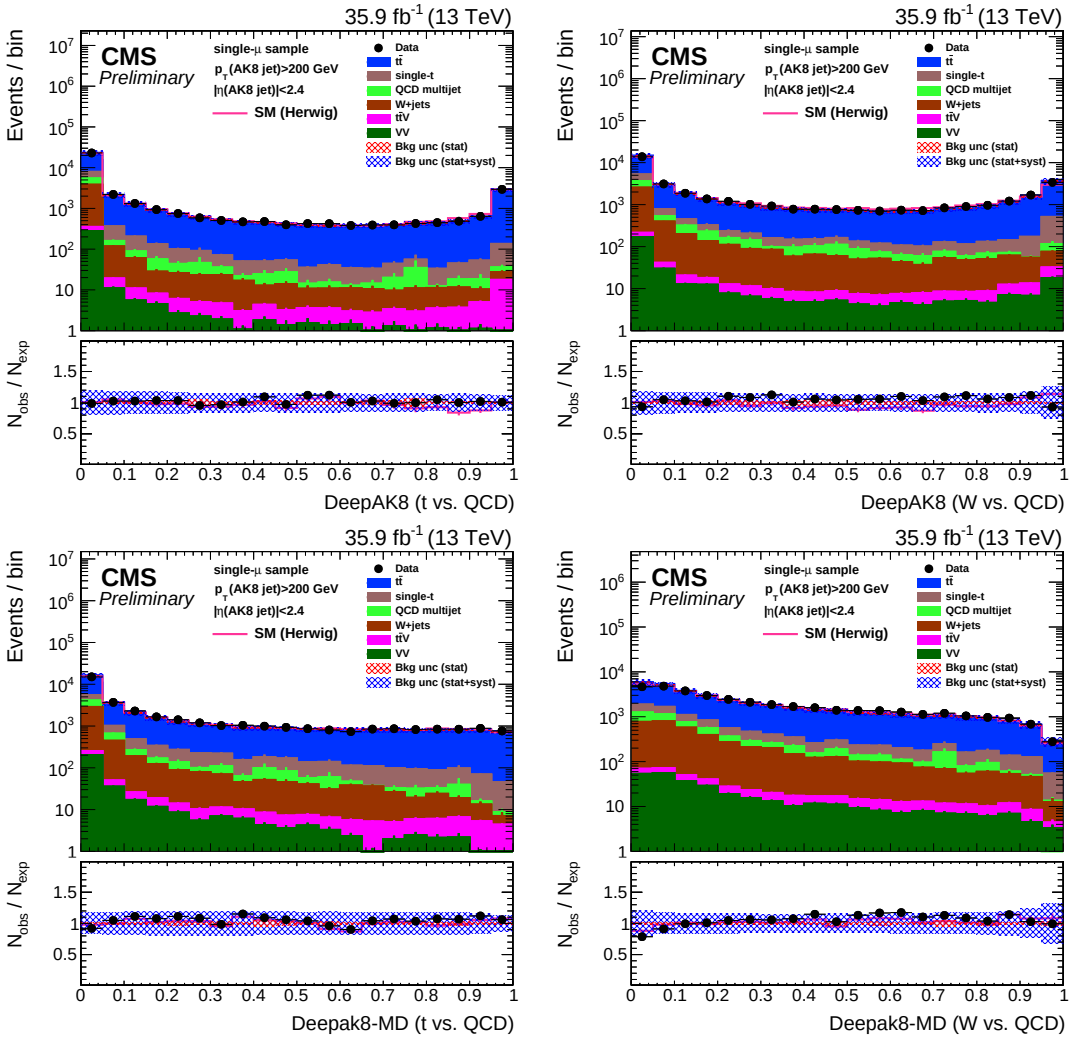


Figure 5.17: Distribution of the top quark (left) and W boson (right) tagging discriminants of the DeepAK8 (upper row) and DeepAK8-MD (lower row) algorithms in data and simulation in the single-μ sample. The nominal $t\bar{t}$ sample is generated with POWHEG and showered with PYTHIA8, while the solid pink line corresponds to the distribution obtained from the $t\bar{t}$ sample showered with HERWIG++. The background event yield is normalized to the total observed data yield. The lower panel shows the data to simulation ratio. The shaded blue (red) band corresponds to the total uncertainty (statistical uncertainty of the simulated samples), the pink line corresponds to the data to simulation ratio using the alternative $t\bar{t}$ sample, and the vertical lines correspond to the statistical uncertainty of the data. Figures taken from [94].

from the top decay is outside the jet cone. Any other scenario falls in the "unmerged" category. The tagging efficiency in data is extracted via a maximum likelihood fit of the m_{SD} distribution, with both regions, one passing and the other failing the tagging requirement, fitted to data simultaneously. The templates of the m_{SD} distribution are derived from simulation, and the normalization of the three categories in both the "pass" and "fail" regions are allowed to float in the fit. The signal efficiency can be extracted from the fit as:

$$\epsilon_S^{\mathrm{top}} = \frac{N^{\mathrm{pass}}_{\text{merged top quark}}}{N^{\mathrm{pass}}_{\text{merged top quark}} + N^{\mathrm{fail}}_{\text{merged top quark}}} \tag{5.3}$$

for top quark tagging and

$$\epsilon_S^{\mathrm{W}} = \frac{N^{\mathrm{pass}}_{\text{merged W boson}}}{N^{\mathrm{pass}}_{\text{merged W boson}} + N^{\mathrm{fail}}_{\text{merged W boson}}} \tag{5.4}$$

for W boson tagging, respectively. The fit is performed in the range from 50 to 250 GeV with a bin width of 10 GeV. The sources of systematic uncertainties discussed in Ref. [94] are considered and are treated as nuisance parameters in the fit. After calculating the efficiencies in data and simulation, the simulation-to-data scale factor is determined as:

$$\mathrm{SF} = \frac{\epsilon_{\mathrm{data}}}{\epsilon_{\mathrm{simulation}}}. \tag{5.5}$$

The SFs are extracted differentially in jet p_{T}. An example of m_{SD} distributions for data and simulation in the "pass" and "fail" regions for $400 < p_{\mathrm{T}} < 480$ GeV after performing the maximum likelihood fit are displayed in Fig 5.18. The SFs are summarized in Figure 5.19. They are typically consistent with unity, within uncertainties.

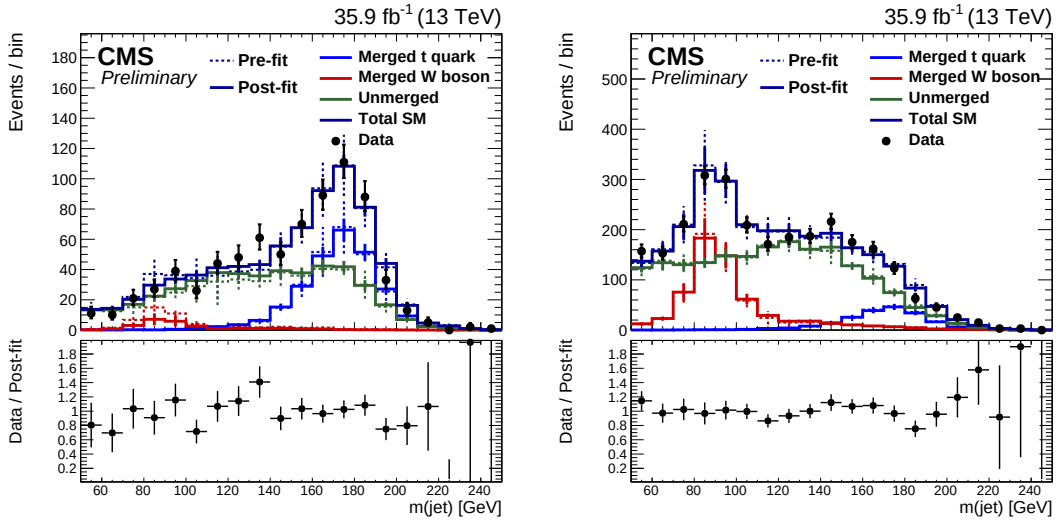


Figure 5.18: The m_{SD} distributions for data and simulation for events that pass (left) and fail (right) the top tagging requirement for AK8 jets with $400 < p_{\mathrm{T}}(\mathrm{jet}) < 480\,\mathrm{GeV}$. The solid (dashed) lines correspond to the contribution of each category after (before) the maximum likelihood fit as described in the text. The lower panel shows the ratio of the observed data to the post-fit expectation from simulation. Figures from [94].

5.6.3 Misidentification Rate in the Multijet and Single-γ Sample

The misidentification rate for the top quark and W boson identification is studied in the multijet and single-γ samples. The data-to-simulation comparison of some fundamental jet substructure variables, such as m_{SD}, the $p_{\mathrm{T}}(\mathrm{jet})$, the N-subjettiness ratios, τ_{32} and τ_{21}, are shown in Figure 5.20 and 5.21 for the multijet and single-γ samples, respectively. Figures 5.22 and 5.23 display the distributions of the top quark and W boson tagging discriminants for the DeepAK8 and DeepAK8-MD algorithms. For both samples, simulated events show a moderate discrepancy in the jet p_{T} distribution compared to data due to the lack of higher-order corrections in the simulation. Therefore, simulated events are reweighted to match the jet p_{T} distribution in data. For the multijet sample, two samples generated with different event generators are considered to study the dependence of the misidentification rate on the modeling of event kinematics and parton showering. The nominal sample uses MadGraph for the event generation and

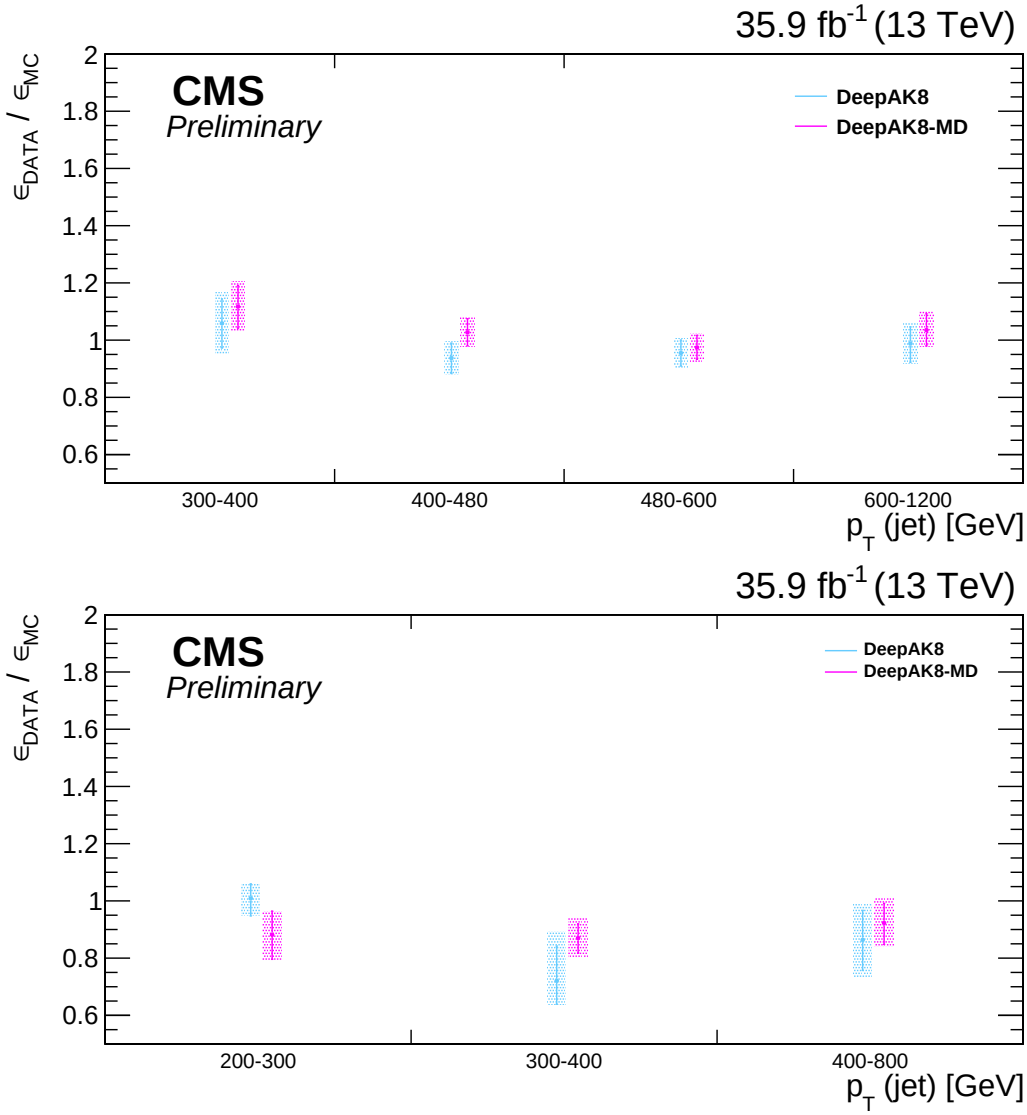


Figure 5.19: Summary of the signal efficiency SFs of the DeepAK8 and DeepAK8-MD algorithms for top quark (upper) and W boson (lower) identification. The markers correspond to the SF value, the error bars to the statistical uncertainty on the SF measurement, and the band is the total (statistical + systematic) uncertainty. Figures adapted from [94].

PYTHIA8 for the parton showering and hadronization, whereas the alternative sample uses HERWIG++ alone for event generation and the modeling of the parton showering. The systematic uncertainties discussed in Ref. [94] are taken into account and are shown via the shaded blue band in the figures. Overall, good agreement is observed between data and simulation within uncertainties. Some discrepancy is observed in the low-value region of τ_{32} and τ_{21} distribution. For the top and W tagging discriminants, the nominal DeepAK8 algorithm generally shows better agreement between data and simulation than the mass decorrelated DeepAK8-MD algorithm. The two multijet samples actually show a large difference here, with a better agreement observed with the sample generated with HERWIG++.

The misidentification rate in data is extracted from the multijet and single-γ samples as the fraction of events in which the probe jet passes the tagging requirement. The small contribution from processes with jets arising from real top quarks or W bosons, such as $t\bar{t}$ and W+jets, is subtracted from the observed yield using the expected yield from simulation with a conservative 100% uncertainty. The SFs for the misidentification rate extracted from the multijet and the single-γ samples are summarized in Figures 5.24 and 5.25 for the top quark and W boson identification, respectively. For top quark tagging, the misidentification rate SFs are typically between 1.0 and 1.2 in both the multijet and the single-γ sample for the DeepAK8 and DeepAK8-MD algorithms, and are fairly consistent between the MADGRAPH+PYTHIA8 and the HERWIG++ QCD multijet sample. On the other hand, the misidentification rate SFs for W boson tagging show larger variations between the two different QCD multijet samples, ranging typically between 0.6 and 1.4 for the DeepAK8 and DeepAK8-MD algorithms, while the SFs extracted from the single-γ sample are in between them and are around 1.2. This is likely due to the difference in the quark/gluon fraction between the multijet and the single-γ sample, and the difference in the modeling of gluons between PYTHIA8 and HERWIG++.

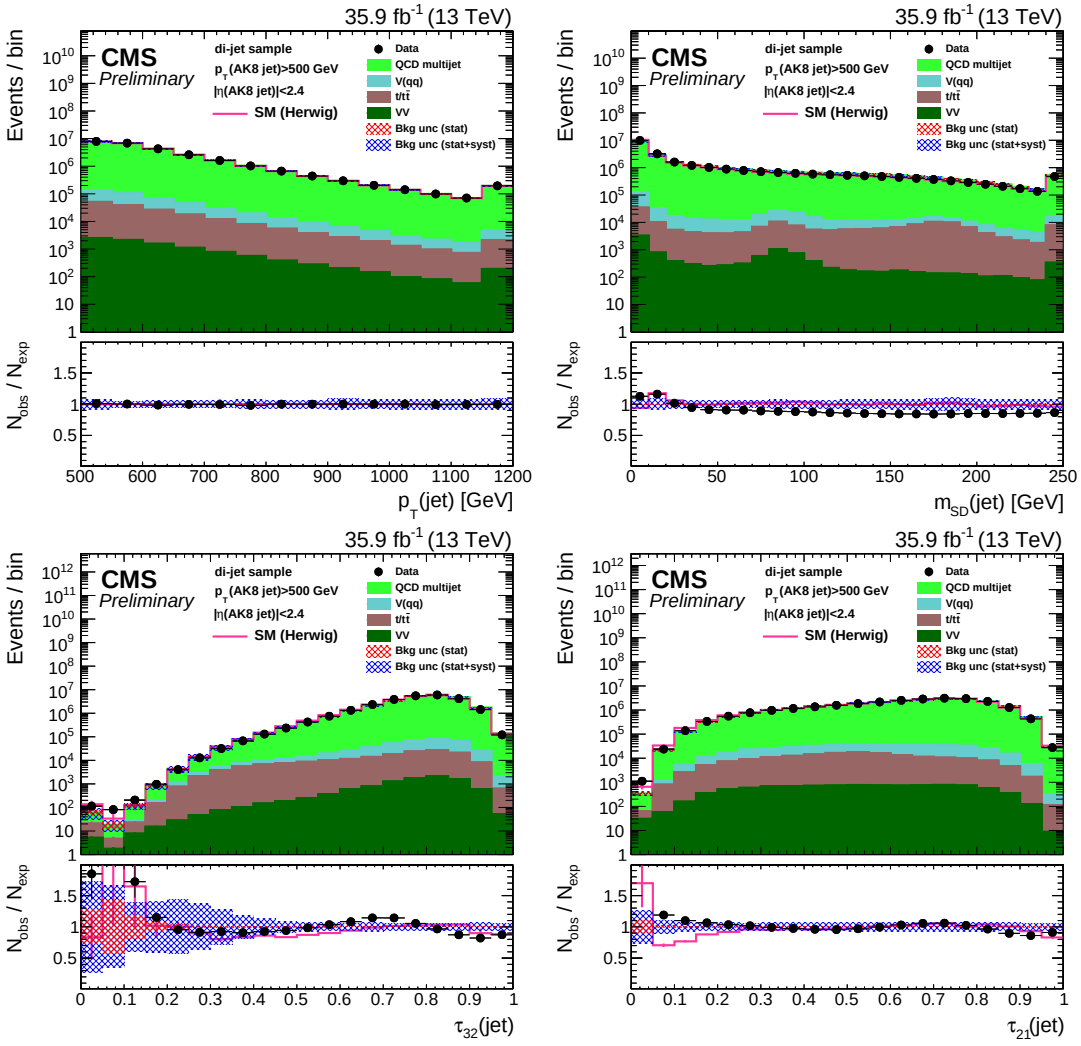


Figure 5.20: Distribution of the jet p_{T} (upper-left), the soft drop mass m_{SD} (upper-right), the N-subjettiness ratios, τ_{32} (lower-left) and τ_{21} (lower-right) in data and simulation in the multijet sample. The nominal QCD multijet sample is generated with MadGraph and showered with pythia8, while the solid pink line corresponds to the simulation distribution obtained using the alternative sample generated and showered with herwig++. The event yield is normalized to the total observed data yield as a function of jet p_{T}. The lower panel shows the data to simulation ratio. The shaded blue (red) band corresponds to the total uncertainty (statistical uncertainty of the simulated samples), the pink line corresponds to the data to simulation ratio using the alternative QCD multijet sample, and the vertical lines correspond to the statistical uncertainty of the data. Figures taken from [94].

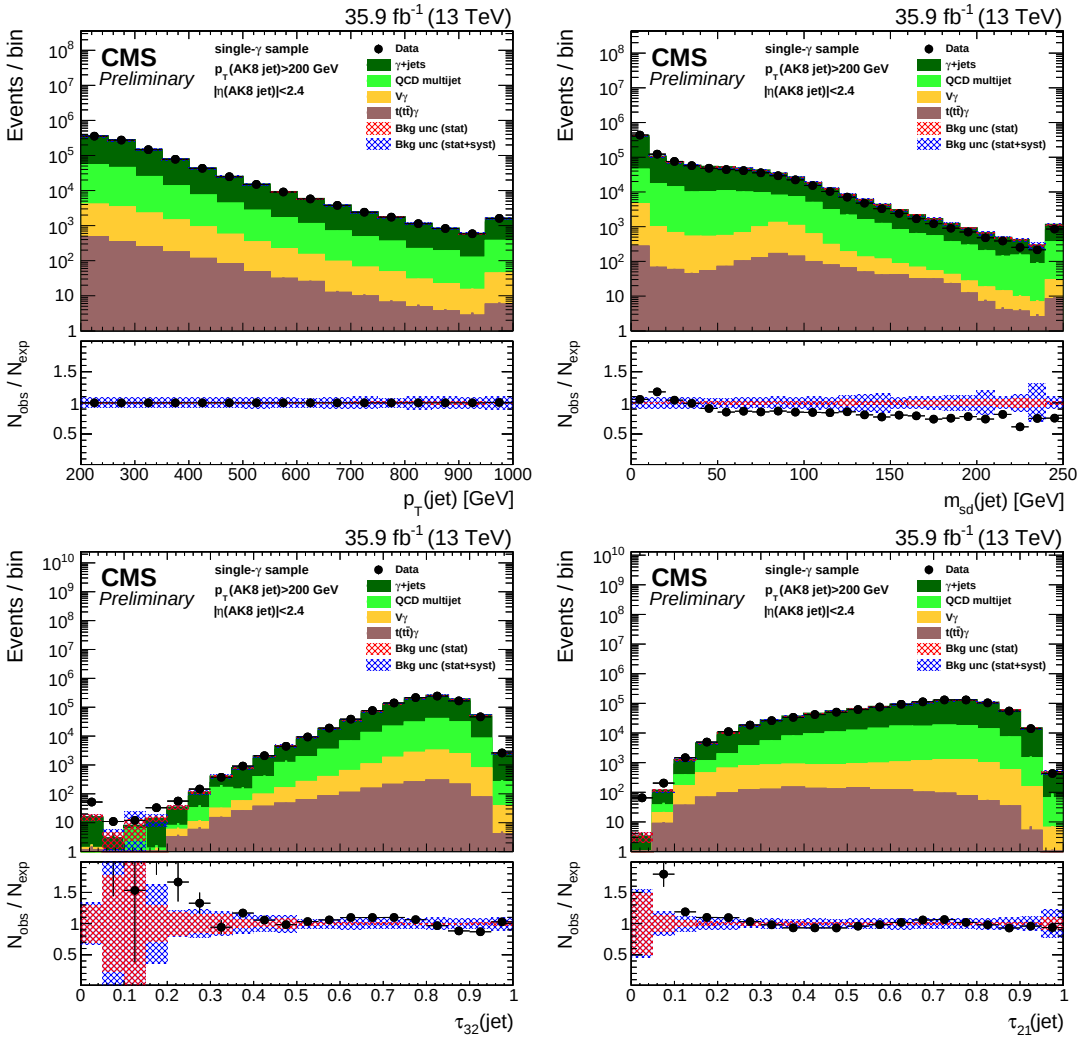


Figure 5.21: Distribution of the jet p_{T} (upper-left), the soft drop mass m_{SD} (upper-right), the N-subjettiness ratios, τ_{32} (lower-left) and τ_{21} (lower-right) in data and simulation in the single-γ sample. The event yield is normalized to the observed data yield as a function of jet p_{T}. The lower panel shows the data to simulation ratio. The shaded blue (red) band corresponds to the total uncertainty (statistical uncertainty of the simulated samples), and the vertical lines correspond to the statistical uncertainty of the data. Figures taken from [94].

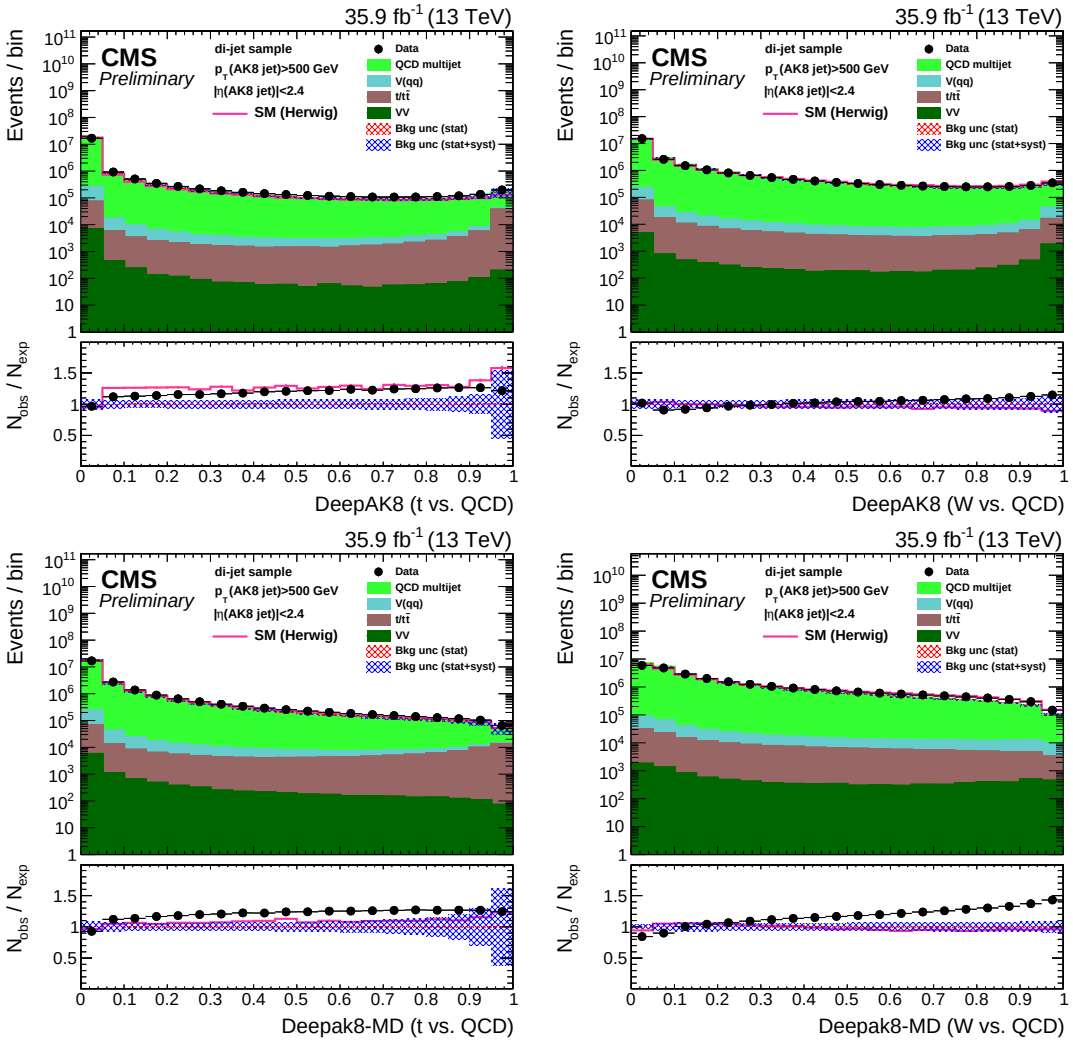


Figure 5.22: Distribution of the top quark (left) and W boson (right) tagging discriminants of the DeepAK8 (upper row) and DeepAK8-MD (lower row) algorithms in data and simulation in the multijet sample. The nominal QCD multijet sample is generated with MadGraph and showered with pythia8, while the solid pink line corresponds to the simulation distribution obtained using the alternative sample generated and showered with herwig++. The event yield is normalized to the total observed data yield as a function of jet p_{T}. The lower panel shows the data to simulation ratio. The shaded blue (red) band corresponds to the total uncertainty (statistical uncertainty of the simulated samples), the pink line corresponds to the data to simulation ratio using the alternative QCD multijet sample, and the vertical lines correspond to the statistical uncertainty of the data. Figures taken from [94].

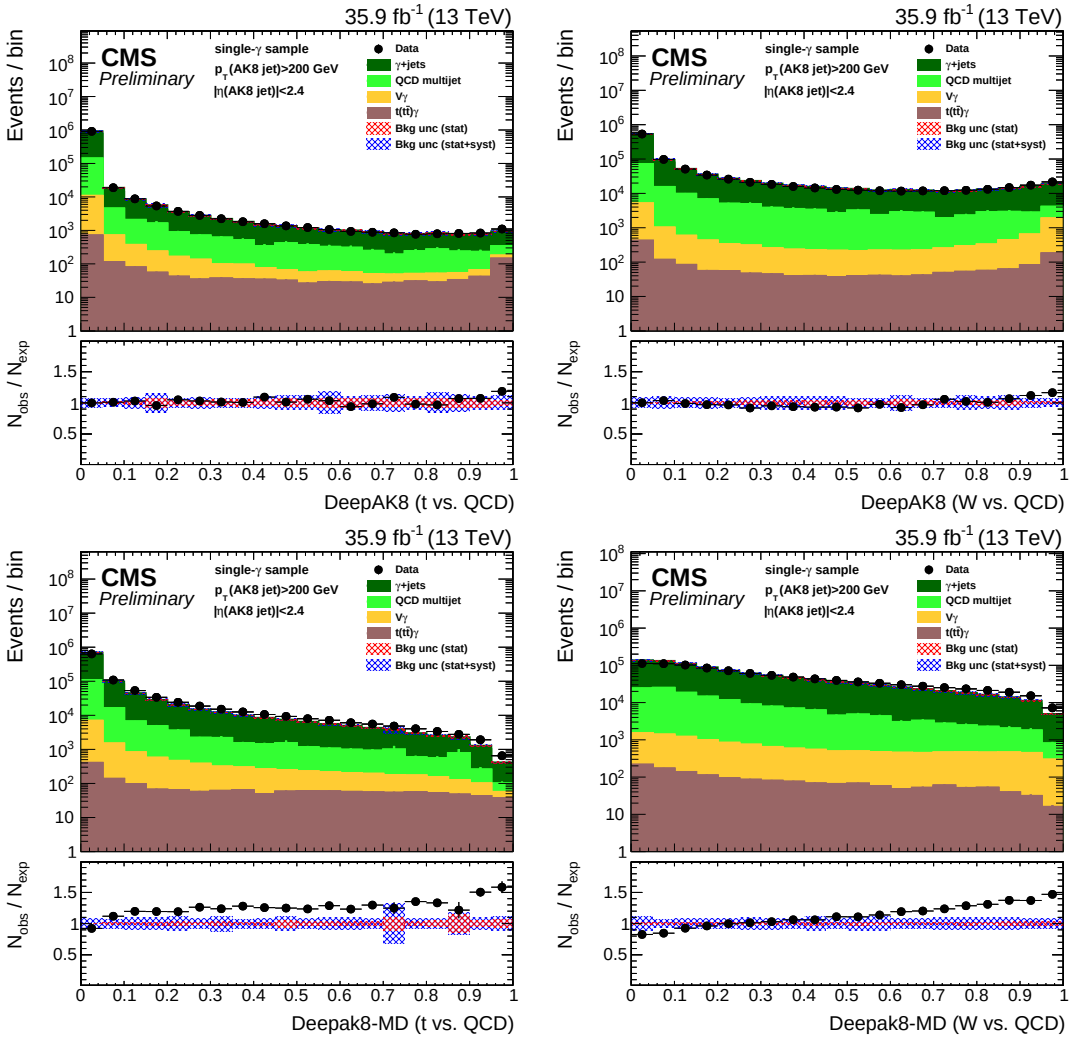


Figure 5.23: Distribution of the top quark (left) and W boson (right) tagging discriminants of the DeepAK8 (upper row) and DeepAK8-MD (lower row) algorithms in data and simulation in the single-γ background sample. The event yield is normalized to the observed data yield as a function of jet p_{T}. The lower panel shows the data to simulation ratio. The shaded blue (red) band corresponds to the total uncertainty (statistical uncertainty of the simulated samples), and the vertical lines correspond to the statistical uncertainty of the data. Figures taken from [94].

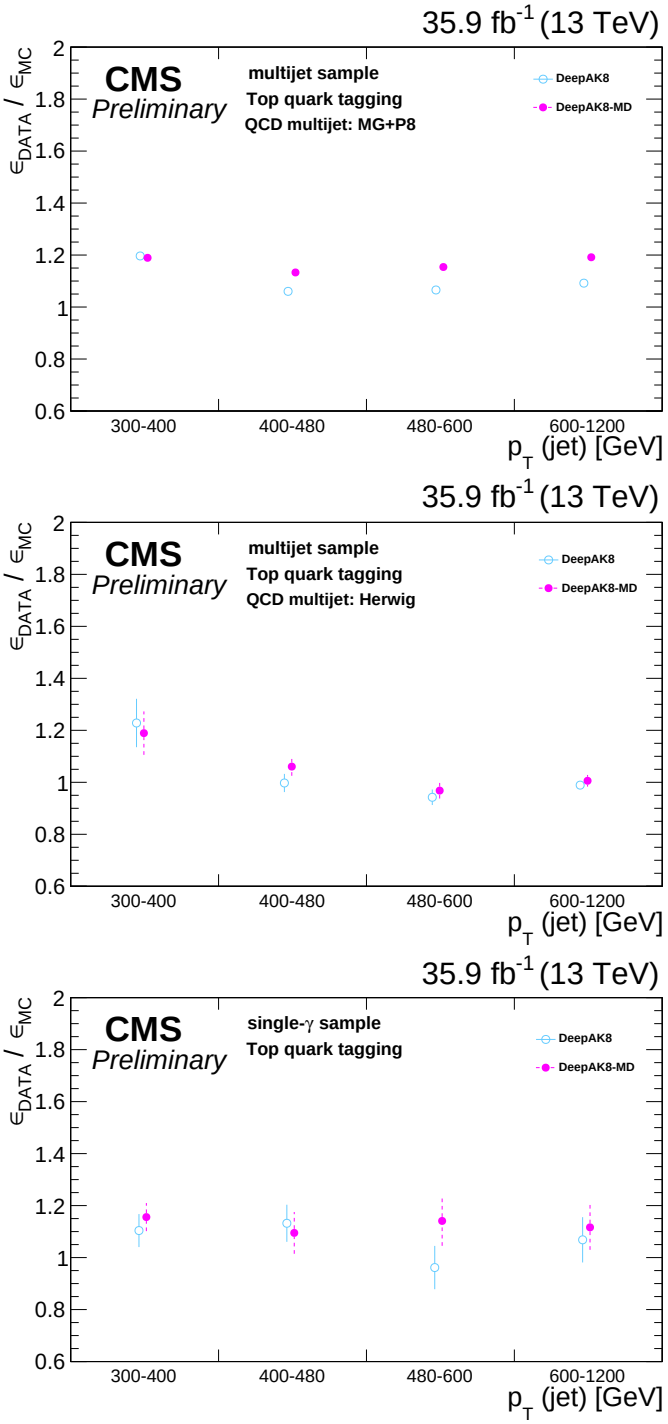


Figure 5.24: The ratio of the misidentification rate in data to that in simulation for top quark identification in the multijet (upper and middle rows) and the single-γ (lower row) samples. The QCD multijet process is simulated using MadGraph (MG) for the hard-scatter process and pythia8 (P8) for parton showering (upper) and herwig++ for both (middle). Figures adapted from [94].

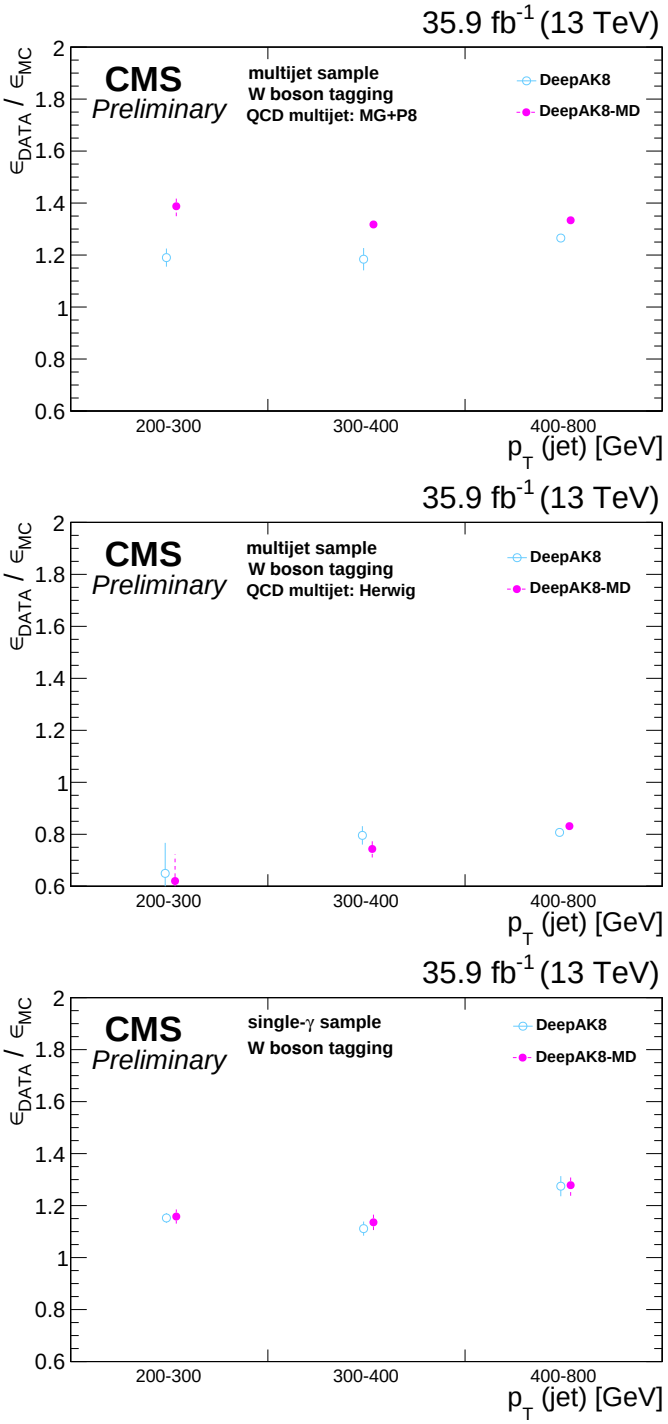


Figure 5.25: The ratio of the misidentification rate in data to that in simulation for W boson tagging in the multijet (upper and middle rows) and the single-γ (lower row) samples. The QCD multijet process is simulated using MadGraph (MG) for the hard process and pythia8 (P8) for parton showering (upper) and herwig++ for both (middle). Figures adapted from [94].

Chapter 6

ParticleNet: Jet Tagging via Particle Clouds

In this chapter, a new deep learning approach for jet tagging using a novel way to represent jets is presented[1]. Instead of organizing a jet's constituent particles into an ordered structure (e.g., a sequence or a tree), a jet is treated as an *unordered* set of particles. This is very analogous to the point cloud representation of 3D shapes used in computer vision, where each shape is represented by a set of points in space, and the points themselves are also unordered. Therefore, a jet can be viewed as a "particle cloud". Based on Dynamic Graph CNN [98], "ParticleNet", a customized neural network architecture is designed that operates directly on particle clouds for jet tagging. The ParticleNet architecture is evaluated on two jet tagging benchmarks and is found to achieve significant improvements over all existing methods.

[1]This chapter is largely based on Ref. [97], with only some small modification.

6.1 Jet Representations

The efficiency and effectiveness of ML techniques on jet physics relies heavily on how a jet is represented. In this section, we review the mainstream jet representations and introduce the particle cloud representation.

6.1.1 Image-Based Representation

The image representation has its root in the reconstruction of jets with calorimeters. A calorimeter measures the energy deposition of a jet on fine-grained spatial cells. Treating the energy deposition on each cell as the pixel intensity naturally creates an image for a jet. When jets are formed by particles reconstructed with the full detector information (e.g., using a particle-flow algorithm [37, 99]), a jet image can be constructed by mapping each particle onto the corresponding calorimeter cell, and sum up the energy if more than one particle is mapped to the same cell.

The image-based approach has been extensively studied for various jet tagging tasks, e.g., W boson tagging [100, 101, 102, 103, 104, 105], top tagging [106, 107, 108] and quark-gluon tagging [109, 110]. Convolutional neural networks (CNNs) with various architectures were explored in these studies, and they were found to achieve sizable improvement in performance compared to traditional multivariate methods using observables motivated by QCD theory. However, the architectures investigated in these papers are in general much shallower compared to state-of-the-art CNN architectures used in image classification tasks (e.g., ResNet [88] or Inception [111]). Therefore it remains to be seen that if deeper architectures can further improve the performance.

Despite the promising performance, the image-based representation has two main shortcomings. While it can include all information without loss when a jet is only measured by the calorimeter, once the jet constituent particles are reconstructed, how to

incorporate additional information about the particles is unclear, as it involves combining non-additive quantities (e.g., the particle type) of multiple particles entering the same cell. Moreover, treating jets as images also lead to a very sparse representation: A typical jet has $\mathcal{O}(10)$ to $\mathcal{O}(100)$ particles, while a jet image typically needs $\mathcal{O}(1000)$ pixels (e.g., 32×32) in order to fully contain the jet. Therefore, more than 90% of the pixels are blank. This makes the CNNs highly computationally inefficient on jet images.

6.1.2 Particle-Based Representation

A more natural way to represent a jet, when particles are reconstructed, is to simply view the jet as a collection of its constituent particles. This approach allows for the inclusion of any kind of features for each particle and therefore is significantly more flexible than the image representation. It is also much more compact compared to the image representation, though at the cost of being variable-length, as each jet may contain a different number of particles.

A collection of particles, though, is a rather general concept. Before applying any deep learning algorithm, a concrete data structure has to be chosen. The prevailing choice is a sequence, where particles are sorted in a specific way (e.g., with decreasing transverse momentum) and organized into a 1D list. Using particle sequences as inputs, jet tagging tasks have been tackled with recurrent neural networks (RNNs) [112, 113, 114, 115, 116], 1D CNNs [117, 118, 119, 120] and physics-oriented neural networks [121, 122, 123]. Another interesting choice is a binary tree, which is well motivated from the QCD theory perspective. Recursive neural networks (RecNNs) are then a natural fit and have been studied [124, 125].

One thing to note about the sequence or tree representation is that they both need the particles to be sorted in some way, as the order of the particles is used implicitly in

the corresponding RNNs, 1D CNNs or the RecNNs. However, the constituent particles in a jet have no intrinsic order. Thus, the manually imposed order may turn out to be suboptimal and impair the performance.

6.1.3 Jet as a Particle Cloud

An even more natural representation than particle sequences or trees would be an unordered, permutation-invariant *set* of particles. As a special case of the particle-based representations, it shares all their advantages, especially the flexibility to include arbitrary features for each particle. We refer to such a representation of a jet as a "particle cloud", analogous to the point cloud representation of 3D shapes used in computer vision. They are actually highly similar, as both are essentially unordered sets of entities distributed irregularly in space. In both clouds, the elements are not unrelated individuals, but are rather correlated, as they represent higher-level objects (i.e., jets or 3D shapes) that have rich internal structures. Therefore, deep-learning algorithms developed for point clouds are likely to be helpful for particle clouds, i.e., jets, as well.

The idea of regarding jets as unordered sets of particles was also recently proposed in [126] in parallel with our work. The Deep Sets framework [127] was adapted to construct the infrared and collinear safe Energy Flow Network, and the more general Particle Flow Network. However, different from the Dynamic Graph CNN [98] approach adopted in our work, the Deep Sets approach does not explicitly exploit the local spatial structure of particle clouds, but only process the particle clouds in a global way. Another closely related approach is to represent a jet as a graph whose vertices are the particles. Message-passing neural networks (MPNNs) with different variants of adjacency matrices were explored on such jet graphs and were found to show better performance than the RecNNs [128]. However, depending on how the adjacency matrix is defined, the MPNNs

may not respect the permutation symmetry of the particles.

6.2 Network Architecture

The permutation symmetry of the particle cloud makes it a natural and promising representation of jets. However, to achieve the best possible performance, the architecture of the neural network has to be carefully designed to fully exploit the potential of this representation. In this section, we introduce ParticleNet, a CNN-like deep neural network for jet tagging with particle cloud data.

6.2.1 Edge Convolution

CNNs have achieved overwhelming success in all kinds of machine learning tasks on visual images. Two key features of CNNs contribute significantly to their success. Firstly, the convolution operation exploits translational symmetry of images by using shared kernels across the whole image. This not only greatly reduces the number of parameters in the network, but also allows the parameters to be learned more effectively, as each set of weights will use all locations of the image for learning. Secondly, CNNs exploit a hierarchical approach [129] for learning image features. The convolution operations can be effectively stacked to form a deep network. Different layers in the CNNs have different receptive fields, therefore can learn features at different scales, with the shallower layers exploiting local neighborhood information and the deeper layers learning more global structures. Such a hierarchical approach proves an effective way of learning on images.

Motivated by the success of CNNs, we would like to adopt a similar approach for learning on point (particle) cloud data. However, regular convolution operations cannot be applied on point clouds, as the points there can be distributed irregularly, rather than following some uniform grids like pixels in an image. Therefore, the basis for a convolu-

tion, i.e., a "local patch" of each point on which the convolution kernel operates, remains to be defined for point clouds. Moreover, a regular convolution operation, typically in the form $\sum_j K_j x_j$ where K is the kernel and x_j denotes the features of each point, is not invariant under permutation of the points. Thus, the form of a convolution also needs to be modified to respect the permutation symmetry of point clouds.

Recently, the edge convolution ("EdgeConv") operation has been proposed in [98] as a convolution-like operation for point clouds. EdgeConv starts by representing a point cloud as a graph, whose vertices are the points themselves, and the edges are constructed as connections between each point to its k nearest neighboring points. In this way, a local patch needed for convolution is defined for each point as the k nearest neighboring points connected to it. The EdgeConv operation for each point x_i then has the form

$$\boldsymbol{x}_i' = \square_{j=1}^{k} \boldsymbol{h}_{\boldsymbol{\Theta}}(\boldsymbol{x}_i, \boldsymbol{x}_{i_j}), \tag{6.1}$$

where $\boldsymbol{x}_i \in \mathbb{R}^F$ denotes the feature vector of the point x_i and $\{i_1, ..., i_k\}$ are the indices of the k nearest neighboring points of the point x_i. The edge function $\boldsymbol{h}_{\boldsymbol{\Theta}} : \mathbb{R}^F \times \mathbb{R}^F \to \mathbb{R}^{F'}$ is some function parameterized by a set of learnable parameters $\boldsymbol{\Theta}$, and $\square$ is a channel-wise symmetric aggregation operation, e.g., max, sum or mean. The parameters $\boldsymbol{\Theta}$ of the edge function are shared for all points in the point cloud. This, together with the choice of a symmetric aggregation operation $\square$, makes EdgeConv a permutationally symmetric operation on point clouds.

In this work, we follow the choice in [98] to use a specialized form of the edge function,

$$\boldsymbol{h}_{\boldsymbol{\Theta}}(\boldsymbol{x}_i, \boldsymbol{x}_{i_j}) = \bar{\boldsymbol{h}}_{\boldsymbol{\Theta}}(\boldsymbol{x}_i, \boldsymbol{x}_{i_j} - \boldsymbol{x}_i), \tag{6.2}$$

where the feature vectors of the neighbors, $\boldsymbol{x}_{i_j}$, are substituted by their differences from

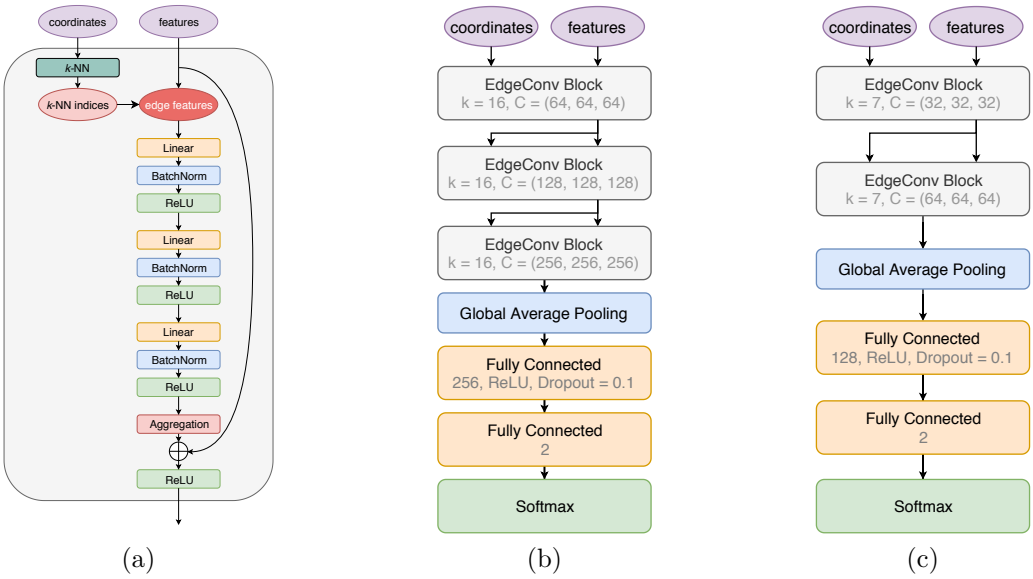
the central point $\boldsymbol{x}_i$, and $\bar{\boldsymbol{h}}_{\boldsymbol{\Theta}}$ can be implemented as a multilayer perceptron (MLP) whose parameters are shared among all edges. For the aggregation operation $\square$, however, we use mean, i.e., $\frac{1}{k}\sum$, throughout this work, which shows better performance than the max operation used in the original paper.

One important feature of the EdgeConv operation is that it can be easily stacked, just like regular convolutions. This is because EdgeConv can be viewed as a mapping from a point cloud to another point cloud with the same number of points, only possibly changing the dimension of the feature vector for each point. Therefore, another EdgeConv operation can be applied subsequently. This allows us to build a deep network using EdgeConv operations which can learn features of point clouds hierarchically.

The stackability of EdgeConv operations also brings another interesting possibility. Basically, the feature vectors learned by EdgeConv can be viewed as new coordinates of the original points in a latent space, and then, the distances between points, used in the determination of the k nearest neighbors, can be computed in this latent space. In other words, the proximity of points can be dynamically learned with EdgeConv operations. This results in the Dynamic Graph CNN [98], where the graph describing the point clouds are dynamically updated to reflect the changes in the edges, i.e., the neighbors of each point. Ref. [98] demonstrates that this leads to better performance than keeping the graph static.

6.2.2 ParticleNet

The ParticleNet architecture makes extensive use of EdgeConv operations and also adopts the dynamic graph update approach. However, a number of different design choices are made in ParticleNet compared to the original Dynamic Graph CNN to better suit the jet tagging task, including the number of neighbors, the configuration of the



(a) (b) (c)

Figure 6.1: The architecture of (a) the EdgeConv block, (b) the ParticleNet network and (c) the ParticleNet-Lite network.

MLP in EdgeConv, the use of shortcut connection, etc.

Figure 6.1a illustrates the structure of the EdgeConv block implemented in this work. The EdgeConv block starts with finding the k nearest neighboring particles for each particle, using the "coordinates" input of the EdgeConv block to compute the distances. Then, inputs to the EdgeConv operation, the "edge features", are constructed from the "features" input using the indices of k nearest neighboring particles. The EdgeConv operation is implemented as a 3-layer MLP. Each layer consists of a linear transformation, followed by a batch normalization [130] and then the ReLU nonlinearity [89]. Inspired by ResNet [88], a shortcut connection running parallel to the EdgeConv operation is also included in each block, allowing the input features to pass through directly. An EdgeConv block is characterized by two hyperparameters, the number of neighbors k, and the number of channels $C = (C_1, C_2, C_3)$, corresponding to the number of units in each linear transformation layer.

The ParticleNet architecture used in this work is shown in Figure 6.1b. It consists

of three EdgeConv blocks. The first EdgeConv block uses the spatial coordinates of the particles in the pseudorapidity–azimuth space to compute the distances, while the subsequent blocks use the learned feature vectors as coordinates. The number of nearest neighbors k is 16 for all three blocks, and the number of channels C for each EdgeConv block is (64, 64, 64), (128, 128, 128), and (256, 256, 256), respectively. After the EdgeConv blocks, a channel-wise global average pooling operation is applied to aggregate the learned features over all particles in the cloud. This is followed by a fully-connected layer with 256 units and the ReLU activation. A dropout layer [90] with a drop probability of 0.1 is included to prevent overfitting. A fully-connected layer with 2 units, followed by a softmax function, is used to generate the output for the binary classification task.

A similar network with reduced complexity is also investigated. Compared to the baseline ParticleNet architecture, only two EdgeConv blocks are used, with the number of nearest neighbors k reduced to 7 and the number of channels C reduced to (32, 32, 32) and (64, 64, 64) for the two blocks, respectively. The number of units in the fully-connected layer after pooling is also lowered to 128. This simplified architecture is denoted as "ParticleNet-Lite" and is illustrated in Figure 6.1c. The number of arithmetic operations is reduced by almost an order of magnitude in ParticleNet-Lite, making it more suitable when computational resources are limited.

The networks are implemented with Apache MXNet [91] and the training is performed on a single Nvidia GTX 1080 Ti graphics card (GPU). A batch size of 384 (1024) is used for the ParticleNet (ParticleNet-Lite) architecture due to GPU memory constraint. The AdamW optimizer [131], with a weight decay of 0.0001, is used to minimize the cross entropy loss. The 1-cycle learning rate (LR) schedule [77] is adopted in the training, with the LR selected following the LR range test described in [77], and slightly tuned afterward with a few trial trainings. The training of ParticleNet (ParticleNet-Lite) network uses an initial LR of 3×10^{-4} (5×10^{-4}), rising to the peak LR of 3×10^{-3} (5×10^{-3}) linearly

Variable	Definition	TOP	QG	QG-PID
$\Delta\eta$	difference in pseudorapidity between the particle and the jet axis	x	x	x
$\Delta\phi$	difference in azimuthal angle between the particle and the jet axis	x	x	x
$\log p_T$	logarithm of the particle's p_T	x	x	x
$\log E$	logarithm of the particle's energy	x	x	x
$\log \frac{p_T}{p_T(\text{jet})}$	logarithm of the particle's p_T relative to the jet p_T	x	x	x
$\log \frac{E}{E(\text{jet})}$	logarithm of the particle's energy relative to the jet energy	x	x	x
ΔR	angular separation between the particle and the jet axis ($\sqrt{(\Delta\eta)^2+(\Delta\phi)^2}$)	x	x	x
q	electric charge of the particle			x
`isElectron`	if the particle is an electron			x
`isMuon`	if the particle is a muon			x
`isChargedHadron`	if the particle is a charged hadron			x
`isNeutralHadron`	if the particle is a neutral hadron			x
`isPhoton`	if the particle is a photon			x

Table 6.1: Input variables used in the top tagging task (TOP) and the quark-gluon tagging task (QG) with and without particle identification (PID) information.

in 8 epochs, and then decreasing to the initial LR linearly in another 8 epochs. This is followed by a cool-down phase of 4 epochs which gradually reduces the LR to 5×10^{-7} (1×10^{-6}) for better convergence. A snapshot of the model is saved at the end of each epoch, and the model snapshot showing the best accuracy on the validation dataset is selected for the final evaluation.

6.3 Results

The performance of the ParticleNet architecture is evaluated on two representative jet tagging tasks: top tagging and quark-gluon tagging. In this section, we show the benchmark results.

6.3.1 Top Tagging

Top tagging, i.e., identifying jets originating from hadronically decaying top quarks, is commonly used in searches for new physics at the LHC. We evaluate the performance of the ParticleNet architecture on this task using the top tagging dataset [132], which is

an extension of the dataset used in [121] with some modifications. Jets in this dataset are generated with PYTHIA8 [82] and passed through DELPHES [133] for fast detector simulation. No multiple parton interaction or pileup is included in the simulation. Jets are clustered from the Delphes E-Flow objects with the anti-k_T algorithm [40] using a distance parameter $R = 0.8$. Only jets with transverse momentum $p_T \in [550, 650]$ and pseudorapidity $|\eta| < 2$ are considered. Each signal jet is required to be matched to a hadronically decaying top quark within $\Delta R = 0.8$, and all three quarks from the top decay must also be within $\Delta R = 0.8$ of the jet axis. The background jets are obtained from a QCD dijet process. This dataset consists of 2 million jets in total, half signal and half background. The official splitting for training (1.2M jets), validation (400k jets) and testing (400k jets) is used in the development of the ParticleNet model for this dataset.

In this dataset, up to 200 jet constituent particles are stored for each jet. Only kinematic information, i.e., the four-momentum (p_x, p_y, p_z, E), of each particle is available. The ParticleNet model takes up to 100 constituent particles with the highest p_T for each jet, and uses 7 variables derived from the four-momentum for each particle as inputs, which are listed in Table 6.1. The $(\Delta\eta, \Delta\phi)$ variables are used as coordinates to compute the distances between particles in the first EdgeConv block. They are also used together with the other 5 variables, $\log p_T$, $\log E$, $\log \frac{p_T}{p_T(\text{jet})}$, $\log \frac{E}{E(\text{jet})}$ and ΔR, to form the input feature vector for each particle.

We compare the performance of ParticleNet with three alternative models: [2]

- **ResNeXt-50:** The ResNeXt-50 model is a very deep 2D CNN using jet images as inputs. The ResNeXt architecture [135] was proposed for generic image classification, and we modify it slightly for the jet tagging task. The model is trained on the top tagging dataset starting from randomly initialized weights. Note that the

[2] A comprehensive comparison between a wide range of machine learning approaches on this top tagging dataset is presented in Ref. [134], where an earlier version of ParticleNet is also included.

ResNeXt-50 architecture is much deeper and therefore has a much larger capacity than most of the CNN architectures [100, 102, 103, 104, 105, 106, 107, 109, 110, 108] explored for jet tagging so far, so evaluating its performance on jet tagging will shed light on whether architectures for generic image classification are also applicable to jet images.

- **P-CNN:** The P-CNN is a 14-layer 1D CNN using particle sequences as inputs. It is essentially the DeepAK8 architecture described in Section 5, but using only input features available in the top tagging dataset (i.e., only the 4-momenta of the particle, without properties of the tracks or secondary vertices). The model is trained on the top tagging dataset from scratch.

- **PFN:** The Particle Flow Network (PFN) [126] is a recent architecture for jet tagging which also treats a jet as an unordered set of particles, same as the particle cloud approach in this work. However, the network is based on the Deep Sets framework [127], which uses global symmetric functions and does not exploit local neighborhood information explicitly as is done with the EdgeConv operation. Since the performance of PFN on this top tagging dataset has already been reported in [126], we did not re-implement it but just include the results for comparison.

The results are summarized in Table 6.2 and also shown in Figure 6.2a in terms of receiver operating characteristic (ROC) curves. A number of metrics are used to evaluate the performance, including the accuracy, the area under the ROC curve (AUC), and the background rejection ($1/\varepsilon_b$, i.e., the reciprocal of the background misidentification rate) at a certain signal efficiency (ε_s) of 50% or 30%. The background rejection metric is particularly relevant to physics analysis at the LHC, as it is directly related to the expected contribution of background, and is commonly used to select the best jet tagging algorithm. The ParticleNet model achieves state-of-the-art performance on the top tagging

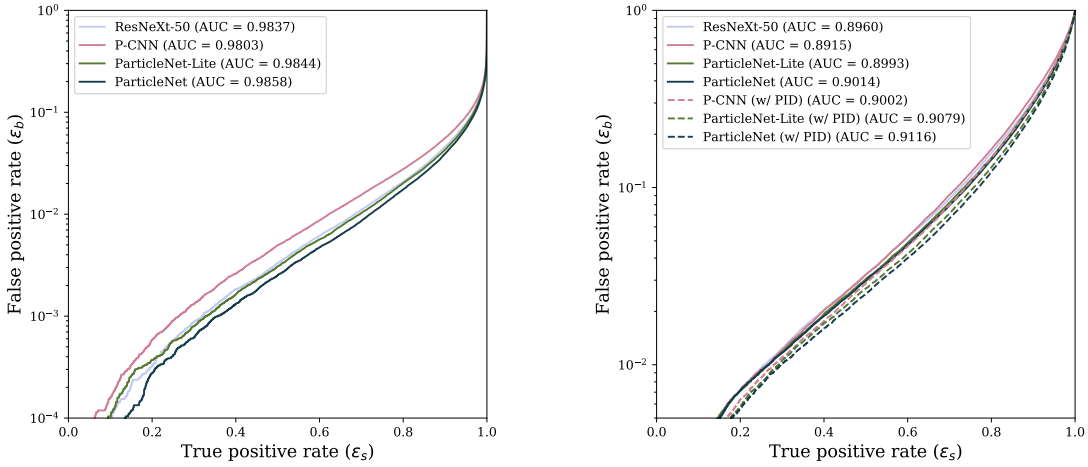
	Accuracy	AUC	$1/\varepsilon_b$ at $\varepsilon_s = 50\%$	$1/\varepsilon_b$ at $\varepsilon_s = 30\%$
ResNeXt-50	0.936	0.9837	302 ± 5	1147 ± 58
P-CNN	0.930	0.9803	201 ± 4	759 ± 24
PFN	-	0.9819	247 ± 3	888 ± 17
ParticleNet-Lite	0.937	0.9844	325 ± 5	1262 ± 49
ParticleNet	**0.940**	**0.9858**	$\mathbf{397 \pm 7}$	$\mathbf{1615 \pm 93}$

Table 6.2: Performance comparison on the top tagging benchmark dataset. The ParticleNet, ParticleNet-Lite, P-CNN and ResNeXt-50 models are trained on the top tagging dataset starting from randomly initialized weights. A total of 9 independent trainings are performed. The table shows the result from the median-accuracy training, and the standard deviation of the 9 trainings is quoted as the uncertainty. Uncertainty on the accuracy and AUC are negligible and therefore omitted. The performance of PFN on this dataset is reported in [126], and the uncertainty corresponds to the spread in 10 trainings.

benchmark dataset and improves over previous methods significantly. Its background rejection power at 30% signal efficiency is roughly 1.8 (2.1) times as good as PFN (P-CNN), and about 40% better than ResNeXt-50. Even the ParticleNet-Lite model, with significantly reduced complexity, outperforms all the previous models, achieving about 10% improvement with respect to ResNeXt-50. The large performance improvement of the ParticleNet architecture over the PFN architecture is likely due to better exploitation of the local neighborhood information with the EdgeConv operation.

6.3.2 Quark-Gluon Tagging

Another important jet tagging task is quark-gluon tagging, i.e., discriminating jets initiated by quarks and by gluons. The quark-gluon tagging dataset from [126] is used to evaluate the performance of the ParticleNet architecture on this task. The signal (quark) and background (gluon) jets are generated with Pythia8 using the $Z(\to \nu\nu) + (u, d, s)$ and $Z(\to \nu\nu) + g$ processes, respectively. No detector simulation is performed. The final state non-neutrino particles are clustered into jets using the anti-k_T algorithm [40] with $R = 0.4$. Only jets with transverse momentum $p_T \in [500, 550]$ and rapidity $|y| < 2$ are



(a) (b)

Figure 6.2: Performance comparison in terms of ROC curves on (a) the top tagging benchmark dataset and (b) the quark-gluon tagging benchmark dataset.

considered. This dataset consists of 2 million jets in total, half signal and half background. We follow the recommended splitting of 1.6M/200k/200k for training, validation and testing in the development of the ParticleNet model on this dataset.

One important difference of the quark-gluon tagging dataset is that it includes not only the four-momentum, but also the type of each particle (i.e., electron, photon, pion, etc.). Such particle identification (PID) information can be quite helpful for jet tagging. Therefore, we include this information in the ParticleNet model and compare it with the baseline version using only the kinematic information. The PID information is included in an experimentally realistic way by using only five particle types (electron, muon, charged hadron, neutral hadron and photon), as well as the electric charge, as inputs. These 6 additional variables, together with the 7 kinematic variables, form the input feature vector of each particle for models with PID information, as shown in Table 6.1.

Table 6.2 compares the performance of the ParticleNet model with a number of alternative models introduced in Section 6.3.1. Model variants with and without PID inputs

	Accuracy	AUC	$1/\varepsilon_b$ at $\varepsilon_s = 50\%$	$1/\varepsilon_b$ at $\varepsilon_s = 30\%$
ResNeXt-50	0.821	0.8960	30.9	80.8
P-CNN	0.818	0.8915	31.0	82.3
PFN	-	0.8911	30.8 ± 0.4	-
ParticleNet-Lite	0.826	0.8993	32.8	84.6
ParticleNet	0.828	0.9014	33.7	85.4
P-CNN (w/ PID)	0.827	0.9002	34.7	91.0
PFN-Ex (w/ PID)	-	0.9005	34.7 ± 0.4	-
ParticleNet-Lite (w/ PID)	0.835	0.9079	37.1	94.5
ParticleNet (w/ PID)	**0.840**	**0.9116**	$\mathbf{39.8 \pm 0.2}$	$\mathbf{98.6 \pm 1.3}$

Table 6.3: Performance comparison on the quark-gluon tagging benchmark dataset. The ParticleNet, ParticleNet-Lite, P-CNN and ResNeXt-50 models are trained on the quark-gluon tagging dataset starting from randomly initialized weights. A total of 9 independent trainings are performed for the ParticleNet model, and the table shows the result from the median-accuracy training, with the standard deviation of the 9 trainings quoted as the uncertainty. Due to limited computational resources, the training of other models is performed only once, but the uncertainty due to random weight initialization is expected to be fairly small. The performance of PFN on this dataset is reported in [126], and the uncertainty corresponds to the spread in ten trainings. Note that a number of PFN models with different levels of PID information are investigated in [126], and "PFN-Ex", also using experimentally realistic PID information, is shown here for comparison.

are also compared. Note that for the ResNeXt-50 model, only the version without PID inputs is presented, as it is based on jet images that cannot incorporate PID information straightforwardly. The corresponding ROC curves are shown in Figure 6.2b. Overall, the addition of PID inputs has a large impact on the performance, increasing the background rejection power by 10–15% compared to the same model without using PID information. This clearly demonstrates the advantage of particle-based jet representations, including the particle cloud representation, as they can easily integrate any additional information for each particle. The best performance is obtained by the ParticleNet model with PID inputs, achieving almost 15% improvement on the background rejection power compared to the PFN-Ex and P-CNN models. The ParticleNet-Lite model achieves the second-best performance and shows about 7% improvement with respect to the PFN-Ex and P-CNN models.

6.4 Model Complexity

Another aspect of machine learning models is the complexity, e.g., the number of parameters and the computational cost. Table 6.4 compares the number of parameters and the computational cost of all the models used in the top tagging task in Section 6.3.1. The computational cost is evaluated using the inference time per object, which is a more relevant metric than the training time for real-life applications of machine learning models. The inference time of each model is measured on both the CPU and the GPU, using the implementations with Apache MXNet. For the CPU, to mimic the event processing workflow typically used in collider experiments, a batch size of 1 is used, and the inference is performed in single-thread mode. For the GPU, a batch size of 100 is used instead, as the full power of the GPU cannot be revealed with a very small batch size (e.g., 1) due to the overhead in data transfer between the CPU and the GPU. The ParticleNet model achieves the best classification performance at the cost of speed, being more than an order of magnitude slower than the PFN and the P-CNN models, but still, it is not prohibitively slow even on the CPU. In addition, the current implementation of the EdgeConv operation used in the ParticleNet model is not as optimized as the regular convolution operation. Therefore, further speed-up is expected from an optimized implementation of EdgeConv. On the other hand, the ParticleNet-Lite model provides a good balance between speed and performance, showing more than 40% improvement in performance while being only a few times slower than the PFN and P-CNN models. Notably, it is also the most economical model, outperforming all previous approaches with only 26k parameters, thanks to the effective exploitation of the permutation symmetry of the particle clouds. Overall, PFN is the fastest model on both the CPU and the GPU, making it a suitable choice for extremely time critical tasks.

	Parameters	Time (CPU) [ms]	Time (GPU) [ms]	$1/\varepsilon_b$ at $\varepsilon_s = 30\%$
ResNeXt-50	1.46M	7.4	0.22	1147 ± 58
P-CNN	348k	1.6	0.020	759 ± 24
PFN	82k	**0.8**	**0.018**	888 ± 17
ParticleNet-Lite	**26k**	2.4	0.084	1262 ± 49
ParticleNet	366k	23	0.92	$\mathbf{1615 \pm 93}$

Table 6.4: Number of parameters, inference time per object, and background rejection of different models. The CPU inference time is measured on an Intel Core i7-6850K CPU with a single thread using a batch size of 1. The GPU inference time is measured on a Nvidia GTX 1080 Ti GPU using a batch size of 100.

6.5 Conclusion

In this chapter, we presented a new approach for machine learning on jets. The core of this approach is to treat jets as particle clouds, i.e., unordered sets of particles. Based on this particle cloud representation, we introduce ParticleNet, a network architecture tailored to jet tagging tasks. The ParticleNet architecture achieves state-of-the-art performance on top tagging and quark-gluon tagging benchmarks and improves significantly over existing methods.

While we only demonstrate the power of the particle cloud representation in jet tagging tasks, we think that it is a natural and generic way of representing jets (and even the whole collision event) and can be applied to a broad range of particle physics problems. Applications of the particle cloud approach to, e.g., pileup identification, jet grooming, jet energy calibration, etc., would be particularly interesting and worth further investigation.

Part III

Search for Lorentz-Boosted Higgs Bosons Decaying to Charm Quarks

The search for the decay of the Higgs boson to charm quarks is crucial in understanding the structure of Yukawa couplings, as it provides a direct probe of the Higgs coupling to second-generation quarks. However, such a search is extremely challenging at the LHC due to the overwhelming jet production background and requires dedicated methods for object reconstruction, e.g., the identification of the charm quarks. This part of the thesis presents a search for the Higgs boson decaying to charm quarks using proton-proton collision data collected by the CMS experiment in 2016. A new approach to reconstruct the Higgs boson is developed, in which both quarks from the Higgs decay are reconstructed with a single large-radius jet and the DNN-based DeepAK8 algorithm, presented in Chapter 5, is adapted to identify the charm quark pair. This new approach significantly improves the signal purity, leading to highly competitive results.

This part is organized as follows: In Chapter 7, we briefly discuss the motivation for this search. Chapter 8 presents the strategy and the full procedure of this search, including a description of the data and simulated samples, the definition of the physics objects, reconstruction of the Higgs boson, the event selection, the background estimation and the signal extraction method, as well as the effects of systematic uncertainties. The results of the search are summarized in Chapter 9.

Chapter 7

Introduction and Motivation

The discovery of the Higgs boson [136, 137] at the LHC has opened a new chapter in the history of particle physics. Investigating the properties of the discovered Higgs boson has become one of the top priorities of the LHC physics program, as every measurement has to be carefully compared with the SM prediction, and any deviation, if observed, would be a clear sign of new physics. So far, the newly discovered Higgs boson has been extensively studied at the LHC using proton-proton collision data at $\sqrt{s} = 7$, 8, and 13 TeV. The mass of the Higgs boson is precisely measured using the $\mathrm{H} \to \gamma\gamma$ and $\mathrm{H} \to \mathrm{ZZ}^* \to 4\ell$ channels, yield a result of $m_{\mathrm{H}} = 125.26 \pm 0.21\,\mathrm{GeV}$ by the CMS experiment [138], and a compatible result of $m_{\mathrm{H}} = 124.97 \pm 0.24\,\mathrm{GeV}$ by the ATLAS experiment [139]. The couplings of the Higgs boson to vector bosons has been firmly established already in Run 1 of the LHC by the observation of the Higgs boson decaying to $\gamma\gamma$, ZZ and WW [140, 141, 142, 143, 144, 145, 146]. As for the Yukawa couplings, so far only couplings to the third-generation fermions have been directly observed. The decay to τ leptons, $\mathrm{H} \to \tau^+\tau^-$, was first observed using a combination of the ATLAS and CMS data in Run 1 [147]. Direct evidence of the Higgs coupling to the top quark is established by the observation of the $\mathrm{t\bar{t}H}$ production by the CMS [148] and the ATLAS

[149] Collaborations using more data collected in Run 2. Most recently, the decay of the Higgs boson to bottom quarks, $H \to b\bar{b}$, has also been observed [150, 151]. So far, all measured properties are consistent with the expectations of the standard model within measurement uncertainties.

Measurement of the Higgs couplings to second-generation fermions is clearly the next milestone. However, due to the much smaller couplings compared to third generation fermions, such measurements are extremely challenging. Searches for the decay of the Higgs boson to a pair of muons, $H \to \mu^+\mu^-$, have been carried out by both the ATLAS and CMS Collaborations. Using Run 2 data corresponding to an integrated luminosity of $139\,\text{fb}^{-1}$ ($35.9\,\text{fb}^{-1}$), the ATLAS [152] (CMS [153]) search reported an observed upper limit on the cross section times branching fraction, $\sigma(pp \to H) \times \mathcal{B}(H \to \mu^+\mu^-)$, of 1.7 (3.0) times of the SM expectation at 95% confidence level (CL). Substantially more data is needed to establish evidence for the muon Yukawa coupling. On the other hand, couplings to second-generation quarks are even more difficult to probe at the LHC due to the overwhelming background. The possibility of probing the Yukawa coupling to the charm quark (y_c) from a global fit to the Higgs data has been discussed in Refs. [154, 155], and an upper bound on y_c/y_c^{SM} of 6.2 at 95% CL is obtained in [155]. However, some assumptions are made in this estimation, particularly, the LEP constraints [156] are taken into account which carry some model dependence. Another approach to probe the charm quark Yukawa coupling is via the decays of the Higgs boson to a final state with charmonium, $H \to J/\psi\gamma$ [157]. Searches for $H \to J/\psi\gamma$ have been performed by the ATLAS and CMS Collaborations [158, 159, 160, 161], and the current upper limit on $\sigma(pp \to H) \times \mathcal{B}(H \to J/\psi\gamma)$ is around 100–200 times the SM expectation at 95% CL. The search for the decay of the Higgs boson to a charm quark-antiquark pair, $H \to c\bar{c}$, provides the most direct probe of the charm quark Yukawa coupling, but the reconstruction and identification of the charms quarks poses significant experimental challenges. A direct

search for $H \to c\bar{c}$ in the $ZH(Z \to \ell\ell)$ channel ($\ell = e, \mu$) has been performed by the ATLAS Collaboration [162]. Using $36.1\,\text{fb}^{-1}$ of data, an observed (expected) upper limit on $\sigma(pp \to ZH) \times \mathcal{B}(H \to c\bar{c})$ at 95% CL of 110 (150) times the SM expectation is obtained.

Recently, the CMS Collaboration has carried out the first direct search for $H \to c\bar{c}$ [163], using proton-proton collision data collected in 2016 corresponding to an integrated luminosity of $35.9\,\text{fb}^{-1}$. The search targets the Higgs boson produced in association with a vector (V) boson, i.e., a W or Z boson. The V boson is required to decay into leptons, i.e., $W \to \ell\nu$, $Z \to \ell\ell$ or $Z \to \nu\nu$. The presence of a charged lepton, or large $p_{\text{T}}^{\text{miss}}$ from the neutrinos, provides a clear signature to trigger the event efficiently and suppress the overwhelming QCD multijet background significantly. To fully explore the kinematic topology of the $H \to c\bar{c}$ decay, two approaches for the reconstruction of the $H \to c\bar{c}$ candidate are adopted. The traditional "resolved-jet" approach uses two well-separated and individually resolved jets to reconstructed the Higgs candidate. The "merged-jet" approach reconstructs both charm quarks from the Higgs decay with a single, large-radius jet. The former approach works well for Higgs candidates with lower p_{T}, while the latter performs better at higher p_{T}, where the two quarks from the Higgs decay become more collimated. The merged-jet approach is presented in detail in this thesis.

Chapter 8

Analysis Method

8.1 Data and Simulated Events

This analysis uses proton-proton collision data collected by the CMS detector during the 2016 run of the LHC at a center-of-mass energy of 13 TeV, corresponding to an integrated luminosity of $35.9\,\mathrm{fb}^{-1}$. Events in the 0-lepton channel are collected with a $p_{\mathrm{T}}^{\mathrm{miss}}$ trigger, while in the 1- and 2-lepton channels, events are collected with a trigger that requires the presence of an isolated electron or muon.

Simulated events are also used in this analysis to study the characteristics of signal and background processes and design the overall search strategy. The quark-induced ZH and WH signal processes are generated at next-to-leading order (NLO) accuracy in QCD using the POWHEG v2 [164, 165, 166] event generator extended with the MiNLO procedure [167, 168], while the gluon-induced ZH process is generated at leading order (LO) accuracy with POWHEG v2. The mass of the Higgs boson is set to 125 GeV in all signal samples. The production cross sections of the signal processes are corrected as a function of $p_{\mathrm{T}}(\mathrm{V})$ to next-to-next-to-leading order (NNLO) QCD + NLO EW accuracy combining the VHNNLO [169, 170, 171, 172], VH@NNLO [173, 174], and HAWK v2.0 [175]

generators as described in Ref. [15]. The cross sections of the signal processes are summarized in Table 8.1.

Process	Cross section [pb]
$qq \to W^+(\to \ell^+\nu)H$	0.2828
$qq \to W^-(\to \ell^-\nu)H$	0.1795
$qq \to Z(\to \ell\ell)H$	0.0770
$gg \to Z(\to \ell\ell)H$	0.0124
$qq \to Z(\to \nu\nu)H$	0.1530
$gg \to Z(\to \nu\nu)H$	0.0246

Table 8.1: List of the simulated VH signal samples and the corresponding cross sections.

The V+jets background samples are generated using MadGraph5_aMC@NLO v2.4.2 [81] at NLO QCD accuracy with up to two additional partons using the FxFx merging scheme [176]. The production cross sections are scaled to the NNLO prediction obtained from FEWZ 3.1 [177]. Events are reweighted as a function of $p_{\mathrm{T}}(\mathrm{V})$ to account for NLO EW corrections [178].

The $t\bar{t}$ process [179] is generated at NLO accuracy with POWHEG v2. The production cross sections for the $t\bar{t}$ samples are scaled to the NNLO prediction with the next-to-next-to-leading-log (NNLL) soft-gluon resummation obtained from Top++ v2.0 [180]. The $t\bar{t}$ samples are reweighted as a function of top quark p_{T} to account for known differences between data and simulation [181].

The single-top process in the t-channel [182, 183] (tW-channel [184]) is generated at NLO accuracy with POWHEG v2 (v1), while the s-channel single-top process is generated with MadGraph5_aMC@NLO v2.4.2. The WW process [185] is generated at NLO accuracy with POWHEG v2. The WZ and ZZ samples are generated with MadGraph5_aMC@NLO v2.4.2 at NLO with the FxFx merging scheme. The same generator is also used at LO accuracy with the MLM matching scheme [186] to generate a sample of QCD multijet events.

The NNPDF3.0 [83] parton distribution functions (PDF) are used to produce all simulated samples, with the NLO (LO) PDF set used for the NLO (LO) samples. In all cases, parton showering and hadronization are simulated in PYTHIA v8.212 [82]. A GEANT4-based model [86] is used to simulate the response of the CMS detector. For all samples, additional pp interactions in the same or neighboring bunch crossings (referred to as "pileup") are simulated and added to the hard-scattering process. The events are then reweighted to match the pileup profile observed in the collected data.

8.2 Object Selection

8.2.1 Charged Leptons

In this analysis, two levels of identification criteria are defined for charged leptons (electrons and muons). The baseline criterion is used to count the lepton multiplicity and categorize events into the mutually exclusive 0-, 1-, and 2-lepton channels, while more stringent requirements are imposed in the 1- and 2-lepton channels to select high-purity samples of isolated leptons. In this section, we introduce the baseline criterion, and the channel-specific requirements are described in Section 8.4.1 and 8.4.2.

Electrons

Reconstructed electron candidates are required to have $p_{\mathrm{T}} > 7\,\mathrm{GeV}$ and $|\eta| < 2.4$, and have a transverse (longitudinal) impact parameter $|d_{xy}| < 0.05\,\mathrm{cm}$ ($|d_z| < 0.2\,\mathrm{cm}$) with respect to the primary vertex. Additional identification criteria based on a multivariate method are imposed to suppress background, using a working point that corresponds to a signal efficiency of $\sim 90\%$ for prompt electrons. In addition, the electrons are required to be isolated from any additional nearby activity in the detector. The PF isolation [187]

is used, which is defined as

$$\mathrm{Iso_{PF}} = \sum p_{\mathrm{T}}^{\mathrm{charged}} + \max\left(0, \sum p_{\mathrm{T}}^{\mathrm{neutral\ hadron}} + \sum p_{\mathrm{T}}^{\gamma} - p_{\mathrm{T}}^{\mathrm{PU}}\right), \tag{8.1}$$

where the sums run over charged PF candidates, neutral hadrons and photons, within a cone of $\Delta R = 0.3$ around the electron direction. To avoid contamination due to pileup, only charged PF candidates originating from the primary vertex are included in the sum, and contribution from neural particles from pileup are corrected by the $p_{\mathrm{T}}^{\mathrm{PU}}$ term, which is estimated from the event pileup energy density multiplied by an "effective" area that takes into account the dependence on pseudorapidity. In this analysis, electron candidates are required to have a relative isolation, defined as

$$\mathrm{Iso_{rel}} = \frac{\mathrm{Iso_{PF}}}{p_{\mathrm{T}}}, \tag{8.2}$$

smaller than 0.4.

Muons

Reconstructed muon candidates are required to have $p_{\mathrm{T}} > 5\,\mathrm{GeV}$, $|\eta| < 2.4$, $|d_{xy}| < 0.5\,\mathrm{cm}$ and $|d_z| < 1.0\,\mathrm{cm}$. In addition, we require the muon candidates to pass the loose muon identification as described in [95]. Similar to electrons, the muons are also required to be isolated. The PF isolation (eq. (8.1)) computation for muons is slightly different from the electrons – it uses a larger cone of $\Delta R = 0.4$, and the $p_{\mathrm{T}}^{\mathrm{PU}}$ term is estimated as half of the sum of charged hadron deposits originating from pileup vertices. In this analysis, muon candidates are required to have a relative isolation $\mathrm{Iso_{rel}} < 0.4$.

8.2.2 Jets

Two types of jets are used in this analysis and are described in this section.

Large-R jets

The primary jet collection used in this analysis is obtained by clustering the PF candidates with the anti-k_{T} algorithm using a fairly large distance parameter $R = 1.5$. This jet collection will be referred to as "large-R jets" in what follows. The PUPPI algorithm [43] is used to correct for pileup. The "modified mass drop tagger" algorithm [55], also known as the "soft drop" (SD) algorithm [56], with angular exponent $\beta = 0$, soft cutoff threshold $z_{\mathrm{cut}} < 0.1$, and characteristic radius $R_0 = 1.5$, is applied to remove soft and wide-angle radiation from the jet. In the default configuration, the SD algorithm identifies two hard subjets of the large-R jet. We use the kinematics of these two subjets to calculate the 4-momentum of the large-R jet. The jet energy scale and resolution corrections derived for PUPPI jets clustered with $R = 0.4$ are applied on the subjets. The large-R jets are required to have $p_{\mathrm{T}} > 200\,\mathrm{GeV}$, $|\eta| < 2.4$ and satisfy the tight PF jet identification criteria [188]. The soft drop mass (m_{SD}), computed as the invariant mass of the two subjets, is required to be between 50 and 200 GeV. To avoid misidentifying leptons as jets, any large-R jets overlapping with an electron or muon with $\Delta R(\text{large-}R\text{ jet}, \ell) < 1.5$ are not considered in this analysis.

Small-R jets

An additional jet collection is used in the event selection. These jets, which will be referred to as "small-R jets", are formed by clustering PF candidates using the anti-k_{T} algorithm with $R = 0.4$. Charged PF candidates associated with pileup vertices are removed from the jet constituents using the charged hadron subtraction algorithm [189].

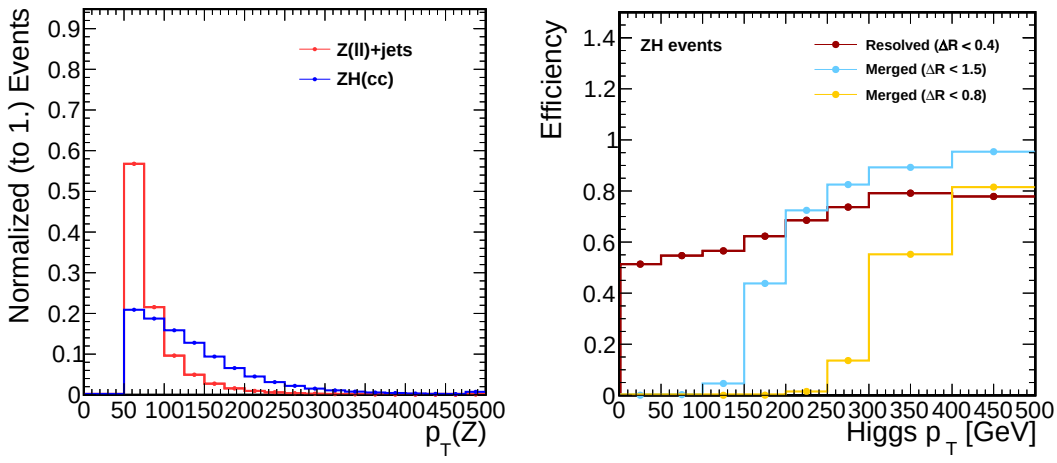


Figure 8.1: Left: The $p_{\mathrm{T}}(\mathrm{V})$ spectrum in the VH signal (blue line) and in the V+jets background (red line) events. Right: Efficiency for both quarks from the Higgs decay to be contained in a single jet clustered with different R ($R = 0.8$ or $R = 1.5$) as a function of $p_{\mathrm{T}}(\mathrm{H})$. For comparison, the corresponding efficiency when the two quarks from the Higgs decay are resolved into two jets clustered with $R = 0.4$ is also shown.

The jet energy scale and resolution are corrected following the method in [42]. The small-R jets are required to have $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ and $|\eta| < 2.4$, and satisfy the tight PF jet identification criteria [188]. To avoid misidentifying leptons as jets, small-R jets overlapping with an electron or muon with $\Delta R(\text{small-}R\text{ jet}, \ell) < 0.4$ are discarded.

8.3 Higgs Boson Reconstruction

Traditionally, when a Higgs boson decays to a pair of quarks, the Higgs boson is reconstructed using two separate small-R jets ("resolved jets"), each corresponding to one quark from the decay. However, in this analysis, the Higgs boson is reconstructed with a single, large-R jet ("merged jet"). This approach targets mainly the Lorentz-boosted Higgs bosons and has a number of advantages over the resolved-jet approach. Firstly, the boosted regime has better signal purity, though at the cost of a significantly lower signal acceptance. This is demonstrated in Figure 8.1 (left) which shows that the

$p_{\mathrm{T}}(\mathrm{V})$ spectrum in the VH signal is much harder than that in the V+jets background. Therefore, the boosted regime (e.g., $p_{\mathrm{T}}(\mathrm{V}) \gtrsim 200\,\mathrm{GeV}$) provides a much higher reduction of background events than signal events. Secondly, the use of a large-R jet is capable of capturing both quarks from the Higgs decay (and potentially also the final state radiation emitted by the quarks) in one jet, which allows for better exploitation of the correlation between the two quarks, and helps avoid combinatorial backgrounds that can arise in the resolved-jet approach. A more detailed discussion on the advantages of the merged-jet approach can be found in the literature (e.g., Refs. [190, 191]).

On the other hand, the acceptance of both the VH signal and the V+jets background drops very rapidly as a function of $p_{\mathrm{T}}(\mathrm{V})$. No more than ∼5% of the signal events survive a selection of $p_{\mathrm{T}}(\mathrm{V}) \gtrsim 200\,\mathrm{GeV}$. Therefore, it is crucial to strike a balance between signal purity and acceptance. In the case of the merged-jet approach, this can be achieved by choosing a suitable distance parameter R of the jet clustering algorithm. In general, when a particle (e.g., the Higgs boson) with a high Lorentz-boost undergoes a two-body decay, the angular separation between the two decay products is approximately $\Delta R \sim 2m/p_{\mathrm{T}}$. For a Higgs boson with $p_{\mathrm{T}}(\mathrm{H}) \simeq 200\,\mathrm{GeV}$, this gives $\Delta R \simeq 1.25$. Figure 8.1 (right) displays the efficiency for both quarks from the Higgs decay to be contained in a single jet clustered with the anti-k_{T} algorithm using different distance parameters ($R = 0.8$ or $R = 1.5$), as a function of $p_{\mathrm{T}}(\mathrm{H})$. For comparison, we also include the corresponding efficiency when the two quarks from the Higgs decay can be resolved into two jets clustered with the anti-k_{T} algorithm using $R = 0.4$. As expected, in the low $p_{\mathrm{T}}(\mathrm{H})$ regime ($p_{\mathrm{T}}(\mathrm{H}) \lesssim 100\,\mathrm{GeV}$), the Higgs boson can only be reconstructed with two resolved jets. However, as $p_{\mathrm{T}}(\mathrm{H})$ increases, the merged-jet approach shows higher efficiency, getting comparable or even better than the resolved-jet approach at $p_{\mathrm{T}}(\mathrm{H}) \simeq 200\ (400)\,\mathrm{GeV}$ for jets clustered with $R = 1.5$ (0.8). Based on these observations, jets clustered with $R = 1.5$ are used in this analysis to reconstruct the Higgs boson.

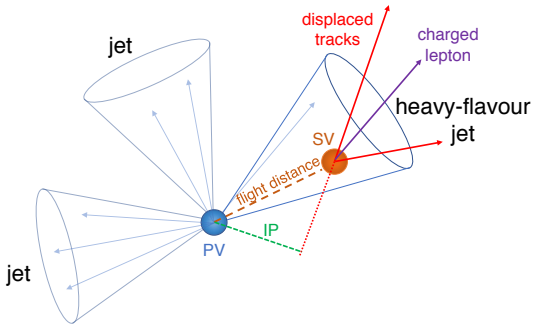


Figure 8.2: Illustration of the characteristics of a c-jet: the presence of displaced tracks, secondary vertices and charged leptons. Image from Ref. [79].

One of the biggest challenges for this analysis is to efficiently reconstruct and identify the charm quark pair from the Higgs decay, while rejecting background jets arising from light quarks (u, d, s) and gluons, as well as from b quarks. The identification of jets arising from c quarks ("c-jets") mainly relies on the long lifetime ($\lesssim 1\,\text{ps}$) of the hadrons containing c quarks, which leads to typical displacements of a few mm to one cm, depending on their momenta. These c-hadrons leave traces in the detector such as displaced tracks, from which a secondary vertex may be reconstructed. In addition, a muon or electron is present in the decay chain of a c-hadron in about 10% of the cases, which can also be exploited for separating c-jets from light-flavor jets. These characteristics of a c-jet are illustrated in Figure 8.2. However, the identification of c-jets is particularly challenging, as these characteristics of c-jets are also present in b-jets, and the properties of the c-jets are actually in between light-flavor jets and b-jets. For example, one of the most important handles for identifying c-jets, i.e., the lifetime of the c-hadrons, though much longer than the light-flavor hadrons, is close to that of b-hadrons ($\simeq 1.5\,\text{ps}$). This makes it particularly challenging to identify c-jets efficiently while rejecting both light-jets and b-jets simultaneously.

In this analysis, advanced machine learning techniques are exploited to tackle this challenge. Specifically, we attempt to identify both charm quarks from the Higgs decay

simultaneously using a large-R jet. This approach has the advantage of being able to exploit the correlation between the decay products of the Higgs boson, thereby improving the performance. The DNN-based identification algorithm, DeepAK8, described in detail in Section 5, is adapted to the identification of charm quark pairs by performing a dedicated training with $R = 1.5$ jets. The mass-decorrelated version is used in this analysis. This adapted version of the DeepAK8 algorithm will be referred to as "DeepAK15" hereafter.

Due to the multi-class nature of the DeepAK15 algorithm (see Section 5.1 for all the output classes), a number of options exist for how to use it for identifying charm quark pairs. The most obvious one is,

$$\text{score}(\text{H} \to \text{c}\overline{\text{c}}) + \text{score}(\text{Z} \to \text{c}\overline{\text{c}}), \tag{8.3}$$

which considers $\text{H} \to \text{c}\overline{\text{c}}$ and $\text{Z} \to \text{c}\overline{\text{c}}$ as signal and all the other classes as background, including not only jets from QCD radiation, but also e.g., jets from hadronic decays of top quarks, W bosons and Z bosons (except for $\text{Z} \to \text{c}\overline{\text{c}}$). This definition provides better sensitivity in the 1-lepton channel where $\text{t}\overline{\text{t}}$ events constitute a large fraction of the background, due to its explicit suppression of jets arising from real hadronic top quarks or W bosons. Another option is,

$$\frac{\text{score}(\text{H} \to \text{c}\overline{\text{c}}) + \text{score}(\text{Z} \to \text{c}\overline{\text{c}})}{\text{score}(\text{H} \to \text{c}\overline{\text{c}}) + \text{score}(\text{Z} \to \text{c}\overline{\text{c}}) + \text{score}(\text{H} \to \text{b}\overline{\text{b}}) + \text{score}(\text{Z} \to \text{b}\overline{\text{b}}) + \text{score}(\text{QCD})}, \tag{8.4}$$

which considers only jets from QCD radiation and from b quark pairs ($\text{H} \to \text{b}\overline{\text{b}}$ and $\text{Z} \to \text{b}\overline{\text{b}}$) as background and "renormalizes" the scores to reflect that. This definition does not attempt to suppress jets from top quarks or W bosons explicitly but rejects jets

from b quark pairs. Similarly, a third way is,

$$\frac{\text{score}(\text{H} \to \text{c}\bar{\text{c}}) + \text{score}(\text{Z} \to \text{c}\bar{\text{c}})}{\text{score}(\text{H} \to \text{c}\bar{\text{c}}) + \text{score}(\text{Z} \to \text{c}\bar{\text{c}}) + \text{score}(\text{QCD})}, \tag{8.5}$$

which essentially considers only jets from QCD radiation as background, and ignores the outputs for other classes. Overall, the definition in eq. (8.5) results in the best sensitivity, especially in the 2-lepton channel where the background is predominantly Z+jets with little contribution from $\text{t}\bar{\text{t}}$. The definition in eq. (8.3) performs much worse in the 2-lepton channel, as the suppression of top and W comes at the cost of signal efficiency, especially the $\text{W} \to \text{cq}$ decay which is very similar to the $\text{Z(H)} \to \text{c}\bar{\text{c}}$ decay. On the other hand, the definition in eq. (8.4) provides a much better suppression of background jets arising from b-quark pairs, especially the largely irreducible $\text{H} \to \text{b}\bar{\text{b}}$ background, but hurts the $\text{Z(H)} \to \text{c}\bar{\text{c}}$ efficiency due to the similarity between b and c quarks. In the end, since light-flavor jets constitute a much larger fraction of the background, a much tighter selection against b-quark pairs does not result in better overall sensitivity. Therefore, the $\text{c}\bar{\text{c}}$-tagging discriminant as defined in eq. (8.5) is used throughout this analysis.

The performance of DeepAK15 for identifying a pair of c quarks using the $\text{c}\bar{\text{c}}$-tagging discriminant defined in eq. (8.5) is displayed in Figure 8.3 in terms of receiver operating characteristic (ROC) curves, for large-R jets with $p_{\text{T}} > 200\,\text{GeV}$ and $m_{\text{SD}} > 50\,\text{GeV}$. Figure 8.3 (left) shows the efficiency of identifying a c quark pair from the Higgs decay and the misidentification rate for jets stemming from QCD radiation in the V+jets process. Three working points are defined with 1%, 2.5% and 5% misidentification rates, and the corresponding efficiencies for identifying a c quark pair are $\simeq 23\%$, $\simeq 35\%$, and $\simeq 46\%$. Another important aspect is the misidentification of jets arising from a pair of b quarks as the signal. The corresponding ROC curve is displayed in Figure 8.3 (right). For the three working points defined above, the corresponding misidentification rates for

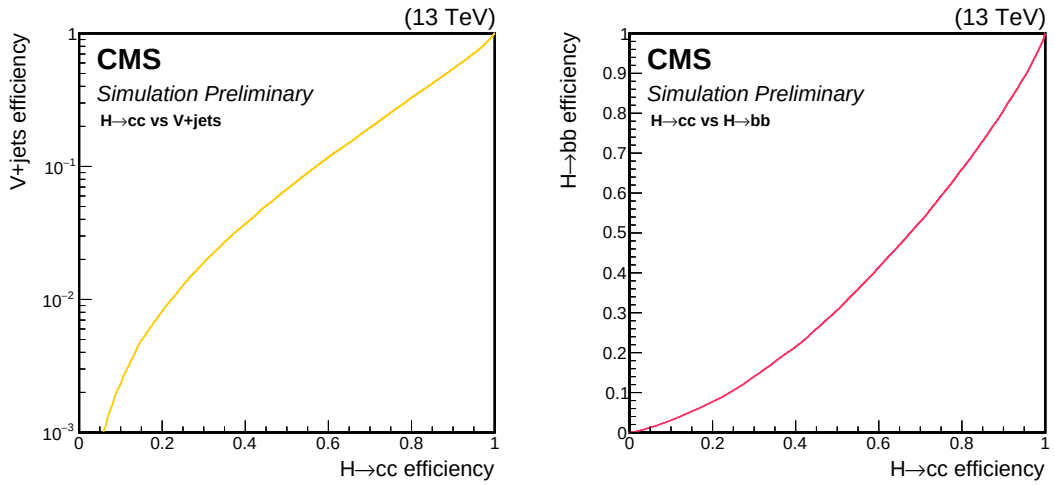


Figure 8.3: The performance of DeepAK15 for identifying a $c\bar{c}$ pair using the $c\bar{c}$-tagging discriminant defined in eq. (8.5) in terms of receiver operating characteristic (ROC) curves, for large-R jets with $p_T > 200\,\text{GeV}$ and $m_{SD} > 50\,\text{GeV}$. Left: the efficiency of correctly identifying a pair of c quarks from Higgs decay vs the efficiency of misidentifying jets from the V+jets process; Right: the efficiency of correctly identifying a pair of c quarks from Higgs decay vs the efficiency of misidentifying a pair of b quarks from Higgs decay.

Category	$c\bar{c}$-tagging discriminant
High purity (HP)	> 0.91
Medium purity (MP)	$(0.83, 0.91]$
Low purity (LP)	$(0.72, 0.83]$

Table 8.2: Definition of the three mutually exclusive categories based on the $c\bar{c}$-tagging discriminant.

$H \to b\bar{b}$ are $\simeq 9\%$, $\simeq 17\%$, and $\simeq 27\%$. To improve the sensitivity of this analysis, three mutually exclusive categories are defined based on the three working points, and the requirements on the $c\bar{c}$-tagging discriminant are listed in Table 8.2.

The DeepAK15 $c\bar{c}$-tagging discriminant is calibrated using data and simulated samples. Ideally, a sample with a high purity of $H \to c\bar{c}$ events should be used to measure the efficiency of the DeepAK15 $c\bar{c}$-tagging discriminant and correct for the difference between data and simulated samples. However, such a high purity $H \to c\bar{c}$ sample is impossible to obtain with the current LHC data. Therefore, an alternative sample enriched

in jets arising from gluon splitting to $c\bar{c}$ is used. In addition, we require the presence of at least one secondary vertex in each of the two subjets of the large-R jet, to make the selected $g \to c\bar{c}$ jets better resemble the signal jets, and also to enhance the overall purity of the $g \to c\bar{c}$ jets in this sample. Using a template fit method, p_T-dependent simulation-to-data scale factors are extracted from this sample. The typical values of the scale factors range from 0.85 to 1.3, with a relative uncertainty of about 15–20%. These scale factors are applied only to the simulated $VH(H \to c\bar{c})$ (and $VZ(Z \to c\bar{c})$) events as will be explained later in Section 8.5.

8.4 Event Selection

A typical VH signal event has a clear signature of a vector boson (W or Z) recoiling against a Higgs boson, with little additional activity in the event. In this analysis, we exploit the leptonic decays of the vector boson, i.e., $W \to \ell\nu$, $Z \to \ell\ell$ or $Z \to \nu\nu$, where ℓ is an electron or muon. By selecting events with one or more charged leptons, or with high missing transverse momentum originating from the neutrinos, the large QCD multijet background can be effectively suppressed. Three analysis channels, the 2-lepton channel, the 1-lepton channel, and the 0-lepton channel are defined in this analysis based on the number of identified charged leptons. The characteristics of the signal and background processes, as well as the detailed event selection criteria, are described separately for each channel in Sections 8.4.1 to 8.4.3. The event selection criteria are summarized in Table 8.4 in Section 8.4.4.

8.4.1 The 2-Lepton Channel

The 2-lepton channel targets ZH events with the Z boson decaying into a pair of electrons or muons. Events are selected with at least two electrons or muons passing the

loose lepton selection criteria described in Section 8.2.1. The leptons are further required to be isolated, with a relative PF isolation below 0.25 (0.15) for muons (electrons). The p_{T} of the leptons are required to be above 20 GeV such that the trigger becomes fully efficient. The two leptons leading in p_{T} are required to be compatible with a Z boson decay, i.e., they must have opposite electric charges and the same flavor (i.e., both are muons or electrons), and the invariant mass of the lepton pair, $m(\ell\ell)$, must be compatible with the mass of the Z boson, with $75 < m(\ell\ell) < 105\,\mathrm{GeV}$. The Z boson is then reconstructed from these two leptons, with the four-momentum of the Z boson computed by summing up the four-momenta of the two leptons.

Since this analysis targets the boosted regime of the VH production, only events with $p_{\mathrm{T}}(\mathrm{Z}) > 200\,\mathrm{GeV}$ are selected. The $\mathrm{H} \to \mathrm{c\overline{c}}$ candidate is reconstructed from the large-R jets passing the requirements described in Section 8.2.2. At $p_{\mathrm{T}}(\mathrm{Z}) > 200\,\mathrm{GeV}$, there are typically zero or only one large-R jets in each signal or background event. Only in very rare cases does an event have two or more large-R jets. The leading large-R jet is chosen as the Higgs boson candidate ($\mathrm{H_{cand}}$) in the event. Events with no large-R jets are discarded. In VH signal events, the vector boson and the Higgs boson are typically back-to-back. Therefore, we require the difference in azimuthal angle between the reconstructed vector boson and the Higgs candidate, $\Delta\phi(\mathrm{V},\mathrm{H})$, to be above 2.5 radians. To avoid double-counting, small-R jets are removed from the event if they overlap with $\mathrm{H_{cand}}$, i.e., satisfying $\Delta R(\text{small-}R\text{ jet}, \mathrm{H_{cand}}) < 1.5$. The number of small-$R$ jets, denoted as N_{j}, is required to be less than 3 in the 2-lepton channel.

After the baseline selection, the dominant background process is Z+jets, with a small contribution from diboson production. The $\mathrm{t\overline{t}}$ contribution is highly suppressed due to the requirement of $p_{\mathrm{T}}(\mathrm{Z}) > 200\,\mathrm{GeV}$. The $\mathrm{VH}(\mathrm{H} \to \mathrm{b\overline{b}})$ process is a very small but largely irreducible background.

To better distinguish the VH signal from the main backgrounds, a kinematic BDT

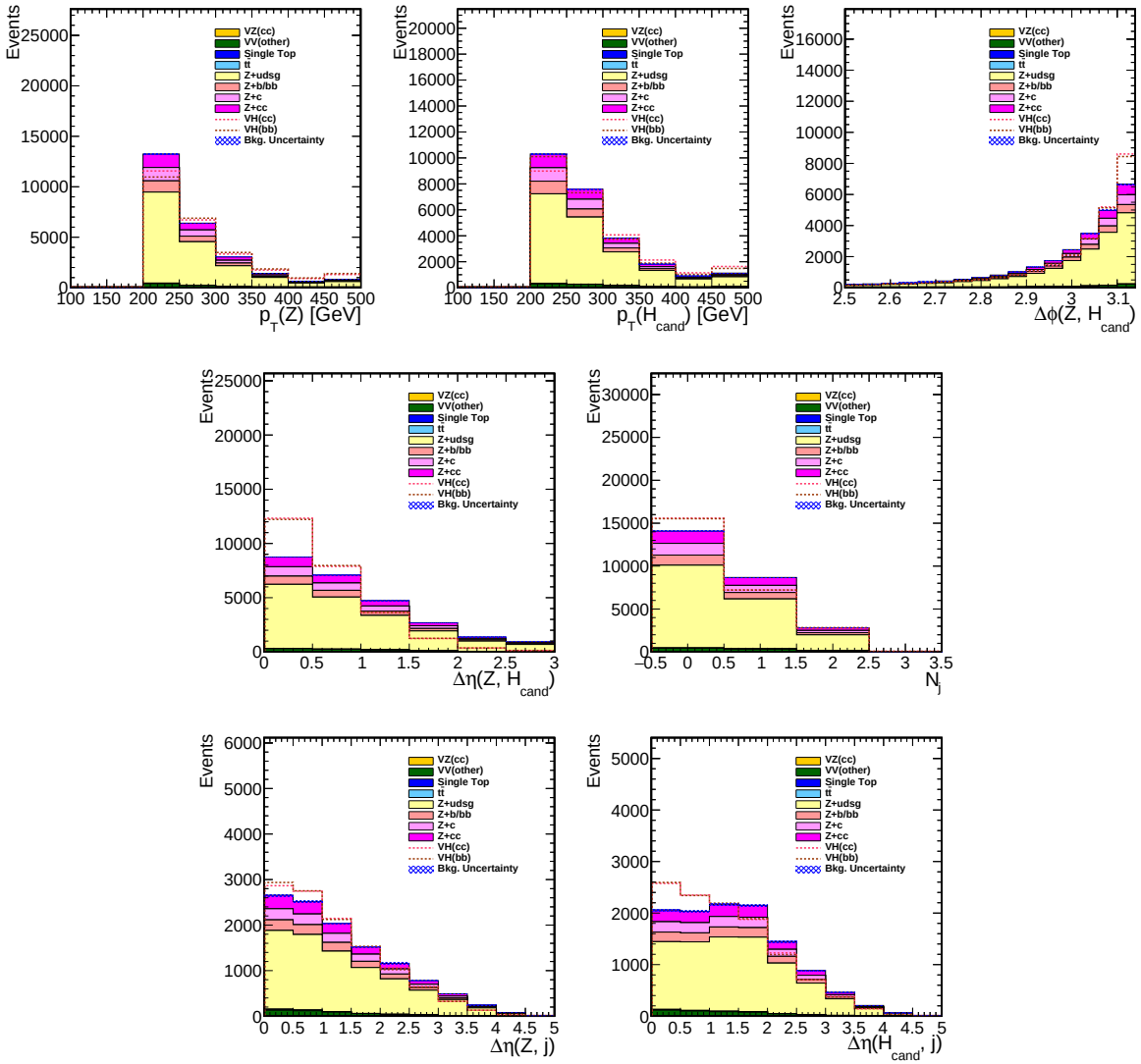


Figure 8.4: Comparison of the kinematic BDT input variables between the VH signal and the background samples in the 2-lepton channel. The VH signal is normalized to the sum of all backgrounds. The definitions of the variables can be found in Table 8.3. For $\min\Delta\eta(\mathrm{Z},\mathrm{j})$ and $\min\Delta\eta(\mathrm{H_{cand}},\mathrm{j})$ variables on the bottom row, only events with at least one additional small-R jet are shown in the plots.

is developed. The kinematic BDT is designed using only the *kinematic* information of the event, i.e., no intrinsic properties of the H_{cand}, like the flavor content, the mass, etc., of the large-R jet are included in the BDT. The goal is to design a BDT that improves the separation between VH signal and the main backgrounds (Z+jets in the 2-lepton channel), while remaining largely independent of the $c\bar{c}$-tagging discriminant and the mass of the H_{cand}. This is achieved by selecting only input variables that are not highly correlated with the intrinsic properties of H_{cand} for the training of the BDT. The input variables used in the kinematic BDT for the 2-lepton channel are defined in Table 8.3, and the distribution of these variables in the VH signal and in the backgrounds after the baseline selection (but without any requirement on the $c\bar{c}$-tagging discriminant) are shown in Figure 8.4.

The kinematic BDT is trained with the XGBoost [192] package. To avoid any potential bias from overfitting the training dataset, dedicated samples are used in the training of the BDT. For the signal process, we use the $VH(H \to b\bar{b})$ signal sample instead of the $VH(H \to c\bar{c})$ signal sample, and use only events with even event numbers in the training, while keeping events with odd event numbers for the main analysis. As the BDT is designed to be insensitive to the flavor content of the H_{cand}, training with the $VH(H \to b\bar{b})$ signal sample results in no loss of performance. For the background process, only the main background, i.e., Z+jets in the case of the 2-lepton channel is used, and we use an alternative Z+jets sample (generated at LO QCD accuracy) for the training. In this way, the events used in the training of the BDT are completely orthogonal to the events used in the main analysis, thereby avoiding any potential bias from overfitting.

Figure 8.5 shows the distribution of the kinematic BDT after the baseline selection for the VH signal and the backgrounds. A good separation is observed. A cut at 0.5 is applied on the output of the kinematic BDT to categorize events, with events having high BDT outputs (above 0.5) entering the signal region, while the other events enter

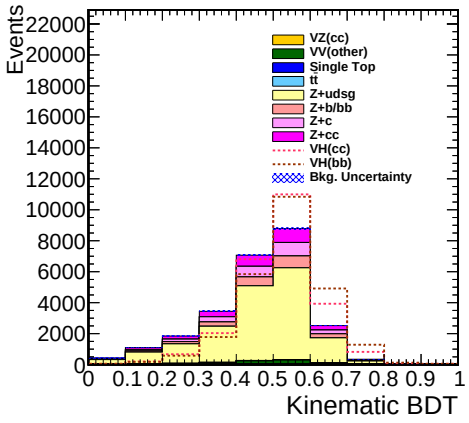


Figure 8.5: Comparison of the kinematic BDT output distribution between the VH signal and the background samples in the 2-lepton channel. The VH signal is normalized to the sum of all backgrounds.

the control region. We do not use the detailed shape of the BDT output in the analysis in order to be robust against the details in the modeling of the event kinematics in the simulated samples.

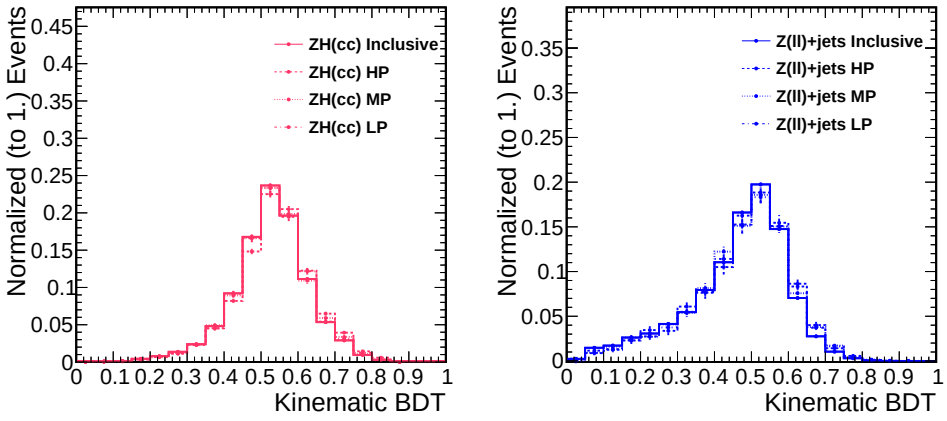


Figure 8.6: Comparison of the kinematic BDT output distribution before and after the requirement on the $c\bar{c}$-tagging discriminant.

Figure 8.6 shows the distributions of the kinematic BDT for the $VH(H \to c\bar{c})$ signal and the V+jets background before and after selections on the $c\bar{c}$-tagging discriminant. The shapes of the BDT are very similar, confirming that the kinematic BDT is indeed

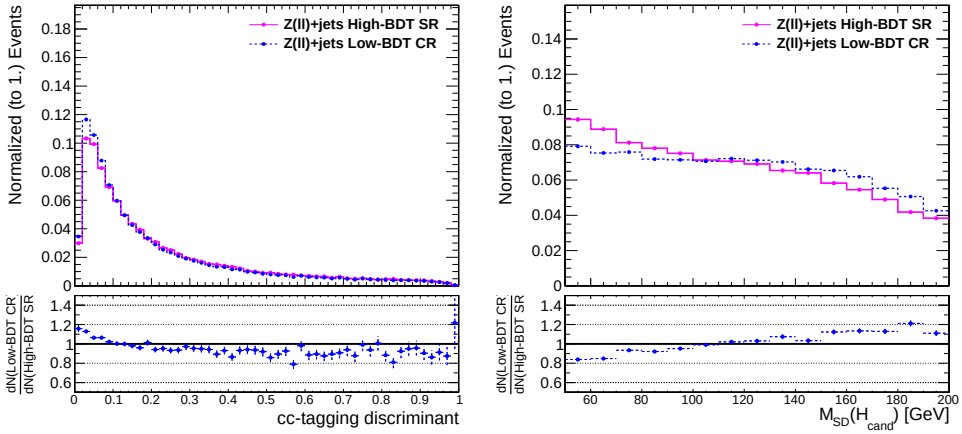


Figure 8.7: Comparison of the c$\bar{\text{c}}$-tagging discriminant (left) and the soft-drop mass shapes (right) between events with high (>0.5) BDT outputs and low (<0.5) BDT outputs for the Z+jets sample in the 2-lepton channel.

independent of the c$\bar{\text{c}}$-tagging discriminant. This is also verified in Figure 8.7, which shows that the distribution of the c$\bar{\text{c}}$-tagging discriminant and the soft-drop mass of the H_{cand} are largely invariant with respect to the selection on the kinematic BDT.

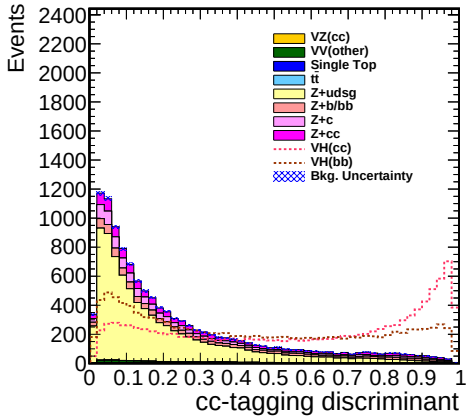


Figure 8.8: Comparison of the c$\bar{\text{c}}$-tagging discriminant between the VH(H $\rightarrow$ c$\bar{\text{c}}$) signal and the backgrounds for events with high kinematic BDT output scores.

After the selection based on the kinematic BDT, events are required to pass the c$\bar{\text{c}}$-tagging requirement. The distribution of the c$\bar{\text{c}}$-tagging discriminant is shown in Figure 8.8 for signal and background events in the high BDT region. The c$\bar{\text{c}}$-tagging discriminant

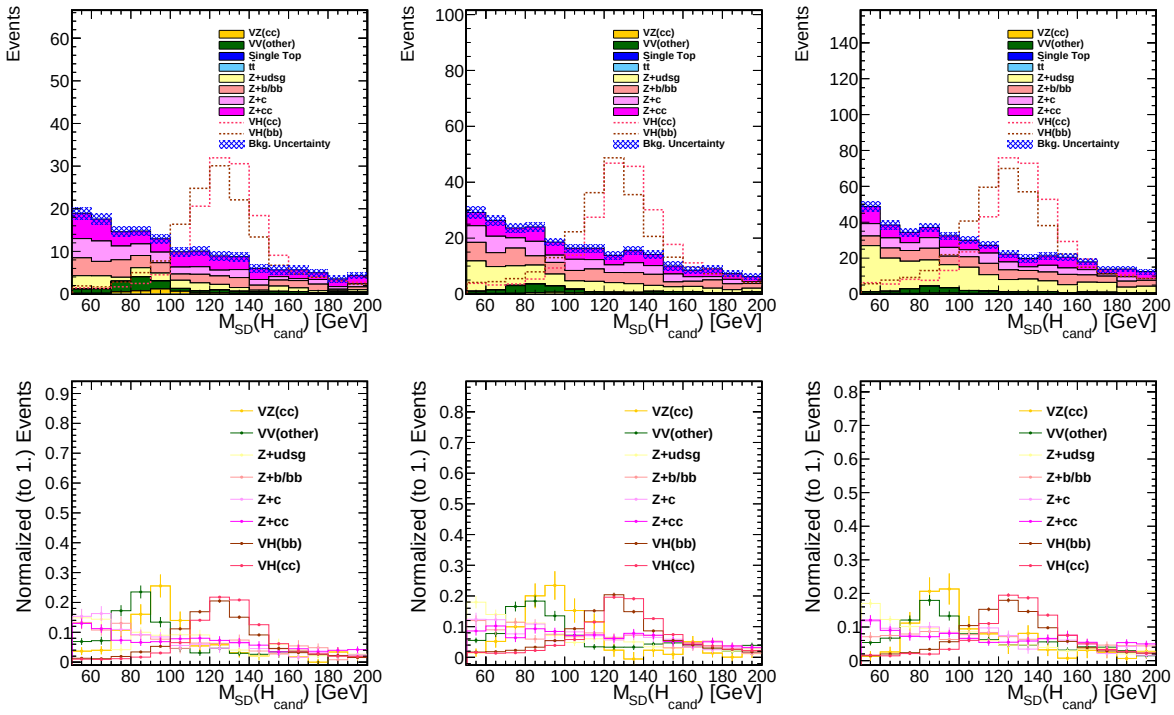


Figure 8.9: Comparison of the soft-drop mass of the H_{cand} between the $VH(H \to c\bar{c})$ signal and the backgrounds for events with high kinematic BDT output scores for the HP (left), MP (middle) and LP (right) categories in the 2-lepton channel. Top: with backgrounds stacked and the $VH(H \to c\bar{c})$ and $VH(H \to b\bar{b})$ processes normalized to the sum of all backgrounds; Bottom: shape comparison between the $VH(H \to c\bar{c})$ signal and the various background processes, where the error bars correspond to the uncertainties due to limited size of the simulated samples.

shows powerful separation power between $H \to c\bar{c}$ signal and Z+jets background, even after the selection on the BDT. Events are categorized into three $c\bar{c}$-tagging categories as defined in Section 8.3. Figure 8.9 displays the distribution of the soft-drop mass of the H_{cand} for events with high kinematic BDT output scores and passing the $c\bar{c}$-tagging requirement for the three categories. Since both the kinematic BDT and the $c\bar{c}$-tagging discriminant are designed to be independent of the mass of the large-R jet, the $VH(H \to c\bar{c})$ signal process and the various background processes show distinct mass spectra, therefore enabling us to extract the $VH(H \to c\bar{c})$ signal by fitting the soft-drop mass of the H_{cand}.

8.4.2 The 1-Lepton Channel

The 1-lepton channel targets WH events where the W bosons decay leptonically. Events are selected with exactly one electron or muon passing the loose lepton definition described in Section 8.2.1. The selected lepton is required to pass a tighter identification and isolation criteria to reject non-prompt leptons from the QCD multijet background more effectively. The p_T of the muon (electron) is required to be above 25 (30) GeV to ensure that the single-lepton trigger is almost 100% efficient. The W boson p_T is reconstructed as the vector sum of the charged lepton momentum and $\vec{p}_T^{\,miss}$. The event configuration is required to be compatible with the leptonic decay of a high-p_T W boson, with the azimuthal separation between $\vec{p}_T^{\,miss}$ and the charged lepton $\Delta\phi(\vec{p}_T^{\,miss}, \ell) < 1.5$.

The H_{cand} is reconstructed from the large-R jets in the same way as in the 2-lepton channel. Small-R jets overlapping with H_{cand} are again removed from the events. The number of small-R jets after removing the overlaps is required to be less than 2 in the 1-lepton channel to suppress the contribution from the $t\bar{t}$ process.

As is done in the 2-lepton channel, a kinematic BDT is also developed for the 1-

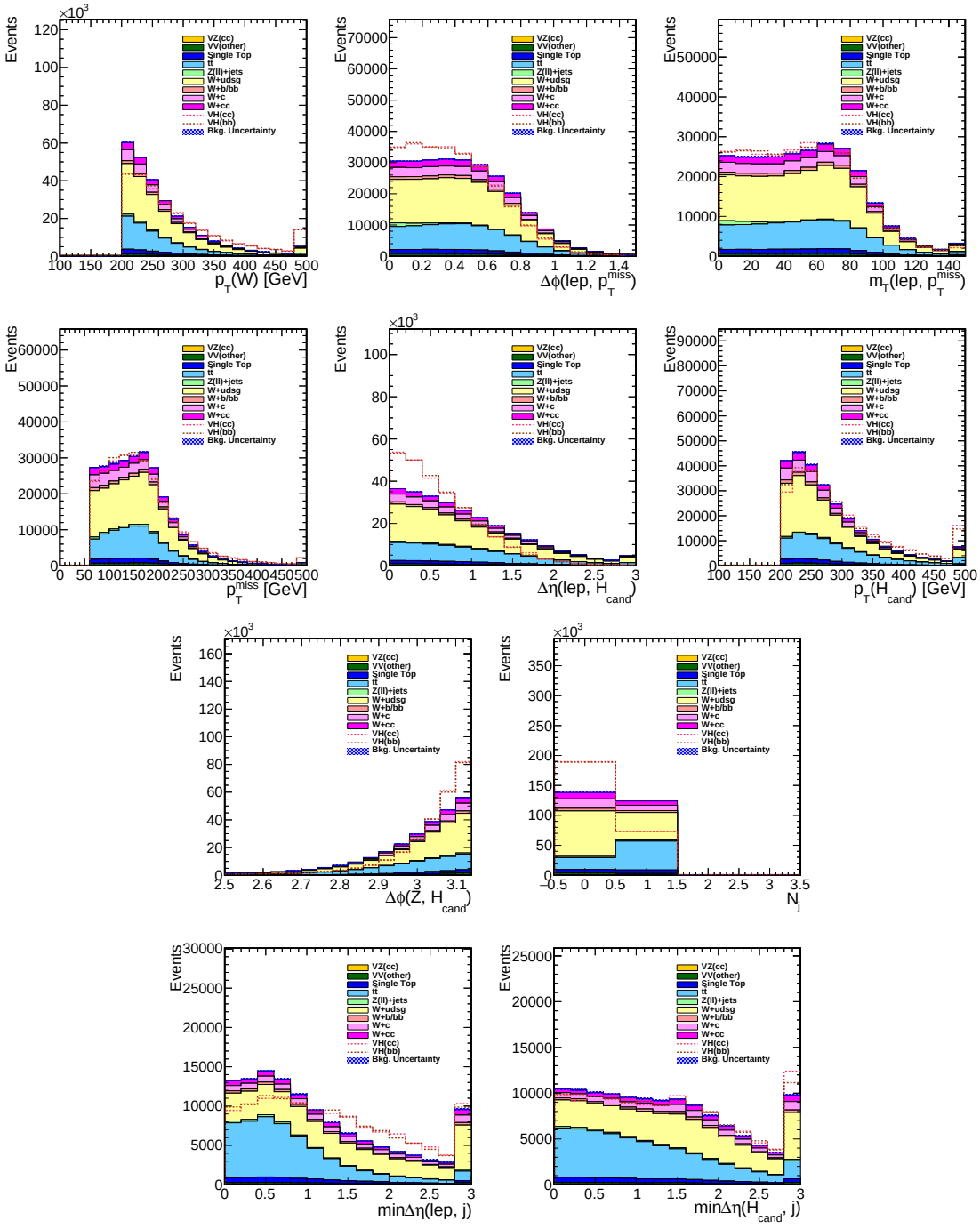


Figure 8.10: Comparison of the kinematic BDT input variables between the VH signal and the background samples in the 1-lepton channel. The VH signal is normalized to the sum of all backgrounds. The definitions of the variables can be found in Table 8.3. For $\min \Delta\eta(\ell, \mathrm{j})$ and $\min \Delta\eta(\mathrm{H_{cand}}, \mathrm{j})$ variables on the bottom row, only events with at least one additional small-R jet are shown in the plots.

lepton channel to reject the main backgrounds, which are W+jets and $t\bar{t}$. The same design philosophy of being independent of the intrinsic properties of H_{cand} is followed, as well as the same training strategy of using fully orthogonal samples in the training of the BDT. The input variables used in the kinematic BDT for the 1-lepton channel are defined in Table 8.3, and the distributions of these variables in the VH signal and in the backgrounds are shown in Figure 8.10.

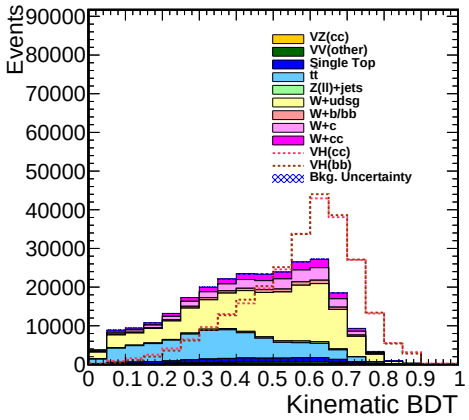


Figure 8.11: Comparison of the kinematic BDT output distribution between the VH signal and the background samples in the 1-lepton channel. The VH signal is normalized to the sum of all backgrounds.

Figure 8.11 shows the distribution of the kinematic BDT after the baseline selection for the VH signal and the backgrounds. A good separation is observed. A cut at 0.5 is applied on the output of the kinematic BDT to categorize events, with events having high BDT outputs (above 0.5) entering the signal region, while the other events enter the control region.

As in the case of the 2-lepton channel, we also compare the shapes of the kinematic BDT before and after selections on the $c\bar{c}$-tagging discriminant as shown in Figure 8.12 to check that the kinematic BDT is indeed independent of the $c\bar{c}$-tagging discriminant. This is also verified in Figure 8.13, which shows that the distribution of the $c\bar{c}$-tagging discriminant and the soft-drop mass of the H_{cand} is largely unchanged with respect to

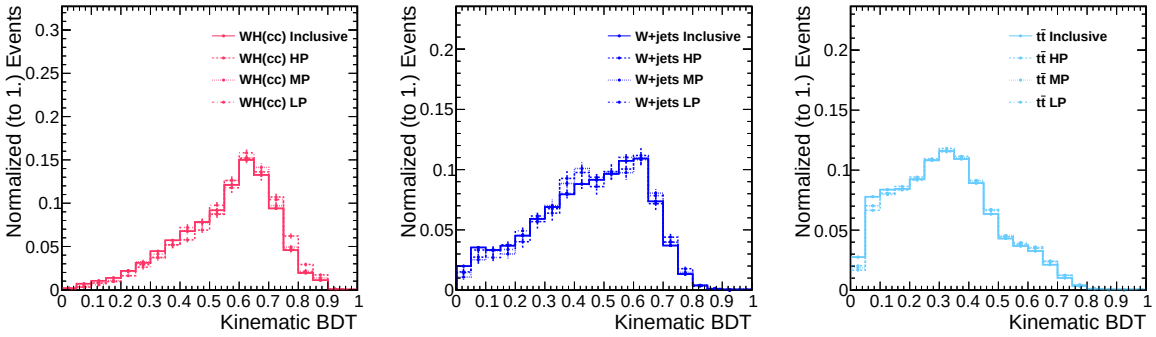


Figure 8.12: Comparison of the kinematic BDT output distribution before and after the requirement on the c$\bar{c}$-tagging discriminant.

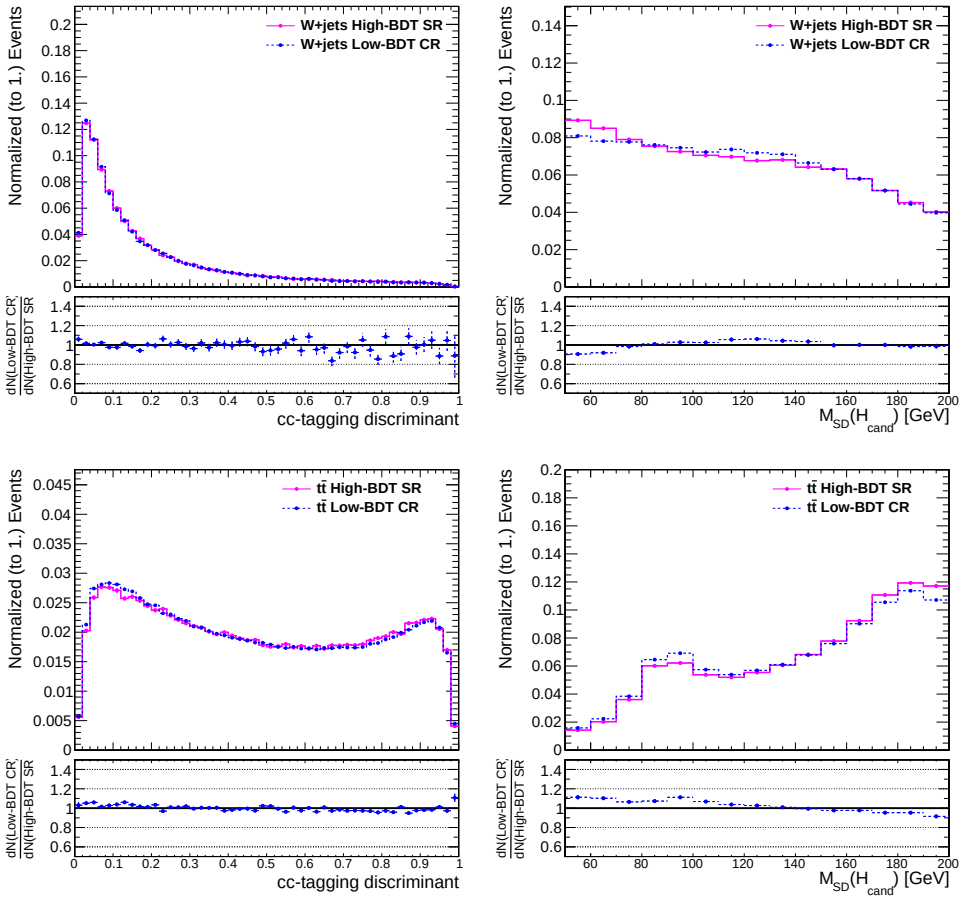


Figure 8.13: Comparison of the c$\bar{c}$-tagging discriminant (left) and the soft-drop mass shapes (right) between events with high (>0.5) BDT outputs and low (<0.5) BDT outputs for W+jets (top) and t$\bar{t}$ (bottom) in the 1-lepton channel.

the selection on the kinematic BDT.

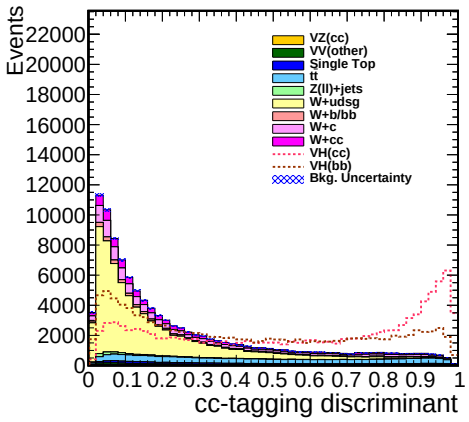


Figure 8.14: Comparison of the c$\bar{c}$-tagging discriminant between the VH(H $\rightarrow$ c$\bar{c}$) signal and the backgrounds for events with high kinematic BDT output scores.

After the selection based on the kinematic BDT, events are required to pass the c$\bar{c}$-tagging requirement. The distribution of the c$\bar{c}$-tagging discriminant is shown in Figure 8.14 for signal and background events in the high BDT region. The c$\bar{c}$-tagging discriminant shows powerful separation power between H $\rightarrow$ c$\bar{c}$ signal and W+jets background which is dominated by light flavor jets. After applying the selection on the c$\bar{c}$-tagging discriminant, the W+jets background is highly suppressed and becomes smaller than the t$\bar{t}$ background. Figure 8.15 displays the distribution of the soft-drop mass of the H_{cand} for events with high BDT scores and passing the c$\bar{c}$-tagging requirements for the three c$\bar{c}$-tagging categories. Again, since both the kinematic BDT and the c$\bar{c}$-tagging discriminant are designed to be independent of the large-R jet mass, the VH(H $\rightarrow$ c$\bar{c}$) signal process and the various background processes show distinct mass spectra.

8.4.3 The 0-Lepton Channel

The 0-lepton channel targets mainly ZH events with the Z boson decaying into a pair of neutrinos, giving rise to large $p_{\mathrm{T}}^{\mathrm{miss}}$. WH events in which the W boson decays

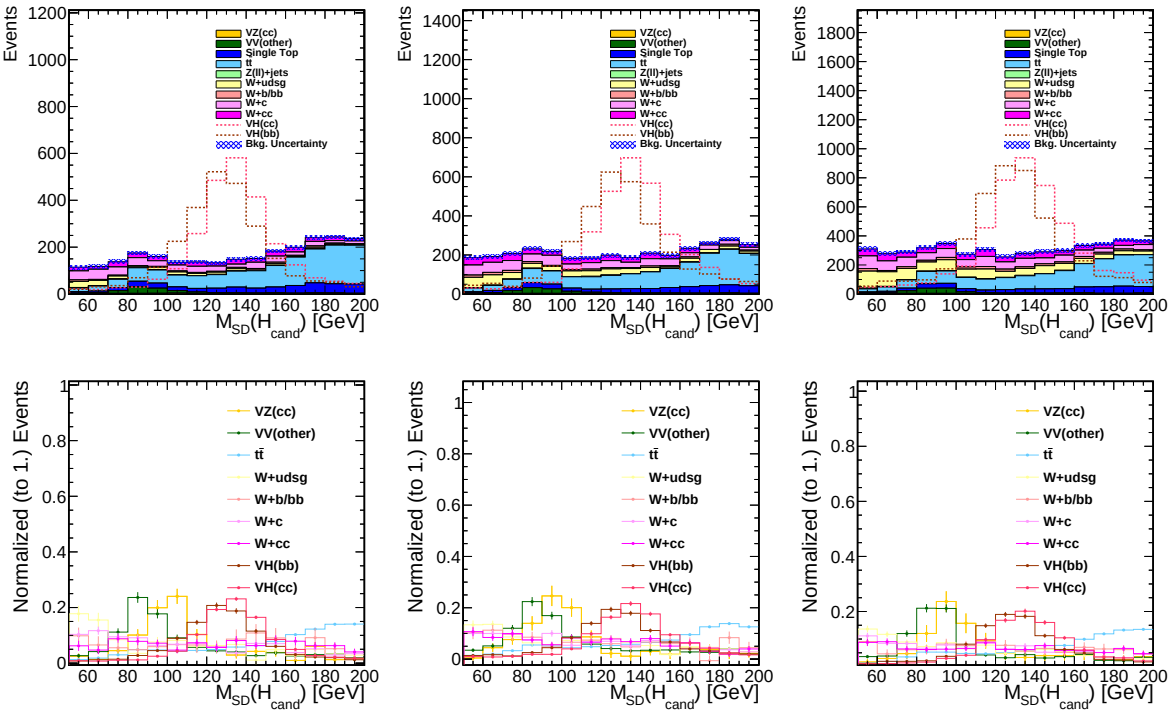


Figure 8.15: Comparison of the soft-drop mass of the H_{cand} between the $VH(H \to c\bar{c})$ signal and the backgrounds for events with high kinematic BDT output scores for the HP (left), MP (middle) and LP (right) categories in the 1-lepton channel. Top: with backgrounds stacked and the $VH(H \to c\bar{c})$ and $VH(H \to b\bar{b})$ processes normalized to the sum of all backgrounds; Bottom: shape comparison between the $VH(H \to c\bar{c})$ signal and the various background processes, where the error bars correspond to the uncertainties due to limited size of the simulated samples.

leptonically can also enter this channel when the charged lepton fails to be identified. For the 0-lepton channel, events are selected with no electrons or muons passing the loose lepton definition described in Section 8.2.1. The $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ corresponds to the transverse momentum of the vector boson in this channel.

The $\mathrm{H}_{\mathrm{cand}}$ is reconstructed in the same way as in the 1- and 2-lepton channels. Small-R jets overlapping with $\mathrm{H}_{\mathrm{cand}}$ are removed from the events. The number of small-R jets after removing the overlaps is required to be less than 2 to suppress the contribution from the $\mathrm{t\bar{t}}$ process, same as in the 1-lepton channel.

Similar to the 1- and 2-lepton channels, a kinematic BDT is developed following the same approach to reject the main backgrounds (Z+jets, W+jets and $\mathrm{t\bar{t}}$) for the 0-lepton channel. The input variables used in the kinematic BDT are defined in Table 8.3, and the distribution of these variables in the VH signal and in the backgrounds are shown in Figure 8.16.

Figure 8.17 shows the distribution of the kinematic BDT after the baseline selection for the VH signal and the backgrounds. The separation power of the kinematic BDT is a bit weaker in the 0-lepton channel than in the 1- and 2-lepton channels, as only the kinematic information of the vector boson in the transverse plane is reconstructed. A cut at 0.5 is applied on the output of the kinematic BDT to categorize events, with events having high BDT outputs (above 0.5) entering the signal region, while the other events entering the control region.

Figure 8.18 shows the distributions of the kinematic BDT for the $\mathrm{VH}(\mathrm{H} \rightarrow \mathrm{c\bar{c}})$ signal and different background processes before and after selections on the $\mathrm{c\bar{c}}$-tagging discriminant. The shapes of the BDT are largely unaffected by the selection on the $\mathrm{c\bar{c}}$-tagging discriminant, confirming that the kinematic BDT is indeed independent of the intrinsic properties (e.g., the flavor content) of the jet. Such independence is also demonstrated in Figure 8.19 by comparing the distribution of the $\mathrm{c\bar{c}}$-tagging discriminant

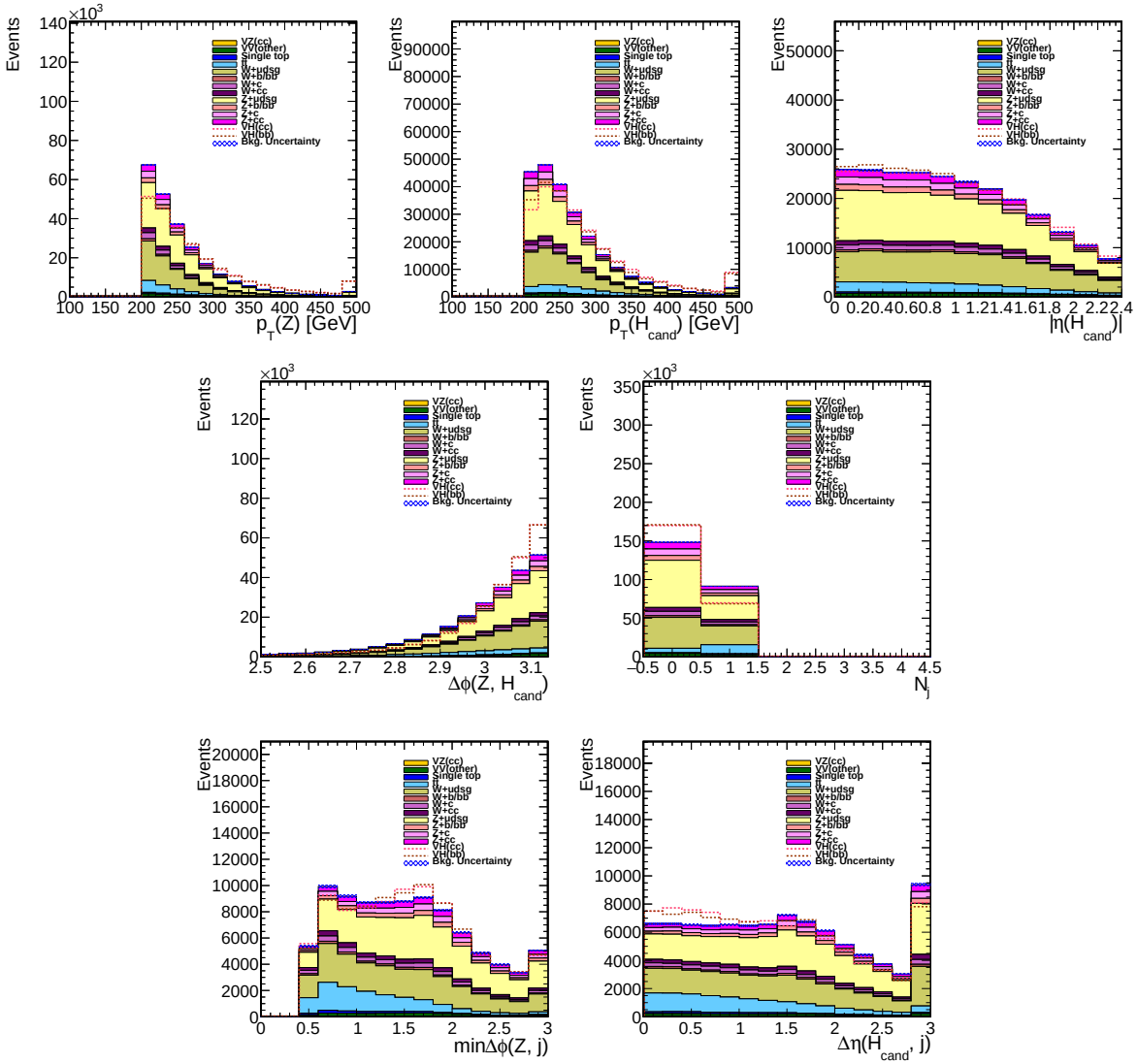


Figure 8.16: Comparison of the kinematic BDT input variables between the VH signal and the background samples in the 0-lepton channel. The VH signal is normalized to the sum of all backgrounds. The definitions of the variables can be found in Table 8.3. For the $\min\Delta\phi(\mathrm{V},\mathrm{j})$ and $\min\Delta\eta(\mathrm{H_{cand}},\mathrm{j})$ variables on the bottom row, only events with at least one additional small-R jet are shown in the plots.

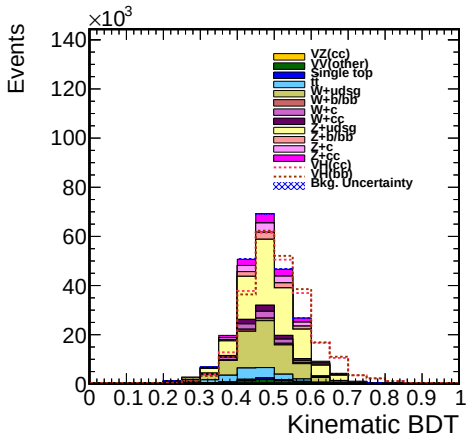


Figure 8.17: Comparison of the kinematic BDT output distribution between the VH signal and the background samples in the 0-lepton channel. The VH signal is normalized to the sum of all backgrounds.

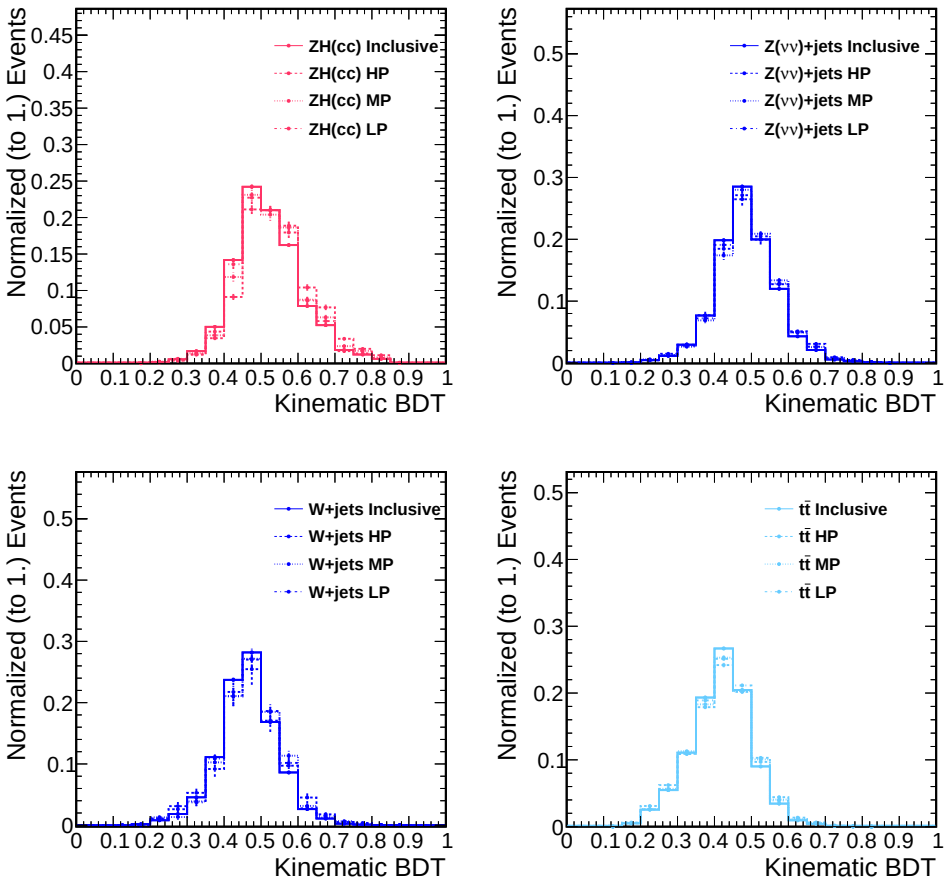


Figure 8.18: Comparison of the kinematic BDT output distribution before and after the requirement on the $c\bar{c}$-tagging discriminant.

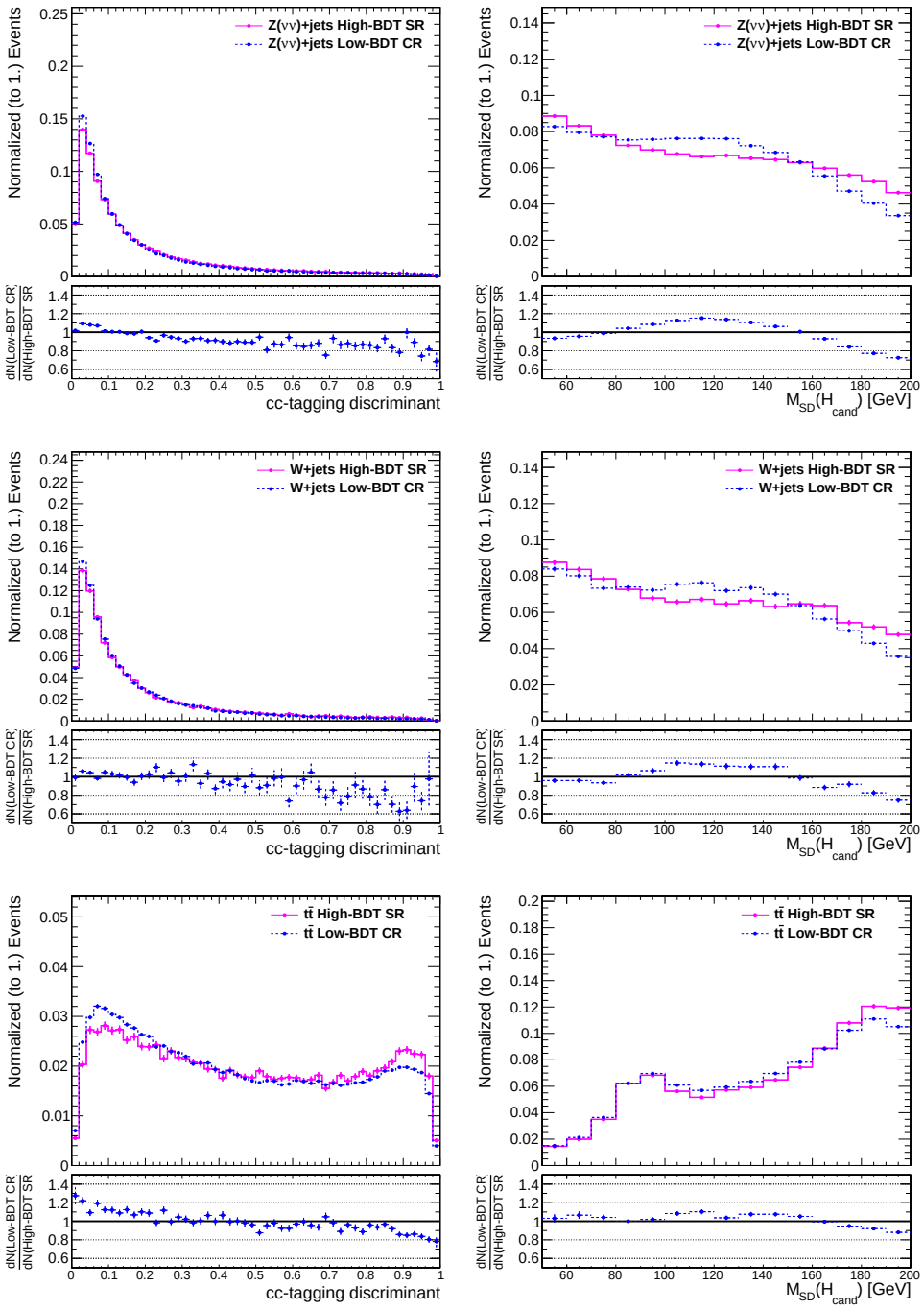


Figure 8.19: Comparison of the $c\bar{c}$-tagging discriminant (left) and the soft-drop mass shapes (right) between events with high (>0.5) BDT outputs and low (<0.5) BDT outputs for Z+jets (top), W+jets (middle) and $t\bar{t}$ (bottom) in the 0-lepton channel.

and the soft-drop mass of the H_{cand} between events with high and low BDT outputs.

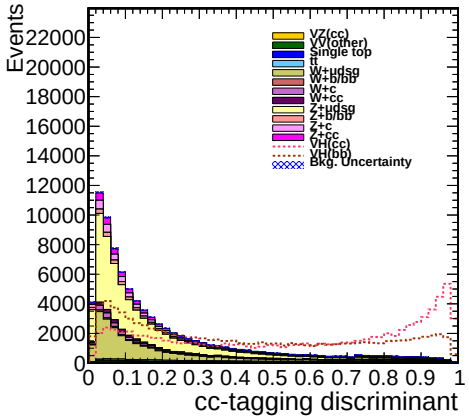


Figure 8.20: Comparison of the $c\bar{c}$-tagging discriminant between the $VH(H \to c\bar{c})$ signal and the backgrounds for events with high kinematic BDT output scores.

After the selection based on the kinematic BDT, events are required to pass the $c\bar{c}$-tagging requirement. The distribution of the $c\bar{c}$-tagging discriminant is shown in Figure 8.20 for signal and background events in the high BDT region. The $c\bar{c}$-tagging discriminant shows powerful separation power between $H \to c\bar{c}$ signal and V+jets background. After applying selection criteria on the $c\bar{c}$-tagging discriminant, the V+light flavor components are highly suppressed, leaving mostly V+heavy flavor components, as well as $t\bar{t}$ events, as the main backgrounds. Figure 8.21 displays the distribution of the soft-drop mass of the H_{cand} for events with high kinematic BDT output scores and satisfying the $c\bar{c}$-tagging requirement for the three $c\bar{c}$-tagging categories. Again, the $VH(H \to c\bar{c})$ signal process and the various background processes show distinct mass shapes.

8.4.4 Summary of the Event Selection

Table 8.3 summarizes kinematic variables used as inputs to the kinematic BDT for the three channels. The full event selections for all three channels are summarized in Table 8.4. Table 8.5 provides a summary of the expected background and signal events

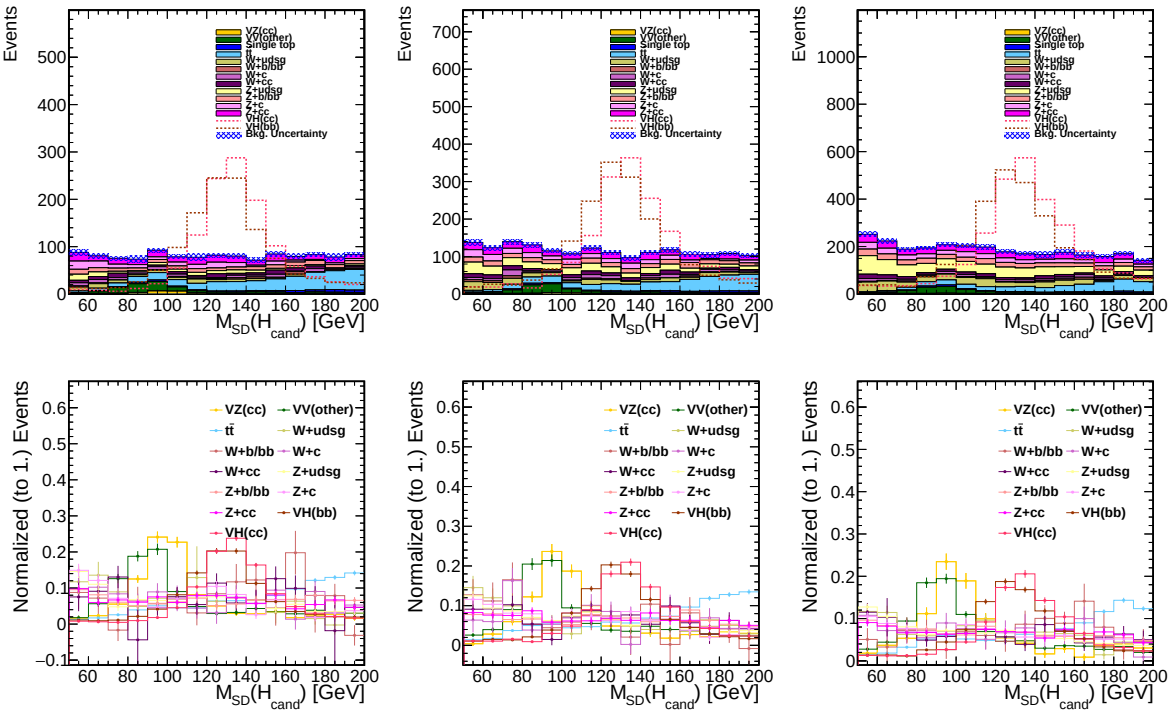


Figure 8.21: Comparison of the soft-drop mass of the H_{cand} between the $VH(H \to c\bar{c})$ signal and the backgrounds for events with high kinematic BDT output scores for the HP (left), MP (middle) and LP (right) categories in the 0-lepton channel. Top: with backgrounds stacked and the $VH(H \to c\bar{c})$ and $VH(H \to b\bar{b})$ processes normalized to the sum of all backgrounds; Bottom: shape comparison between the $VH(H \to c\bar{c})$ signal and the various background processes, where the error bars correspond to the uncertainties due to limited size of the simulated samples.

Variable	Description	0-lepton	1-lepton	2-lepton
$p_{\mathrm{T}}(\mathrm{V})$	vector boson transverse momentum	✓	✓	✓
$p_{\mathrm{T}}(\mathrm{H_{cand}})$	$\mathrm{H_{cand}}$ transverse momentum	✓	✓	✓
$\lvert\eta(\mathrm{H_{cand}})\rvert$	absolute value of the $\mathrm{H_{cand}}$ pseudorapidity	✓		
$\Delta\phi(\mathrm{V},\mathrm{H})$	azimuthal angle between vector boson and $\mathrm{H_{cand}}$	✓	✓	✓
$p_{\mathrm{T}}^{\mathrm{miss}}$	missing transverse momentum		✓	
$\Delta\eta(\mathrm{H_{cand}},\ell)$	difference in pseudorapidity between $\mathrm{H_{cand}}$ and the lepton		✓	
$\Delta\eta(\mathrm{H_{cand}},\mathrm{V})$	difference in pseudorapidity between $\mathrm{H_{cand}}$ and vector boson			✓
$\Delta\eta(\mathrm{H_{cand}},\mathrm{j})$	min. difference in pseudorapidity between $\mathrm{H_{cand}}$ and small-R jets	✓	✓	✓
$\Delta\eta(\ell,\mathrm{j})$	min. difference in pseudorapidity between the lepton and small-R jets		✓	
$\Delta\eta(\mathrm{V},\mathrm{j})$	min. difference in pseudorapidity between vector boson and small-R jets			✓
$\Delta\phi(\vec{p}_{\mathrm{T}}^{\mathrm{miss}},\mathrm{j})$	azimuthal angle between $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ and closest small-R jet	✓		
$\Delta\phi(\vec{p}_{\mathrm{T}}^{\mathrm{miss}},\ell)$	azimuthal angle between $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ and lepton		✓	
m_{T}	transverse mass of lepton $\vec{p}_{\mathrm{T}}$ + $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$		✓	
N_{j}	number of small-R jets	✓	✓	✓

Table 8.3: Variables used in the training of the kinematic BDT.

in the signal regions of each channel, obtained directly from the simulated samples.

8.5 Background Estimation and Signal Extraction

Dedicated control regions enriched in different backgrounds are defined to study the agreement between data and simulation and aid the background estimation in each channel. Two types of control regions are defined in this analysis, as illustrated in Figure 8.22: the low-BDT control region is defined by selecting events with kinematic BDT outputs below 0.5, which is enriched in the V+jets background; the high-N_{j} control region is defined by inverting the selection on the number of small-R jets, which leads to a high-purity $\mathrm{t\bar{t}}$ sample. Note that the latter is not used for the 2-lepton channel, as the background is dominated by Z+jets, while the contribution from $\mathrm{t\bar{t}}$ is negligible. In both types of control regions, events are required to satisfy the same $\mathrm{c\bar{c}}$-tagging discriminant criteria as in the signal regions for each $\mathrm{c\bar{c}}$-tagging category in order to probe events with a flavor composition similar to that in the signal regions. As a result, the efficiency of the $\mathrm{c\bar{c}}$-tagging discriminant can be estimated directly from the data using the control regions, and no additional corrections are needed.

Variable	0-lepton	1-lepton	2-lepton
$p_{\mathrm{T}}(\mathrm{V})$	>200	>200	>200
$m(\ell\ell)$	—	—	$[75, 105]$
p_{T}^{ℓ}	—	$(> 25, > 30)$	>20
$p_{\mathrm{T}}(\mathrm{H_{cand}})$	>200	>200	>200
$m_{\mathrm{SD}}(\mathrm{H_{cand}})$	$[50, 200]$	$[50, 200]$	$[50, 200]$
$\Delta\phi(\mathrm{V}, \mathrm{H})$	>2.5	>2.5	>2.5
N_{j}	<2	<2	<3
$N_{\mathrm{a}\ell}$	=0	=0	—
$p_{\mathrm{T}}^{\mathrm{miss}}$	>200	—	—
$\Delta\phi(\vec{p}_{\mathrm{T}}^{\mathrm{miss}}, \mathrm{j})$	>0.5	—	—
$\Delta\phi(\vec{p}_{\mathrm{T}}^{\mathrm{miss}}, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}(\mathrm{trk}))$	<0.5	—	—
$\Delta\phi(\vec{p}_{\mathrm{T}}^{\mathrm{miss}}, \ell)$	—	<1.5	—
Lepton isolation	—	<0.06	$(< 0.25, < 0.15)$
Kinematic BDT	>0.5	>0.5	>0.5
$\mathrm{c\bar{c}}$-tagging discriminant			
High purity (HP)	> 0.91	> 0.91	> 0.91
Medium purity (MP)	$(0.83, 0.91]$	$(0.83, 0.91]$	$(0.83, 0.91]$
Low purity (MP)	$(0.72, 0.83]$	$(0.72, 0.83]$	$(0.72, 0.83]$

Table 8.4: Event selection criteria for the signal region. Entries marked with "—" indicate that the variable is not used in the given channel. The values listed for kinematic variables are in units of GeV, and for angles in units of radians. Where selection differs between lepton flavors, the selection is listed as (muon, electron).

The $\mathrm{VH}(\mathrm{H} \to \mathrm{c\bar{c}})$ signal is extracted via a binned maximum likelihood fit on the m_{SD} shape of the $\mathrm{H_{cand}}$. As shown in Figures 8.9, 8.15 and 8.21, the m_{SD} of $\mathrm{H_{cand}}$ provides excellent separation between the $\mathrm{VH}(\mathrm{H} \to \mathrm{c\bar{c}})$ signal and the various background processes. Therefore, it is a powerful fit variable for extracting the rates of different processes. The m_{SD} shapes for signal and backgrounds are obtained directly from simulation, as the modeling of m_{SD} in simulation was validated in various control regions, and good agreements between data and simulation were observed. A total of 15 bins for the range of $m_{\mathrm{SD}} \in [50, 200]\,\mathrm{GeV}$ are used in the fit, with a bin width of 10 GeV corresponding roughly to the m_{SD} resolution.

The low-BDT and the high-N_{j} control regions are included together with the signal

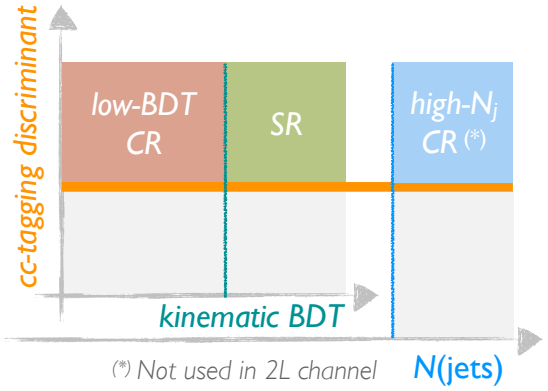


Figure 8.22: Illustration of the setup of the signal region and control regions.

regions in the maximum likelihood fit. This is to correct for any difference between data and simulation in the production rate of the V+jets and $t\bar{t}$ processes in the phase space selected by this analysis. A parameter that scales the overall normalization of a background process is assigned to each of W+jets, Z+jets and $t\bar{t}$ processes in each $c\bar{c}$-tagging category and is allowed to float freely in the fit. The same parameter scales the background rate in the control regions and the signal regions in the same way. With this approach, any potential difference in the $c\bar{c}$-tagging efficiency between data and simulation is also taken into account, as the same selection on the $c\bar{c}$-tagging discriminant is applied across all signal and control regions for each $c\bar{c}$-tagging category. Different parameters are defined for the three $c\bar{c}$-tagging categories, as the difference in $c\bar{c}$-tagging

Process	High purity			Medium purity			Low purity		
	0-lepton	1-lepton	2-lepton	0-lepton	1-lepton	2-lepton	0-lepton	1-lepton	2-lepton
V+jets	621.9	778.7	120.2	1011.3	1199.4	180.9	1622.4	1976.5	293.9
$t\bar{t}$	277.8	1191.1	0.6	294.7	1328.8	0.8	346.5	1636.1	0.9
Single top	41.5	346.3	0.1	47.8	359.8	0.2	54.4	428.0	0.5
$VZ(Z \to c\bar{c})$	32.7	19.6	5.8	21.6	11.2	3.4	20.0	11.8	3.6
VV(other)	96.0	122.0	13.4	109.8	141.7	16.8	144.8	181.8	21.4
$VH(H \to b\bar{b})$	4.3	5.0	1.2	4.8	6.4	1.7	5.2	6.6	1.6
Total background	1074.2	2462.7	141.4	1490.0	3047.3	203.7	2193.4	4240.9	321.8
$VH(H \to c\bar{c})$	0.8	1.0	0.2	0.5	0.6	0.1	0.5	0.6	0.1

Table 8.5: Expected background and signal events in the signal regions of each channel, predicted directly from simulated samples.

efficiency between data and simulation is likely to vary between the three different $c\bar{c}$-tagging categories. The parameters are defined separately for each channel, with the exception that the same scale factor is assumed for W+jets process in the 0-lepton and the 1-lepton channel. As the selection criteria are very similar between the two channels and the difference is only whether the charged lepton from the W boson decay is identified or not, using the same scale factor does not introduce any bias but can significantly improve the precision of the W+jets background estimation in the 0-lepton channel. The background normalization scale factors, determined from the simultaneous fit of the signal and the control regions, are summarized in Table 8.6.

Category	Channel	W+jets	Z+jets	$t\bar{t}$
High purity	0-lepton	1.10 ± 0.09	1.23 ± 0.13	1.11 ± 0.06
	1-lepton	1.10 ± 0.09	—	0.99 ± 0.04
	2-lepton	—	1.06 ± 0.11	—
Medium purity	0-lepton	1.17 ± 0.08	1.25 ± 0.10	1.12 ± 0.06
	1-lepton	1.17 ± 0.08	—	0.99 ± 0.04
	2-lepton	—	1.21 ± 0.09	—
Low purity	0-lepton	1.23 ± 0.07	1.45 ± 0.10	1.02 ± 0.06
	1-lepton	1.23 ± 0.07	—	0.99 ± 0.04
	2-lepton	—	1.24 ± 0.10	—

Table 8.6: Background normalization scale factors determined from the simultaneous fit of the signal and the control regions.

8.6 Systematic Uncertainties

In this section, various categories of systematic uncertainties that affect the analysis are discussed, and the full list of systematic uncertainties is summarized in Table 8.7.

Jet energy scale: uncertainties from the jet energy scale are assessed by varying the energy scale up and down for each jet within one standard deviation. This is done

individually for each of the 26 sources of uncertainties prescribed in Ref. [42]. The effect of the jet energy scale uncertainties is treated as correlated between the small-R jets and the large-R jets by varying the energy scales of the small-R jets and the subjets of the large-R jets simultaneously. The effects of the variation are also propagated to $p_{\mathrm{T}}^{\mathrm{miss}}$.

Jet energy resolution: uncertainties from the jet energy resolution are assessed by varying the simulation-to-data scale factors within their uncertainties when correcting the resolution of the jets in simulation. A single source of uncertainty is considered, including both the statistical and systematic uncertainties of the measured resolution scale factors. The effects of the jet energy resolution uncertainties are treated as correlated between the small-R jets and the large-R jets. The effects of the variation are also propagated to $p_{\mathrm{T}}^{\mathrm{miss}}$.

$p_{\mathrm{T}}^{\mathrm{miss}}$ unclustered energy: uncertainty in $p_{\mathrm{T}}^{\mathrm{miss}}$ related to the energy scale of the unclustered particles is evaluated based on the energy resolution of each type of particle as prescribed in detail in Ref. [45].

Lepton efficiency: the efficiency of the muon and electron trigger, reconstruction, identification, and isolation requirements are determined in data using the standard tag-and-probe technique with Z bosons. Simulation-to-data scale factors are applied to correct the efficiency in simulation. Uncertainties in the measured scale factors are evaluated from the statistical uncertainties of the measurement and the differences in the measured efficiency using alternative samples and selections, and are propagated to the analysis.

$p_{\mathrm{T}}^{\mathrm{miss}}$ trigger efficiency: the efficiency of the $p_{\mathrm{T}}^{\mathrm{miss}}$ triggers affects the expected yields of signals and of background processes that are predicted directly from simulated

samples in the 0-lepton channel. The trigger efficiency is measured in data, and an uncertainty of 2% is estimated.

Pileup reweighting: effects due to the uncertainty in the distribution of the number of pileup interactions are evaluated by varying the cross section used to predict the number of pileup interactions in simulation by 4.6% from its nominal value.

Top-p_{T} reweighting: a discrepancy in the top quark p_{T} spectrum in $\mathrm{t\bar{t}}$ events was observed in the measurement of differential cross sections for top quark pair production [181]. To mitigate this discrepancy, $\mathrm{t\bar{t}}$ events are reweighted to correct for the known difference. Uncertainty due to this reweighting is assessed by comparing the difference in the $\mathrm{t\bar{t}}$ yields and shapes before and after applying the reweighting. The top-p_{T} reweighting not only affects the shape of the top quark p_{T} distribution, but also has a moderate ($\sim 10\%$) effect on the overall yields in the high-p_{T} phase space selected by this analysis. Since the normalization of the $\mathrm{t\bar{t}}$ process will be determined via the simultaneous fit of the signal and control regions (using free-floating rate parameters), we keep only the shape-changing effect of this uncertainty, while the rate-changing effect is factored out by rescaling the $\mathrm{t\bar{t}}$ yields in the up and down variations to preserve the overall $\mathrm{t\bar{t}}$ yields after the baseline selection.

p_{T}(V) reweighting: NLO EW correction parametrized in p_{T}(V) is applied to the V+jets samples. Uncertainty on this NLO EW correction is estimated conservatively by comparing the shape difference with and without the correction, for W+jets and Z+jets processes separately. The overall yields are normalized after the baseline selection to factor out the moderate ($\sim 10\%$) effect on the normalization, as the normalizations of the V+jets processes are determined from the fit. The uncertainties for W+jets and Z+jets processes are treated as uncorrelated.

Parton distribution function (PDF): effects of the uncertainty in the PDF is assessed by reweighting each event according to the 100 replicas of the NNPDF3.0 set, and then taking the standard deviation of the 100 variations, following the prescription in [193]. Only effects on the shape are considered, i.e., the weights are rescaled to preserve the cross section of each process. The uncertainty is treated as uncorrelated for different processes.

Renormalization and factorization scales: uncertainties due to the renormalization scale, μ_R, and the factorization scale, μ_F, are assessed by varying the scales independently by a factor of 0.5 or 2. The effects of such variations are implemented as weights in the event generators. Only effects on the shape are considered, i.e., the weights are rescaled to preserve the cross section of each process. The uncertainty is treated as uncorrelated for different processes.

Size of simulated samples: uncertainties due to the limited size of the simulated samples are taken into account via the Barlow-Beeston method [194].

Uncertainty on the normalization of the backgrounds: the limited number of observed events in the signal and control regions leads to uncertainty in the normalizations of the V+jets and $t\bar{t}$ processes, as they are constrained by the observed data. The uncertainty is taken into account in the fit.

Cross section of the single top and diboson processes: the single top and the diboson backgrounds are predicted directly from simulation. An uncertainty of 15% is assigned for the single top process based on the cross section measurements [195, 196]. For the diboson process, an uncertainty of 10% is estimated based on the cross section measurements [197, 198, 199].

Luminosity: a 2.5% uncertainty is assigned on the integrated luminosity measured by

the CMS collaboration for the 2016 data-taking period [200].

Efficiency of $c\bar{c}$-tagging with DeepAK15: efficiency of tagging a $c\bar{c}$ pair with the DeepAK15 algorithm is measured in data, and simulation-to-data scale factors are applied to the $VH(H \to c\bar{c})$ and the $VZ(Z \to c\bar{c})$ processes. Efficiency of misidentifying a $b\bar{b}$ pair as a $c\bar{c}$ pair is also measured in data, and scale factors are applied to the $VH(H \to b\bar{b})$ and the $VZ(Z \to b\bar{b})$ processes. Uncertainty on the measured scale factors are taken into account in the analysis.

Uncertainties on the VH cross sections: uncertainties in the VH production cross section due to uncertainty of the PDF and the choice of the renormalization and factorization scales are evaluated in [15] and are taken into account in this analysis.

$p_T(V)$ distribution of the VH processes: the simulated VH samples used in this analysis are generated at NLO QCD accuracy. Higher-order NLO EW and NNLO QCD corrections, parametrized in $p_T(V)$, are used to reweight the simulated samples. Uncertainties due to this reweighting are taken into account.

Source	Type	0-lepton	1-lepton	2-lepton
Size of simulated samples	shape	✓	✓	✓
Jet energy scale	shape	✓	✓	✓
Jet energy resolution	shape	✓	✓	✓
Unclustered energy of $p_{\mathrm{T}}^{\mathrm{miss}}$	shape	✓	✓	
$c\bar{c}$ identification efficiency	shape	✓	✓	✓
Lepton efficiency	shape	✓	✓	
Pileup reweighting	shape	✓	✓	✓
top-p_{T} reweighting	shape	✓	✓	✓
$p_{\mathrm{T}}(\mathrm{V})$ reweighting	shape	✓	✓	✓
PDF	shape	✓	✓	✓
Renormalization and factorization scales	shape	✓	✓	✓
VH: $p_{\mathrm{T}}(\mathrm{V})$ NLO EW correction	shape	✓	✓	✓
Luminosity	rate	2.5%	2.5%	2.5%
MET trigger efficiency	rate	2%		
Single top cross section	rate	15%	15%	15%
Diboson cross section	rate	10%	10%	10%
VH: cross section (PDF)	rate	✓	✓	✓
VH: cross section (scale)	rate	✓	✓	✓

Table 8.7: Summary of the systematic uncertainties for each channel. The uncertainties are implemented in two ways in the fit: a "rate" uncertainty affects only the overall yields of a process but does not alter the relative shapes of the fit variable, while a "shape" uncertainty can modify both the overall yields and the shapes of the fit variable for a process.

Chapter 9

Results

To extract the VH($H \to c\bar{c}$) signal, a binned maximum likelihood fit is performed on the m_{SD} distribution of the H_{cand}, with the signal and the control regions fitted to the data simultaneously, as described in Section 8.5. The parameter of interest is the signal strength μ, which is defined as the ratio of the measured signal yield to the SM prediction. Figure 9.1 to 9.3 show the expected m_{SD} distributions of the VH($H \to c\bar{c}$) signal and the background processes in the signal regions after the fit, as well as the observed m_{SD} distributions in data. The corresponding plots for the control regions are shown in Figure 9.4 to 9.8. Overall, good agreement is observed between the predicted background and the observed data within uncertainties. The signal strength of VH($H \to c\bar{c}$) is determined to be $\mu_{\mathrm{VH}(H\to c\bar{c})} = 21^{+26}_{-24}$ in the fit. The post-fit expected yields for the background and signal processes and the observation in data for the signal regions are summarized in Table 9.1.

The systematic uncertainties discussed in Section 8.6 are taken into account in the fit via nuisance parameters, which allow for variations in the shape and normalization of the m_{SD} distributions during the fit. Figure 9.9 shows the post-fit values of the nuisance parameters and their impacts on the VH($H \to c\bar{c}$) signal strength. Only the first 60

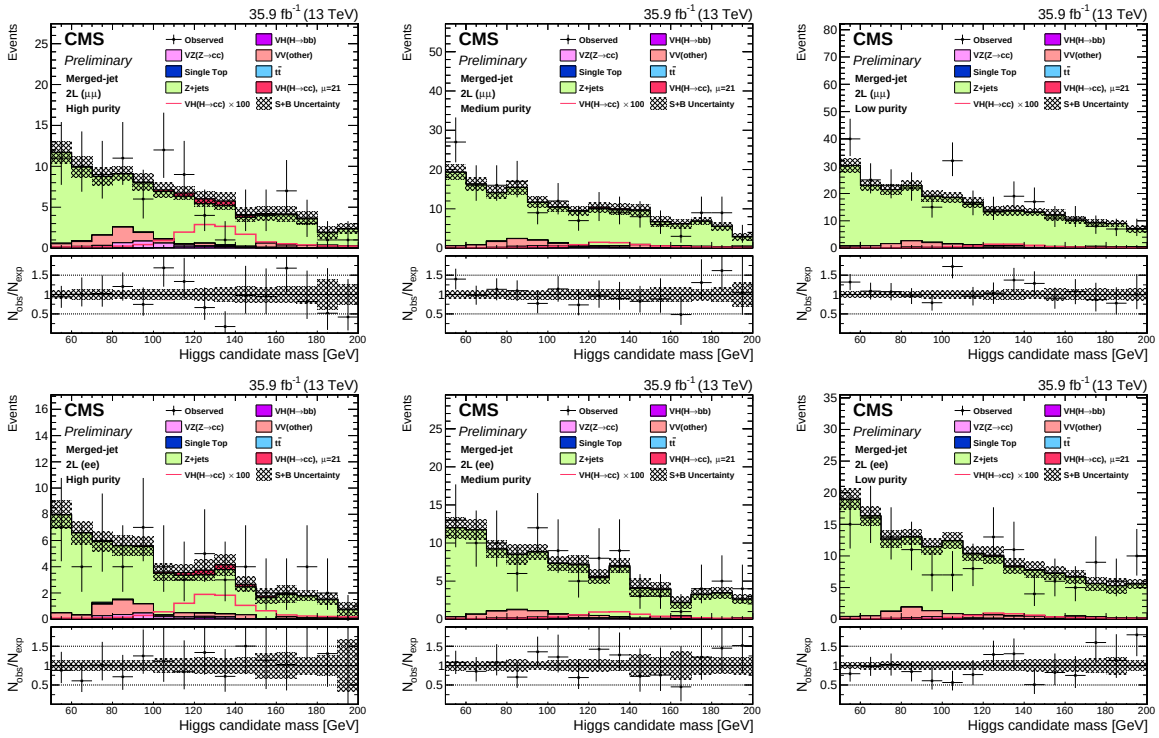


Figure 9.1: Post-fit m_{SD} distributions in the signal regions for the HP (left), MP (middle) and LP (right) categories of the 2-lepton channel. Top: $\mathrm{Z} \to \mu\mu$; Bottom: $\mathrm{Z} \to \mathrm{ee}$.

Process	High purity			Medium purity			Low purity		
	0-lepton	1-lepton	2-lepton	0-lepton	1-lepton	2-lepton	0-lepton	1-lepton	2-lepton
V+jets	777.8	860.6	126.6	1285.1	1431.9	228.5	2322.1	2470.8	363.0
$\mathrm{t\bar{t}}$	295.4	1073.3	0.7	309.0	1203.0	0.7	335.2	1488.4	0.7
Single top	36.7	312.2	0.1	44.4	323.8	0.2	50.0	389.8	0.3
$\mathrm{VZ(Z \to c\bar{c})}$	31.6	18.3	5.3	21.2	10.4	2.9	20.0	10.2	2.9
VV(other)	95.8	123.2	12.6	112.6	139.9	16.2	147.1	178.7	21.1
$\mathrm{VH(H \to b\bar{b})}$	4.2	4.9	1.1	4.8	6.3	1.6	5.0	6.5	1.5
Total background	1241.3	2392.4	146.4	1777.0	3115.3	250.0	2879.5	4544.4	389.6
$\mathrm{VH(H \to c\bar{c})}$ (SM)	16.5 (0.8)	20.7 (1.0)	4.5 (0.2)	9.7 (0.5)	12.1 (0.6)	2.5 (0.1)	9.9 (0.5)	11.8 (0.5)	2.5 (0.1)
Data	1329	2401	150	1833	3163	263	2959	4577	408

Table 9.1: Post-fit expected yields of background and signal processes and the observation in data in the signal regions of each channel. The $\mathrm{VH(H \to c\bar{c})}$ signal yields are scaled by the best-fit signal strength, $\mu_{\mathrm{VH(H \to c\bar{c})}} = 21^{+26}_{-24}$. The SM expected $\mathrm{VH(H \to c\bar{c})}$ signal yields are indicated in the parenthesis.

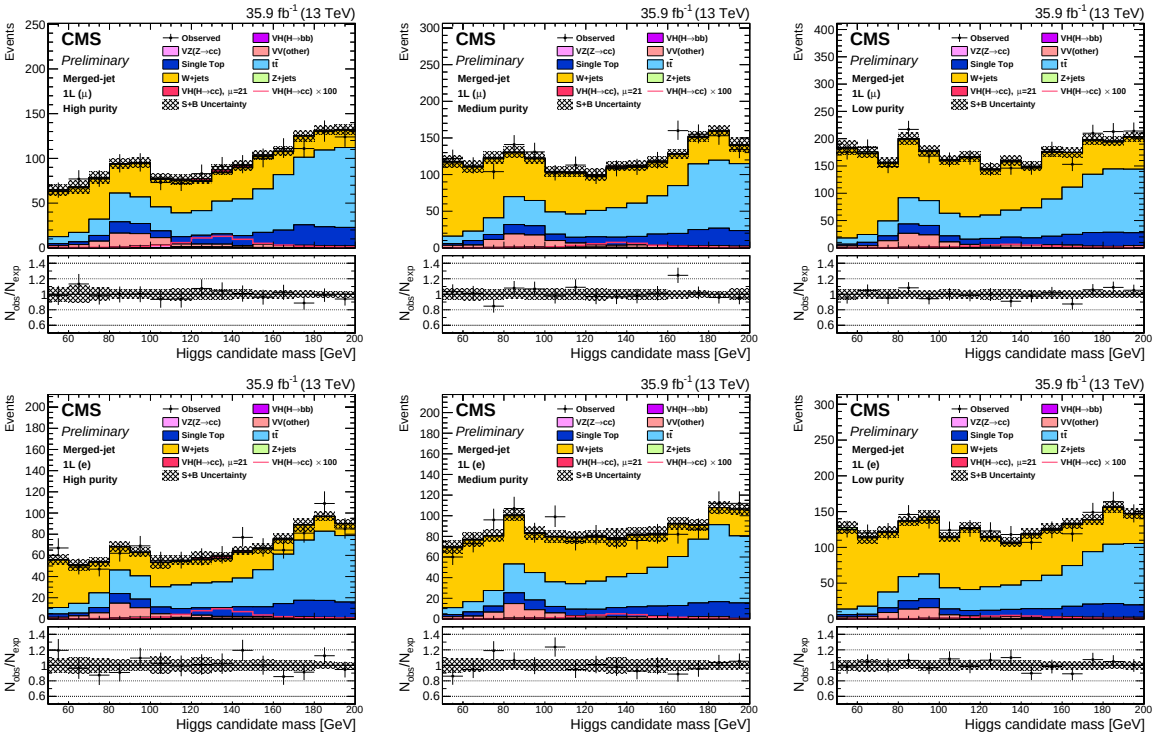


Figure 9.2: Post-fit m_{SD} distributions in the signal regions for the HP (left), MP (middle) and LP (right) categories of the 1-lepton channel. Top: W $\rightarrow \mu\nu$; Bottom: W $\rightarrow$ eν.

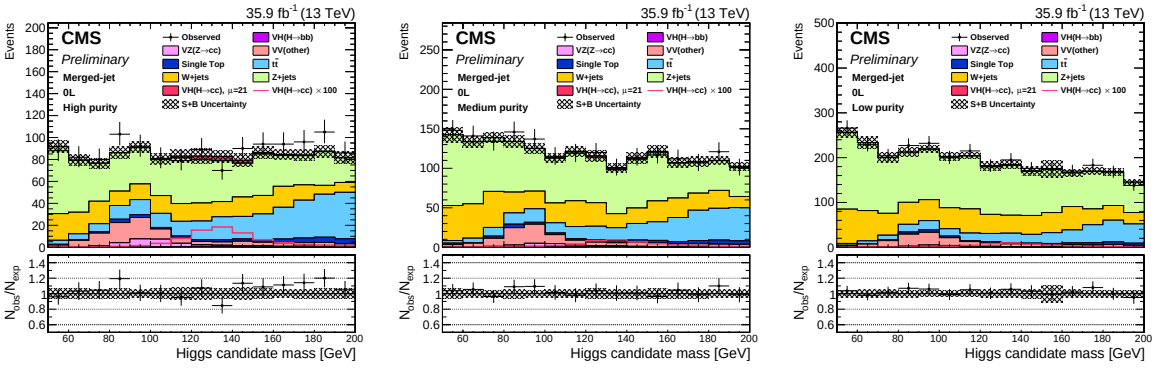


Figure 9.3: Post-fit m_{SD} distributions in the signal regions for the HP (left), MP (middle) and LP (right) categories of the 0-lepton channel.

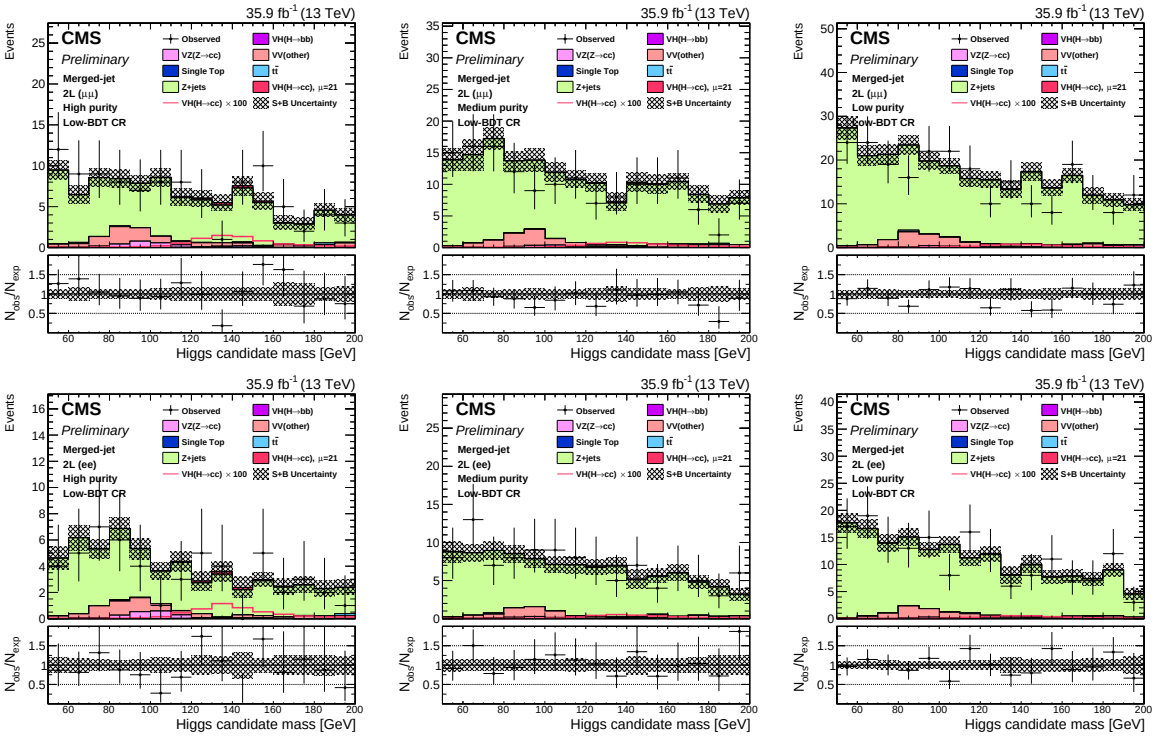


Figure 9.4: Post-fit m_{SD} distributions in the low-BDT control regions for the HP (left), MP (middle) and LP (right) categories of the 2-lepton channel. Top: $\mathrm{Z} \to \mu\mu$; Bottom: $\mathrm{Z} \to \mathrm{ee}$.

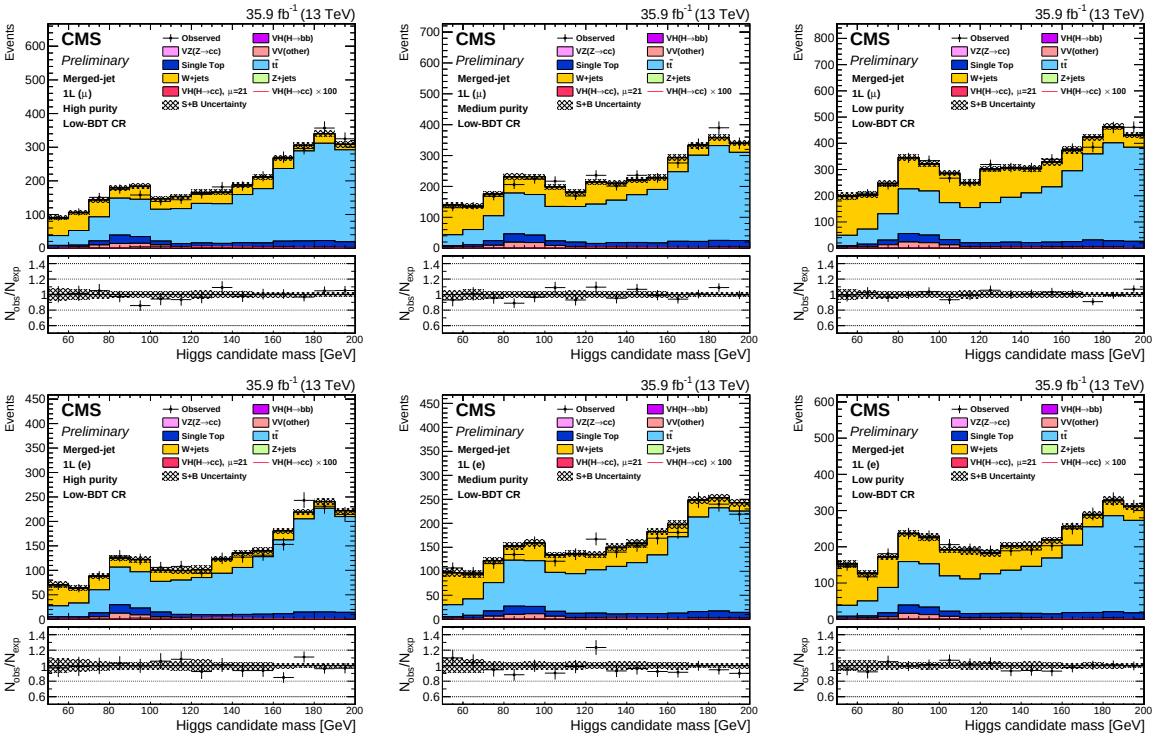


Figure 9.5: Post-fit m_{SD} distributions in the low-BDT control regions for the HP (left), MP (middle) and LP (right) categories of the 1-lepton channel. Top: $\mathrm{W} \to \mu\nu$; Bottom: $\mathrm{W} \to e\nu$.

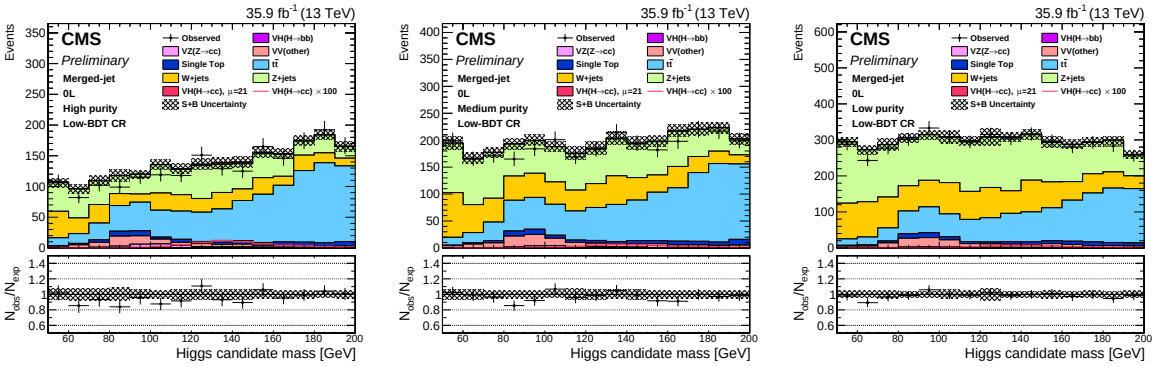


Figure 9.6: Post-fit m_{SD} distributions in the low-BDT control regions for HP (left), MP (middle) and LP (right) categories of the 0-lepton channel.

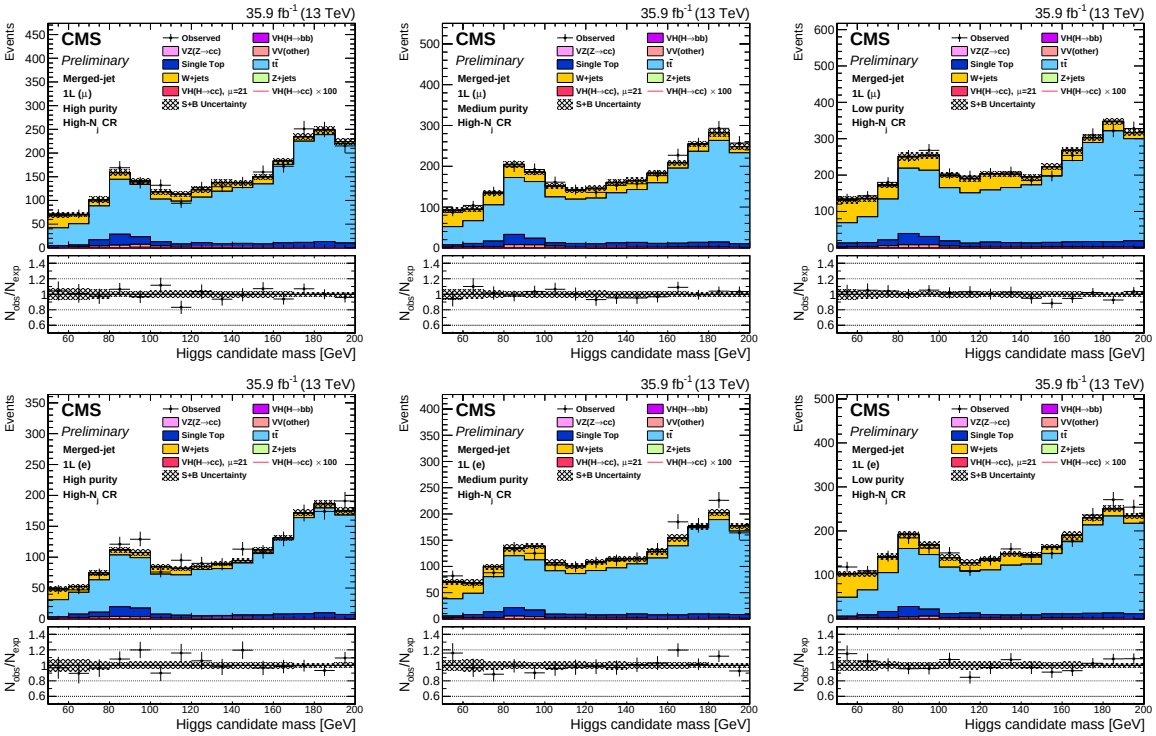


Figure 9.7: Post-fit m_{SD} distributions in the high-N_{j} control regions for the HP (left), MP (middle) and LP (right) categories of the 1-lepton channel. Top: W $\rightarrow \mu\nu$; Bottom: W $\rightarrow$ eν.

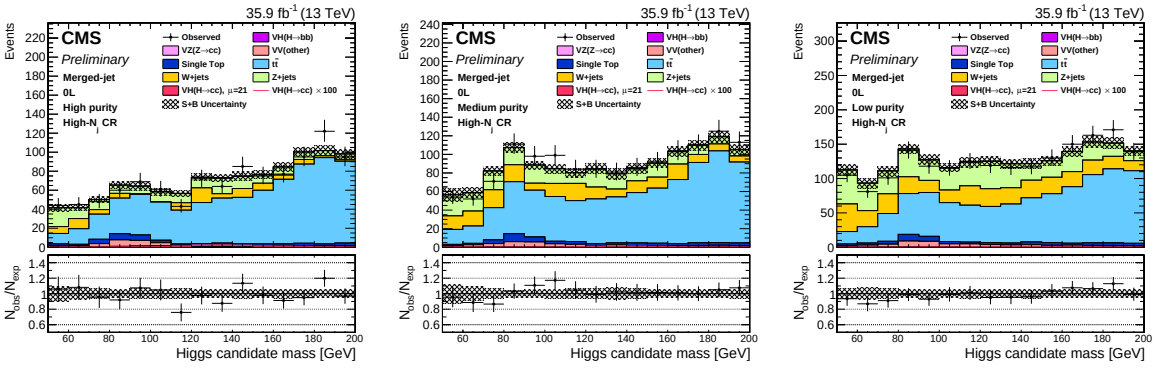


Figure 9.8: Post-fit m_{SD} distributions in the high-N_{j} control regions for the HP (left), MP (middle) and LP (right) categories of the 0-lepton channel.

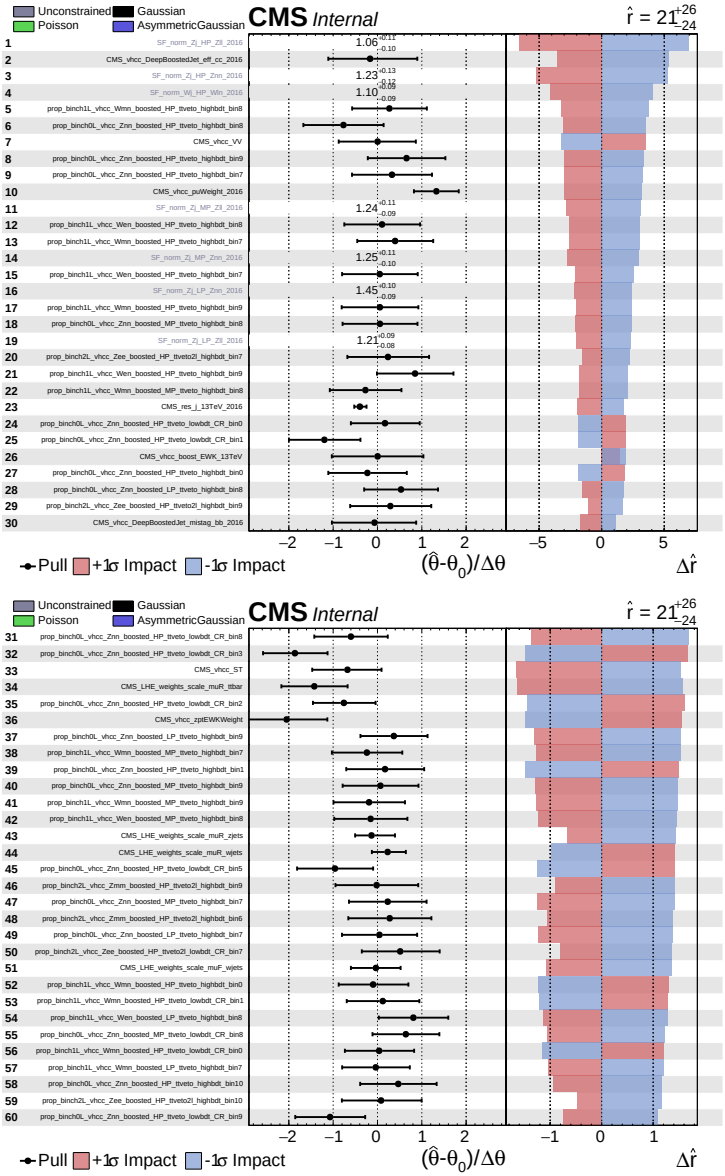


Figure 9.9: Post-fit values of the nuisance parameters and their impacts on the extracted VH($H \to c\bar{c}$) signal strength for the first 60 nuisance parameters with the highest impact on the signal strength. The pulls of the nuisance parameters (black markers with error bars) are computed as the differences between the best-fit values $\hat{\theta}$ and the pre-fit values θ_0, and then divided by the pre-fit uncertainties $\Delta\theta$. The impact $\Delta\hat{r}$ for each nuisance parameter is computed as the difference between the best-fit signal strength value of the nominal fit, $\hat{r}$, and the best-fit value obtained when the nuisance parameter under scrutiny is fixed to its best-fit value plus/minus its post-fit uncertainty.

nuisance parameters with the highest impacts on the VH($H \to c\bar{c}$) signal strength are shown. The best-fit values of the nuisance parameters are typically within 1 standard deviation of the pre-fit values. The nuisance parameter related to the $p_T(V)$ reweighting ("`CMS_vhcc_zptEWKWeight`") is pulled by around 2 standard deviations though, due to some residual mismodelling of the event kinematics in the simulated V+jets samples. As shown in the right panel of Figure 9.9, uncertainties associated with the background normalizations obtained from the control regions, the limited sizes of the simulated samples, and the efficiency of the DeepAK15 $c\bar{c}$-tagging discriminant have the largest impacts on the extracted signal strength.

The full procedure of this analysis is validated by measuring the cross section of VZ production with subsequent $Z \to c\bar{c}$ decay. The VZ($Z \to c\bar{c}$) process is similar to the VH($H \to c\bar{c}$) process and has a much higher cross section, therefore providing a good test bed of the analysis strategy. The event selection criteria, including the baseline selection, the kinematic BDT, and the three categories based on the $c\bar{c}$-tagging discriminant, as well as the signal extraction procedure, are the same in the nominal VH($H \to c\bar{c}$) analysis. The only difference is that the VZ($Z \to c\bar{c}$) process is treated as the signal instead of the VH($H \to c\bar{c}$) process, while the VH($H \to c\bar{c}$) process is considered as a background and fixed to its SM prediction. The best-fit signal strength for VZ($Z \to c\bar{c}$) is found to be $\mu_{VZ(Z\to c\bar{c})} = 0.69^{+0.89}_{-0.75}$, consistent with the SM expectation ($\mu_{VZ(Z\to c\bar{c})} = 1$) within uncertainty. The observed (expected) significance for the VZ($Z \to c\bar{c}$) process is 0.9 (1.3) standard deviations.

As no statistically significant excess is observed above the predicted background in the search of the SM VH($H \to c\bar{c}$) production, an upper limit (UL) on the signal strength $\mu_{VH(H\to c\bar{c})}$ at 95% confidence level (CL) is extracted based on a modified frequentist approach using the CLs criterion [201, 202] under the asymptotic approximation for the test statistic [203, 204]. The observed (expected) UL on $\mu_{VH(H\to c\bar{c})}$ is found to be 71

(49^{+24}_{-15}) at 95% CL. The results for each channel and their combination are shown in Table 9.2.

	0-lepton	1-lepton	2-lepton	Combined
Expected UL	81^{+39}_{-24}	88^{+43}_{-27}	90^{+48}_{-29}	49^{+24}_{-15}
Observed UL	74	120	76	71

Table 9.2: Observed (expected) UL at 95% CL on the signal strength $\mu_{\mathrm{VH}(\mathrm{H}\to\mathrm{c\bar{c}})}$.

Chapter 10

Conclusions and Outlook

In this thesis, a search for the standard model Higgs boson decaying to charm quarks with the CMS experiment is presented. The search targets the Higgs boson produced in association with a W or Z boson, whose leptonic decays provide a clear signature for suppressing the multijet background as well as for event triggering. The search is designed for a phase space where the Higgs boson is produced with a fairly high Lorentz boost ($p_{\mathrm{T}} > 200\,\mathrm{GeV}$). The dominant background to this search, such as W+jets and Z+jets, is more suppressed at higher p_{T}. A novel approach for the reconstruction and identification of the boosted $\mathrm{H} \to \mathrm{c\overline{c}}$ candidates is developed for this search. With a high boost, the two quarks from the Higgs decay become collimated and can be contained in a single large-radius jet. Therefore, a collection of large-radius jets, formed by clustering the particle-flow candidates using the anti-k_{T} algorithm with a distance parameter $R = 1.5$, is used to reconstruct both quarks from the Higgs decay as one jet. The large distance parameter is carefully chosen to ensure that a good fraction of the boosted $\mathrm{H} \to \mathrm{c\overline{c}}$ candidate can be reconstructed in this way. This approach not only has the benefit of exploiting the high-momentum phase space which has a higher signal purity, but can also potentially improves the identification of the $\mathrm{H} \to \mathrm{c\overline{c}}$ decay since the correlations

between the two charm quark showers can be better exploited now that they are clustered into the same jet.

A major challenge of this analysis is the identification of the charm quark pair from the Higgs boson decay. While the charm hadrons from the fragmentation and hadronization of the charm quarks can give rise to displaced tracks and secondary vertices, similar to bottom quarks, these signatures are not as distinct as that of the bottom quarks. Therefore, the identification of charm quarks is more subtle. In this search, an advanced deep learning–based algorithm is adopted for this task. This algorithm is a tailored version of the DeepAK8 algorithm, which features the use of all particle-flow candidates and reconstructed secondary vertices associated with a large-radius jet as inputs. The direct use of the full set of low-level information of a jet, together with customized neural network architecture, results in state-of-the-art performance for charm quark pair identification.

This advanced $\mathrm{H} \to \mathrm{c\overline{c}}$ reconstruction and identification method has led to highly competitive results. Using proton-proton collision data collected in the 2016 run of the LHC that correspond to an integrated luminosity of $35.9\,\mathrm{fb}^{-1}$, an observed (expected) upper limit on the VH production cross section times the $\mathrm{H} \to \mathrm{c\overline{c}}$ branching fraction of 71 (49) times the SM expectation is obtained at 95% confidence level, significantly surpassing previous experimental limits. Combined with the CMS search using two resolved jets to reconstruct the $\mathrm{H} \to \mathrm{c\overline{c}}$ decay, an observed (expected) upper limit of 70 (37) times the SM is obtained, providing the world's most stringent limit to date on the Higgs coupling to charm quarks obtained from direct searches.

The identification of the hadronically decaying boosted particles (top quarks, W, Z and Higgs bosons) itself is also an important topic at the LHC. On the one hand, it is a powerful tool for many searches for heavy (TeV-scale) exotic resonances, as these resonances can possibly be produced directly at the LHC and then decay to highly boosted

SM particles. On the other hand, it can also be exploited to probe the SM in the high-momentum phase space, a territory that cannot be fully explored without such techniques. In this thesis, the development and evaluation of a deep learning–based boosted jet tagging algorithm, DeepAK8, is discussed in detail. The DeepAK8 algorithm is designed as a versatile multi-class classifier for the identification of boosted top quarks, W, Z and Higgs bosons via large-radius ($R = 0.8$) jets. It aims at exploiting the substructure information (such as energy distribution patterns, energy correlations between particles) and the flavor information (such as the presence of displaced tracks, secondary vertices, which are characteristics of heavy-flavor quark fragmentation) simultaneously. While jet substructure–based observables have been extensively used for boosted jet tagging, the flavor information has not been so widely used yet, but it is actually very powerful for identifying particle decays involving heavy flavor quarks, such as the top quark (whose decay always gives rise to a bottom quark), or Z or Higgs bosons decaying to $\mathrm{b\overline{b}}$ or $\mathrm{c\overline{c}}$. The two aspects typically do not fully factorize; thus, a multivariate approach is needed to maximize the performance.

The rise of machine learning, particularly deep learning, in recent years, has clearly boosted the development of new and more powerful jet tagging techniques. In the DeepAK8 algorithm, a customized deep neural network is developed to access and process the constituents of a jet directly for jet tagging. A significant performance improvement compared to traditional approaches based on jet-level observables has been observed. A dedicated version of DeepAK8 whose outputs are largely independently of the jet mass is also developed, allowing this high-performance algorithm to be used by analyses that tolerate no or very little change in the background mass distribution. The performance of the DeepAK8 algorithm has been validated in data, and in general, good agreement has been observed. The development of the DeepAK8 algorithm has demonstrated that, with a good understanding of the underlying physics and a careful design and validation pro-

cess, advances in machine learning techniques are likely to bring significant improvements to particle physics experiments.

A more recent development for jet tagging is also presented in this thesis. A jet is treated as a "particle cloud", i.e., an unordered set of particles with "coordinates" in the (η, ϕ) space and other properties such as momentum, charge, etc. This jet representation is inspired by the "point cloud" concept in computer vision, which refers to a set of points in space, typically measured by a 3D scanner, to represent the external surfaces of objects around the 3D scanner. A key feature of both point clouds or particle clouds is that the elements, i.e., points or particles, in the cloud are intrinsically unordered. Based on the particle cloud representation of jets, ParticleNet, a new graph neural network architecture is developed to process the particles in a manifestly permutation-invariant way and perform jet tagging. Benchmarked on two open simulation datasets, the ParticleNet model achieves significant improvement in performance, reducing the misidentification rate by almost a factor of two compared to the neural network architecture used in DeepAK8.

Despite the impressive progress in machine learning–based boosted jet tagging techniques, several challenges remain. One major challenge is related to the discrepancy between data and simulation, mainly in the modeling of parton showering and hadronization in the event generators. Particularly, gluons are less well modeled compared to quarks in general in current event generators. Such discrepancies pose serious challenges to many analyses for the estimation of background contributions, as well as for the determination of the signal efficiency. Possible ways to cope with this challenge include, e.g., systematic improvement of the event generators, development of new ML-based algorithms that are insensitive to such discrepancy, or more clever methods for background estimation and tagging efficiency calibration. Another challenge is the development of mass-decorrelated tagging algorithms. Forcing the classifier outputs to be independent of one of the most powerful variables, the jet mass, tends to affect the performance adversely, leading to

highly volatile training. While the current mass-decorrelated version of DeepAK8 shows overall excellent performance and mass decorrelation, there is still much room for future improvements. New ideas based on both ML techniques and theoretical considerations are likely to be crucial for this.

With a current upper limit of about 70 times the SM expectation, there is obviously a long way ahead until the Higgs coupling to charm quarks can actually be measured. Improvements in the charm pair identification techniques, the background estimation and signal extraction strategy, and the exploitation of other Higgs production channels such as the vector boson fusion (VBF) or the associated production with a top quark pair ($\mathrm{t\bar{t}H}$), are all likely to contribute significantly to an improved result. Upgrades on the detector, particularly improvements in the tracking system, or the addition of a timing detector, are likely to bring a substantial improvement to the charm quark identification performance as well. So far, only a small amount of data corresponding to an integrated luminosity of $35.9\,\mathrm{fb}^{-1}$ has been explored. This is only about one-quarter of the available data collected in Run 2 of the LHC, and around 1% of the total data expected to be collected across the whole life of the LHC. Clearly, the charming journey has just started.

Huilin Qu

Geneva, Switzerland, September 2019

Bibliography

[1] M. Srednicki, *Quantum field theory*. Cambridge University Press, 2007. (Cited on page 4.)

[2] M. D. Schwartz, *Quantum Field Theory and the Standard Model*. Cambridge University Press, 2014. (Cited on page 4.)

[3] S. Weinberg, *The Quantum theory of fields. Vol. 1: Foundations*. Cambridge University Press, 2005. (Cited on page 4.)

[4] M. Thomson, *Modern particle physics*. Cambridge University Press, New York, 2013. (Cited on page 4.)

[5] W. N. Cottingham and D. A. Greenwood, *An introduction to the standard model of particle physics*. Cambridge University Press, 2007. (Cited on page 4.)

[6] I. Neutelings, "Timeline." https://wiki.physik.uzh.ch/cms/latex:tikz:timescales. (Cited on page 4.)

[7] A. Purcell, *Go on a particle quest at the first CERN webfest. Le premier webfest du CERN se lance à la conquête des particules*, Aug, 2012. https://cds.cern.ch/record/1473657. (Cited on page 5.)

[8] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, *Phys. Rev. Lett.* **13** (1964) 321. (Cited on page 7.)

[9] P. W. Higgs, *Broken symmetries, massless particles and gauge fields*, *Phys. Lett.* **12** (1964) 132. (Cited on page 7.)

[10] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, *Phys. Rev. Lett.* **13** (1964) 508. (Cited on page 7.)

[11] G. S. Guralnik, C. R. Hagen and T. W. B. Kibble, *Global Conservation Laws and Massless Particles*, *Phys. Rev. Lett.* **13** (1964) 585. (Cited on page 7.)

[12] P. W. Higgs, *Spontaneous Symmetry Breakdown without Massless Bosons*, *Phys. Rev.* **145** (1966) 1156. (Cited on page 7.)

[13] T. W. B. Kibble, *Symmetry breaking in non-Abelian gauge theories*, *Phys. Rev.* **155** (1967) 1554. (Cited on page 7.)

[14] Particle Data Group collaboration, *Review of Particle Physics*, *Phys. Rev.* **D98** (2018) 030001. (Cited on pages 9 and 10.)

[15] LHC Higgs Cross Section Working Group collaboration, *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, 1610.07922. (Cited on pages 10, 11, 13, 120, and 156.)

[16] O. S. Brüning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, J. Poole et al., *LHC Design Report. v.1 : the LHC Main Ring*, CERN Yellow Reports: Monographs. CERN, Geneva, 2004, 10.5170/CERN-2004-003-V-1. (Cited on page 15.)

[17] O. S. Brüning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, J. Poole et al., *LHC Design Report. v.2 : the LHC Infrastructure and General Services*, CERN Yellow Reports: Monographs. CERN, Geneva, 2004, 10.5170/CERN-2004-003-V-2. (Cited on page 15.)

[18] M. Benedikt, P. Collier, V. Mertens, J. Poole and K. Schindl, *LHC Design Report. v.3 : the LHC Injector Chain*, CERN Yellow Reports: Monographs. CERN, Geneva, 2004, 10.5170/CERN-2004-003-V-3. (Cited on page 15.)

[19] L. Evans and P. Bryant, *LHC Machine*, *JINST* **3** (2008) S08001. (Cited on page 15.)

[20] C. Lefèvre, *The CERN accelerator complex. Complexe des accélérateurs du CERN*, Dec, 2008. https://cds.cern.ch/record/1260465. (Cited on page 15.)

[21] CMS collaboration, T. Sakuma, *Cutaway diagrams of CMS detector*, May, 2019. https://cds.cern.ch/record/2665537. (Cited on page 17.)

[22] CMS collaboration, *The CMS Experiment at the CERN LHC*, *JINST* **3** (2008) S08004. (Cited on pages 17 and 21.)

[23] I. Neutelings, "Example spherical coordinates." https://wiki.physik.uzh.ch/cms/latex:example_spherical_coordinates. (Cited on page 18.)

[24] CMS collaboration, *The CMS tracker system project: Technical Design Report*, Tech. Rep. CERN-LHCC-98-006, 1997. (Cited on page 18.)

[25] CMS collaboration, *The CMS tracker: addendum to the Technical Design Report*, Tech. Rep. CERN-LHCC-2000-016, 2000. (Cited on page 18.)

[26] CMS collaboration, *Description and performance of track and primary-vertex reconstruction with the CMS tracker*, *JINST* **9** (2014) P10009 [1405.6569]. (Cited on pages 19, 25, and 26.)

[27] CMS collaboration, *Primary vertex resolution in 2016*, Tech. Rep. CMS-DP-2016-041, Jul, 2016. (Cited on page 19.)

[28] CMS collaboration, *The CMS electromagnetic calorimeter project: Technical Design Report*, Tech. Rep. CERN-LHCC-97-033, 1997. (Cited on page 20.)

[29] CMS collaboration, *Changes to CMS ECAL electronics: addendum to the Technical Design Report*, Tech. Rep. CERN-LHCC-2002-027, 2002. (Cited on page 20.)

[30] P. Adzic et al., *Energy resolution of the barrel of the CMS electromagnetic calorimeter*, *JINST* **2** (2007) P04004. (Cited on page 20.)

[31] CMS collaboration, *The CMS hadron calorimeter project: Technical Design Report*, Tech. Rep. CERN-LHCC-97-031, 1997. (Cited on page 21.)

[32] CMS ECAL/HCAL collaboration, *The CMS barrel calorimeter response to particle beams from 2-GeV/c to 350-GeV/c*, *Eur. Phys. J.* **C60** (2009) 359. (Cited on page 22.)

[33] CMS collaboration, *The Performance of the CMS Muon Detector in Proton-Proton Collisions at $\sqrt{s}$ = 7 TeV at the LHC*, *JINST* **8** (2013) P11002 [1306.6905]. (Cited on pages 22 and 23.)

[34] CMS collaboration, *The CMS muon project: Technical Design Report*, Tech. Rep. CERN-LHCC-97-032, 1997. (Cited on page 22.)

[35] CMS collaboration, *CMS. The TriDAS project. Technical design report, vol. 1: The trigger systems*, Tech. Rep. CERN-LHCC-2000-038, 2000. (Cited on page 24.)

[36] CMS collaboration, *CMS The TriDAS Project: Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger*, Tech. Rep. CERN-LHCC-2002-026, 2002. (Cited on page 24.)

[37] CMS collaboration, *Particle-flow reconstruction and global event description with the CMS detector*, *JINST* **12** (2017) P10003 [1706.04965]. (Cited on pages 25 and 98.)

[38] W. Adam, B. Mangano, T. Speer and T. Todorov, *Track Reconstruction in the CMS tracker*, Tech. Rep. CMS-NOTE-2006-041, Dec, 2006. (Cited on page 25.)

[39] W. Adam, R. Fruhwirth, A. Strandlie and T. Todorov, *Reconstruction of electrons with the Gaussian sum filter in the CMS tracker at LHC*, *eConf* **C0303241** (2003) TULT009 [physics/0306087]. (Cited on page 28.)

[40] M. Cacciari, G. P. Salam and G. Soyez, *The anti-k_t jet clustering algorithm*, *JHEP* **04** (2008) 063 [0802.1189]. (Cited on pages 33, 107, and 109.)

[41] M. Cacciari, G. P. Salam and G. Soyez, *FastJet User Manual*, *Eur. Phys. J.* **C72** (2012) 1896 [1111.6097]. (Cited on page 33.)

[42] CMS collaboration, *Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV*, *JINST* **12** (2017) P02014 [1607.03663]. (Cited on pages 34, 35, 124, and 153.)

[43] D. Bertolini, P. Harris, M. Low and N. Tran, *Pileup Per Particle Identification*, *JHEP* **10** (2014) 059 [1407.6013]. (Cited on pages 34, 56, and 123.)

[44] CMS collaboration, *Pileup mitigation at CMS in 13 TeV data*, Tech. Rep. CMS-PAS-JME-18-001, 2019. (Cited on page 35.)

[45] CMS collaboration, *Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13$ TeV using the CMS detector*, *JINST* **14** (2019) P07004 [1903.06078]. (Cited on pages 36 and 153.)

[46] CMS collaboration, *Search for direct production of supersymmetric partners of the top quark in the all-jets final state in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *JHEP* **10** (2017) 005 [1707.03316]. (Cited on pages 40, 49, 53, and 54.)

[47] CMS collaboration, *Search for production of Higgs boson pairs in the four b quark final state using large-area jets in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *JHEP* **01** (2019) 040 [1808.01473]. (Cited on page 40.)

[48] CMS collaboration, *Search for resonant $\mathrm{t\bar{t}}$ production in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *JHEP* **04** (2019) 031 [1810.05905]. (Cited on page 40.)

[49] CMS collaboration, *A multi-dimensional search for new heavy resonances decaying to boosted WW, WZ, or ZZ boson pairs in the dijet final state at 13 TeV*, 1906.05977. (Cited on page 40.)

[50] CMS collaboration, *Measurement of the integrated and differential $t\bar{t}$ production cross sections for high-p_t top quarks in pp collisions at $\sqrt{s} = 8$ TeV*, *Phys. Rev.* **D94** (2016) 072002 [1605.00116]. (Cited on page 40.)

[51] CMS collaboration, *Inclusive search for a highly boosted Higgs boson decaying to a bottom quark-antiquark pair*, *Phys. Rev. Lett.* **120** (2018) 071802 [1709.05543]. (Cited on page 40.)

[52] A. J. Larkoski, I. Moult and B. Nachman, *Jet Substructure at the Large Hadron Collider: A Review of Recent Advances in Theory and Machine Learning*, 1709.04464. (Cited on page 40.)

[53] L. Asquith et al., *Jet Substructure at the Large Hadron Collider : Experimental Review*, 1803.06991. (Cited on page 40.)

[54] S. Marzani, G. Soyez and M. Spannowsky, *Looking inside jets: an introduction to jet substructure and boosted-object phenomenology*, 1901.10342. (Cited on page 40.)

[55] M. Dasgupta, A. Fregoso, S. Marzani and G. P. Salam, *Towards an understanding of jet substructure*, *JHEP* **09** (2013) 029 [1307.0007]. (Cited on pages 41 and 123.)

[56] A. J. Larkoski, S. Marzani, G. Soyez and J. Thaler, *Soft Drop*, *JHEP* **05** (2014) 146 [1402.2657]. (Cited on pages 41 and 123.)

[57] Y. L. Dokshitzer, G. D. Leder, S. Moretti and B. R. Webber, *Better jet clustering algorithms*, *JHEP* **08** (1997) 001 [hep-ph/9707323]. (Cited on page 41.)

[58] M. Wobisch and T. Wengler, *Hadronization corrections to jet cross-sections in deep inelastic scattering*, in *Monte Carlo generators for HERA physics. Proceedings, Workshop, Hamburg, Germany, 1998-1999*, pp. 270–279, 1998, hep-ph/9907280. (Cited on page 41.)

[59] J. Thaler and K. Van Tilburg, *Identifying Boosted Objects with N-subjettiness*, *JHEP* **03** (2011) 015 [1011.2268]. (Cited on page 43.)

[60] J. Thaler and K. Van Tilburg, *Maximizing Boosted Top Identification by Minimizing N-subjettiness*, *JHEP* **02** (2012) 093 [1108.2701]. (Cited on page 43.)

[61] S. Catani, Y. L. Dokshitzer, M. H. Seymour and B. R. Webber, *Longitudinally invariant K_t clustering algorithms for hadron hadron collisions*, *Nucl. Phys.* **B406** (1993) 187. (Cited on page 43.)

[62] S. D. Ellis and D. E. Soper, *Successive combination jet algorithm for hadron collisions*, *Phys. Rev.* **D48** (1993) 3160 [hep-ph/9305266]. (Cited on page 43.)

[63] https://wordstream-files-prod.s3.amazonaws.com/s3fs-public/machine-learning.png. (Cited on page 45.)

[64] E. M. Metodiev and J. Thaler, *On the Topic of Jets: Disentangling Quarks and Gluons at Colliders*, *Phys. Rev. Lett.* **120** (2018) 241602 [1802.00008]. (Cited on page 45.)

[65] P. T. Komiske, E. M. Metodiev and J. Thaler, *An operational definition of quark and gluon jets*, *JHEP* **11** (2018) 059 [1809.01140]. (Cited on page 45.)

[66] S. Carrazza and F. A. Dreyer, *Jet grooming through reinforcement learning*, *Phys. Rev.* **D100** (2019) 014014 [1903.09644]. (Cited on page 45.)

[67] Y. LeCun, Y. Bengio and G. Hinton, *Deep learning*, *Nature* **521** (2015) 436. (Cited on page 45.)

[68] A. Zhang, Z. C. Lipton, M. Li and A. J. Smola, *Dive into Deep Learning*. 2019. (Cited on page 46.)

[69] I. Goodfellow, Y. Bengio and A. Courville, *Deep Learning*. MIT Press, 2016. (Cited on page 46.)

[70] V. Nair and G. E. Hinton, *Rectified Linear Units Improve Restricted Boltzmann Machines*, in *Proceedings of the 27th International Conference on International Conference on Machine Learning*, ICML'10, pp. 807–814, 2010. (Cited on page 47.)

[71] D.-A. Clevert, T. Unterthiner and S. Hochreiter, *Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs)*, 1511.07289. (Cited on page 47.)

[72] G. Klambauer, T. Unterthiner, A. Mayr and S. Hochreiter, *Self-Normalizing Neural Networks*, 1706.02515. (Cited on page 47.)

[73] A. L. Maas, A. Y. Hannun and A. Y. Ng, *Rectifier nonlinearities improve neural network acoustic models*, in *ICML*, vol. 30, 2013. (Cited on page 47.)

[74] K. He, X. Zhang, S. Ren and J. Sun, *Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification*, in *Proceedings of the 2015 IEEE International Conference on Computer Vision*, ICCV '15, pp. 1026–1034, 2015. (Cited on page 47.)

[75] D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, 1412.6980. (Cited on pages 48 and 66.)

[76] N. Qian, *On the momentum term in gradient descent learning algorithms*, *Neural Networks* **12** (1999) 145. (Cited on page 48.)

[77] L. N. Smith, *A disciplined approach to neural network hyper-parameters: Part 1 - learning rate, batch size, momentum, and weight decay*, 1803.09820. (Cited on pages 48 and 105.)

[78] CMS collaboration, *Performance of quark/gluon discrimination in 8 TeV pp data*, Tech. Rep. CMS-PAS-JME-13-002, 2013. (Cited on page 50.)

[79] CMS collaboration, *Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV*, *JINST* **13** (2018) P05011 [1712.07158]. (Cited on pages 51, 59, 70, and 126.)

[80] A. Hocker et al., *TMVA - Toolkit for Multivariate Data Analysis*, physics/0703039. (Cited on page 53.)

[81] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, *JHEP* **07** (2014) 079 [1405.0301]. (Cited on pages 58 and 120.)

[82] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten et al., *An Introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* **191** (2015) 159 [1410.3012]. (Cited on pages 58, 107, and 121.)

[83] NNPDF collaboration, *Parton distributions for the LHC Run II*, *JHEP* **04** (2015) 040 [1410.8849]. (Cited on pages 58 and 121.)

[84] R. D. Ball et al., *Parton distributions with LHC data*, *Nucl. Phys.* **B867** (2013) 244 [1207.1303]. (Cited on page 59.)

[85] CMS collaboration, *Event generator tunes obtained from underlying event and multiparton scattering measurements*, *Eur. Phys. J.* **C76** (2016) 155 [1512.00815]. (Cited on page 59.)

[86] GEANT4 collaboration, *GEANT4: A Simulation toolkit*, *Nucl. Instrum. Meth.* **A506** (2003) 250. (Cited on pages 59 and 121.)

[87] CMS collaboration, *Measurement of $B\bar{B}$ Angular Correlations based on Secondary Vertex Reconstruction at $\sqrt{s} = 7$ TeV*, *JHEP* **03** (2011) 136 [1102.3194]. (Cited on page 59.)

[88] K. He, X. Zhang, S. Ren and J. Sun, *Deep residual learning for image recognition*, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016. (Cited on pages 66, 98, and 104.)

[89] X. Glorot, A. Bordes and Y. Bengio, *Deep sparse rectifier neural networks*, in *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pp. 315–323, 2011. (Cited on pages 66 and 104.)

[90] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, *Dropout: a simple way to prevent neural networks from overfitting*, *The Journal of Machine Learning Research* **15** (2014) 1929. (Cited on pages 66 and 105.)

[91] T. Chen, M. Li, Y. Li, M. Lin, N. Wang, M. Wang et al., *MXNet: A Flexible and Efficient Machine Learning Library for Heterogeneous Distributed Systems*, 1512.01274. (Cited on pages 66 and 105.)

[92] G. Louppe, M. Kagan and K. Cranmer, *Learning to Pivot with Adversarial Networks*, in *Advances in Neural Information Processing Systems 30*, p. 981, 2017, http://papers.nips.cc/paper/6699-learning-to-pivot-with-adversarial-networks.pdf. (Cited on page 67.)

[93] CMS collaboration, *Search for low-mass resonances decaying into bottom quark-antiquark pairs in proton-proton collisions at $\sqrt{s}$ = 13 TeV*, *Phys. Rev.* **D99** (2019) 012005 [1810.11822]. (Cited on page 78.)

[94] CMS collaboration, *Machine learning-based identification of highly Lorentz-boosted hadronically decaying particles at the CMS experiment*, Tech. Rep. CMS-PAS-JME-18-002, CERN, 2019. (Cited on pages 82, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, and 96.)

[95] CMS collaboration, *Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s}$ = 13 TeV*, *JINST* **13** (2018) P06015 [1804.04528]. (Cited on pages 82, 83, and 122.)

[96] CMS collaboration, *Performance of Photon Reconstruction and Identification with the CMS Detector in Proton-Proton Collisions at sqrt(s) = 8 TeV*, *JINST* **10** (2015) P08010 [1502.02702]. (Cited on page 84.)

[97] H. Qu and L. Gouskos, *ParticleNet: Jet Tagging via Particle Clouds*, 1902.08570. (Cited on page 97.)

[98] Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein and J. M. Solomon, *Dynamic Graph CNN for Learning on Point Clouds*, 1801.07829. (Cited on pages 97, 100, 102, and 103.)

[99] ATLAS collaboration, *Jet reconstruction and performance using particle flow with the ATLAS Detector*, *Eur. Phys. J.* **C77** (2017) 466 [1703.10485]. (Cited on page 98.)

[100] J. Cogan, M. Kagan, E. Strauss and A. Schwartzman, *Jet-Images: Computer Vision Inspired Techniques for Jet Tagging*, *JHEP* **02** (2015) 118 [1407.5675]. (Cited on pages 98 and 108.)

[101] L. G. Almeida, M. Backović, M. Cliche, S. J. Lee and M. Perelstein, *Playing Tag with ANN: Boosted Top Identification with Pattern Recognition*, *JHEP* **07** (2015) 086 [1501.05968]. (Cited on page 98.)

[102] L. de Oliveira, M. Kagan, L. Mackey, B. Nachman and A. Schwartzman, *Jet-images – deep learning edition*, *JHEP* **07** (2016) 069 [1511.05190]. (Cited on pages 98 and 108.)

[103] P. Baldi, K. Bauer, C. Eng, P. Sadowski and D. Whiteson, *Jet Substructure Classification in High-Energy Physics with Deep Neural Networks*, *Phys. Rev.* **D93** (2016) 094034 [1603.09349]. (Cited on pages 98 and 108.)

[104] J. Barnard, E. N. Dawe, M. J. Dolan and N. Rajcic, *Parton Shower Uncertainties in Jet Substructure Analyses with Deep Neural Networks*, *Phys. Rev.* **D95** (2017) 014018 [1609.00607]. (Cited on pages 98 and 108.)

[105] F. A. Dreyer, G. P. Salam and G. Soyez, *The Lund Jet Plane*, *JHEP* **12** (2018) 064 [1807.04758]. (Cited on pages 98 and 108.)

[106] G. Kasieczka, T. Plehn, M. Russell and T. Schell, *Deep-learning Top Taggers or The End of QCD?*, *JHEP* **05** (2017) 006 [1701.08784]. (Cited on pages 98 and 108.)

[107] S. Macaluso and D. Shih, *Pulling Out All the Tops with Computer Vision and Deep Learning*, *JHEP* **10** (2018) 121 [1803.00107]. (Cited on pages 98 and 108.)

[108] S. Choi, S. J. Lee and M. Perelstein, *Infrared Safety of a Neural-Net Top Tagging Algorithm*, 1806.01263. (Cited on pages 98 and 108.)

[109] P. T. Komiske, E. M. Metodiev and M. D. Schwartz, *Deep learning in color: towards automated quark/gluon jet discrimination*, *JHEP* **01** (2017) 110 [1612.01551]. (Cited on pages 98 and 108.)

[110] ATLAS collaboration, *Quark versus Gluon Jet Tagging Using Jet Images with the ATLAS Detector*, Tech. Rep. ATL-PHYS-PUB-2017-017, CERN, Jul, 2017. (Cited on pages 98 and 108.)

[111] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, *Rethinking the inception architecture for computer vision*, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818–2826, 2016. (Cited on page 98.)

[112] D. Guest, J. Collado, P. Baldi, S.-C. Hsu, G. Urban and D. Whiteson, *Jet Flavor Classification in High-Energy Physics with Deep Neural Networks*, *Phys. Rev.* **D94** (2016) 112002 [1607.08633]. (Cited on page 99.)

[113] ATLAS collaboration, *Identification of Jets Containing b-Hadrons with Recurrent Neural Networks at the ATLAS Experiment*, Tech. Rep. ATL-PHYS-PUB-2017-003, CERN, Mar, 2017. (Cited on page 99.)

[114] J. Pearkes, W. Fedorko, A. Lister and C. Gay, *Jet Constituents for Deep Neural Network Based Top Quark Tagging*, 1704.02124. (Cited on page 99.)

[115] S. Egan, W. Fedorko, A. Lister, J. Pearkes and C. Gay, *Long Short-Term Memory (LSTM) networks with jet constituents for boosted top tagging at the LHC*, 1711.09059. (Cited on page 99.)

[116] K. Fraser and M. D. Schwartz, *Jet Charge and Machine Learning*, *JHEP* **10** (2018) 093 [1803.08066]. (Cited on page 99.)

[117] CMS collaboration, *CMS Phase 1 heavy flavour identification performance and developments*, Tech. Rep. CMS-DP-2017-013, May, 2017. (Cited on page 99.)

[118] CMS collaboration, *Performance of the DeepJet b tagging algorithm using 41.9/fb of data from proton-proton collisions at 13TeV with Phase 1 CMS detector*, Tech. Rep. CMS-DP-2018-058, Nov, 2018. (Cited on page 99.)

[119] CMS collaboration, *Boosted jet identification using particle candidates and deep neural networks*, Tech. Rep. CMS-DP-2017-049, Nov, 2017. (Cited on page 99.)

[120] M. Stoye, J. Kieseler, H. Qu, L. Gouskos and M. Verzetti, *DeepJet: Generic physics object based jet multiclass classification for LHC experiments*, in *Deep Learning for Physical Sciences Workshop at the 31st Conference on Neural Information Processing Systems (NIPS)*, 2017. (Cited on page 99.)

[121] A. Butter, G. Kasieczka, T. Plehn and M. Russell, *Deep-learned Top Tagging with a Lorentz Layer*, *SciPost Phys.* **5** (2018) 028 [1707.08966]. (Cited on pages 99 and 107.)

[122] G. Kasieczka, N. Kiefer, T. Plehn and J. M. Thompson, *Quark-Gluon Tagging: Machine Learning meets Reality*, 1812.09223. (Cited on page 99.)

[123] M. Erdmann, E. Geiser, Y. Rath and M. Rieger, *Lorentz Boost Networks: Autonomous Physics-Inspired Feature Engineering*, 1812.09722. (Cited on page 99.)

[124] G. Louppe, K. Cho, C. Becot and K. Cranmer, *QCD-Aware Recursive Neural Networks for Jet Physics*, *JHEP* **01** (2019) 057 [1702.00748]. (Cited on page 99.)

[125] T. Cheng, *Recursive Neural Networks in Quark/Gluon Tagging*, *Comput. Softw. Big Sci.* **2** (2018) 3 [1711.02633]. (Cited on page 99.)

[126] P. T. Komiske, E. M. Metodiev and J. Thaler, *Energy Flow Networks: Deep Sets for Particle Jets*, *JHEP* **01** (2019) 121 [1810.05165]. (Cited on pages 100, 108, 109, and 111.)

[127] M. Zaheer, S. Kottur, S. Ravanbakhsh, B. Poczos, R. R. Salakhutdinov and A. J. Smola, *Deep sets*, in *Advances in neural information processing systems*, pp. 3391–3401, 2017. (Cited on pages 100 and 108.)

[128] I. Henrion, J. Brehmer, J. Bruna, K. Cho, K. Cranmer, G. Louppe et al., *Neural Message Passing for Jet Physics*, in *Deep Learning for Physical Sciences Workshop at the 31st Conference on Neural Information Processing Systems (NIPS)*, 2017. (Cited on page 100.)

[129] M. D. Zeiler and R. Fergus, *Visualizing and understanding convolutional networks*, in *European conference on computer vision*, pp. 818–833, Springer, 2014. (Cited on page 101.)

[130] S. Ioffe and C. Szegedy, *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*, 1502.03167. (Cited on page 104.)

[131] I. Loshchilov and F. Hutter, *Fixing Weight Decay Regularization in Adam*, 1711.05101. (Cited on page 105.)

[132] G. Kasieczka, T. Plehn and M. Russel, "Top Quark Tagging Reference Dataset." https://doi.org/10.5281/zenodo.2603256. (Cited on page 106.)

[133] DELPHES 3 collaboration, *DELPHES 3, A modular framework for fast simulation of a generic collider experiment*, *JHEP* **02** (2014) 057 [1307.6346]. (Cited on page 107.)

[134] A. Butter et al., *The Machine Learning Landscape of Top Taggers*, *SciPost Phys.* **7** (2019) 014 [1902.09914]. (Cited on page 107.)

[135] S. Xie, R. Girshick, P. Dollár, Z. Tu and K. He, *Aggregated residual transformations for deep neural networks*, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1492–1500, 2017. (Cited on page 107.)

[136] ATLAS collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, *Phys. Lett.* **B716** (2012) 1 [1207.7214]. (Cited on page 116.)

[137] CMS collaboration, *Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC*, *Phys. Lett.* **B716** (2012) 30 [1207.7235]. (Cited on page 116.)

[138] CMS collaboration, *Measurements of properties of the Higgs boson decaying into the four-lepton final state in pp collisions at* $\sqrt{s} = 13$ *TeV*, *JHEP* **11** (2017) 047 [1706.09936]. (Cited on page 116.)

[139] ATLAS collaboration, *Measurement of the Higgs boson mass in the* $H \to ZZ^* \to 4\ell$ *and* $H \to \gamma\gamma$ *channels with* $\sqrt{s} = 13$ *TeV pp collisions using the ATLAS detector*, *Phys. Lett.* **B784** (2018) 345 [1806.00242]. (Cited on page 116.)

[140] ATLAS collaboration, *Measurement of Higgs boson production in the diphoton decay channel in pp collisions at center-of-mass energies of 7 and 8 TeV with the ATLAS detector*, *Phys. Rev.* **D90** (2014) 112015 [1408.7084]. (Cited on page 116.)

[141] CMS collaboration, *Observation of the diphoton decay of the Higgs boson and measurement of its properties*, *Eur. Phys. J.* **C74** (2014) 3076 [1407.0558]. (Cited on page 116.)

[142] ATLAS collaboration, *Measurements of Higgs boson production and couplings in the four-lepton channel in pp collisions at center-of-mass energies of 7 and 8 TeV with the ATLAS detector*, *Phys. Rev.* **D91** (2015) 012006 [1408.5191]. (Cited on page 116.)

[143] CMS collaboration, *Measurement of the properties of a Higgs boson in the four-lepton final state*, *Phys. Rev.* **D89** (2014) 092007 [1312.5353]. (Cited on page 116.)

[144] ATLAS collaboration, *Observation and measurement of Higgs boson decays to WW^* with the ATLAS detector*, *Phys. Rev.* **D92** (2015) 012006 [1412.2641]. (Cited on page 116.)

[145] ATLAS collaboration, *Study of (W/Z)H production and Higgs boson couplings using $H \to WW^*$ decays with the ATLAS detector*, *JHEP* **08** (2015) 137 [1506.06641]. (Cited on page 116.)

[146] CMS collaboration, *Measurement of Higgs boson production and properties in the WW decay channel with leptonic final states*, *JHEP* **01** (2014) 096 [1312.1129]. (Cited on page 116.)

[147] ATLAS, CMS collaboration, *Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV*, *JHEP* **08** (2016) 045 [1606.02266]. (Cited on page 116.)

[148] CMS collaboration, *Observation of $\mathrm{t\bar{t}}H$ production*, *Phys. Rev. Lett.* **120** (2018) 231801 [1804.02610]. (Cited on page 116.)

[149] ATLAS collaboration, *Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector*, *Phys. Lett.* **B784** (2018) 173 [1806.00425]. (Cited on page 117.)

[150] ATLAS collaboration, *Observation of $H \to b\bar{b}$ decays and VH production with the ATLAS detector*, *Phys. Lett.* **B786** (2018) 59 [1808.08238]. (Cited on page 117.)

[151] CMS collaboration, *Observation of Higgs boson decay to bottom quarks*, *Phys. Rev. Lett.* **121** (2018) 121801 [1808.08242]. (Cited on page 117.)

[152] ATLAS collaboration, *A search for the dimuon decay of the Standard Model Higgs boson in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS Detector*, Tech. Rep. ATLAS-CONF-2019-028, Jul, 2019. (Cited on page 117.)

[153] CMS collaboration, *Search for the Higgs boson decaying to two muons in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *Phys. Rev. Lett.* **122** (2019) 021801 [1807.06325]. (Cited on page 117.)

[154] C. Delaunay, T. Golling, G. Perez and Y. Soreq, *Enhanced Higgs boson coupling to charm pairs*, *Phys. Rev.* **D89** (2014) 033014 [1310.7029]. (Cited on page 117.)

[155] G. Perez, Y. Soreq, E. Stamou and K. Tobioka, *Constraining the charm Yukawa and Higgs-quark coupling universality*, *Phys. Rev.* **D92** (2015) 033016 [1503.00290]. (Cited on page 117.)

[156] A. Falkowski, F. Riva and A. Urbano, *Higgs at last*, *JHEP* **11** (2013) 111 [1303.1812]. (Cited on page 117.)

[157] G. T. Bodwin, F. Petriello, S. Stoynev and M. Velasco, *Higgs boson decays to quarkonia and the $H\bar{c}c$ coupling*, *Phys. Rev.* **D88** (2013) 053003 [1306.5770]. (Cited on page 117.)

[158] ATLAS collaboration, *Search for Higgs and Z Boson Decays to $J/\psi\gamma$ and $\Upsilon(nS)\gamma$ with the ATLAS Detector*, *Phys. Rev. Lett.* **114** (2015) 121801 [1501.03276]. (Cited on page 117.)

[159] CMS collaboration, *Search for a Higgs boson decaying into $\gamma^*\gamma \to \ell\ell\gamma$ with low dilepton mass in pp collisions at $\sqrt{s} = 8$ TeV*, *Phys. Lett.* **B753** (2016) 341 [1507.03031]. (Cited on page 117.)

[160] ATLAS collaboration, *Searches for exclusive Higgs and Z boson decays into $J/\psi\gamma$, $\psi(2S)\gamma$, and $\Upsilon(nS)\gamma$ at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *Phys. Lett.* **B786** (2018) 134 [1807.00802]. (Cited on page 117.)

[161] CMS collaboration, *Search for rare decays of Z and Higgs bosons to J/ψ and a photon in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *Eur. Phys. J.* **C79** (2019) 94 [1810.10056]. (Cited on page 117.)

[162] ATLAS collaboration, *Search for the Decay of the Higgs Boson to Charm Quarks with the ATLAS Experiment*, *Phys. Rev. Lett.* **120** (2018) 211802 [1802.04329]. (Cited on page 118.)

[163] CMS collaboration, *Search for the standard model Higgs boson decaying to charm quarks*, Tech. Rep. CMS-PAS-HIG-18-031, 2019. (Cited on page 118.)

[164] P. Nason, *A New method for combining NLO QCD with shower Monte Carlo algorithms*, *JHEP* **11** (2004) 040 [hep-ph/0409146]. (Cited on page 119.)

[165] S. Frixione, P. Nason and C. Oleari, *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*, *JHEP* **11** (2007) 070 [0709.2092]. (Cited on page 119.)

[166] S. Alioli, P. Nason, C. Oleari and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, *JHEP* **06** (2010) 043 [1002.2581]. (Cited on page 119.)

[167] K. Hamilton, P. Nason and G. Zanderighi, *MINLO: Multi-Scale Improved NLO*, *JHEP* **10** (2012) 155 [1206.3572]. (Cited on page 119.)

[168] G. Luisoni, P. Nason, C. Oleari and F. Tramontano, *$HW^{\pm}$/HZ + 0 and 1 jet at NLO with the POWHEG BOX interfaced to GoSam and their merging within MiNLO*, *JHEP* **10** (2013) 083 [1306.2542]. (Cited on page 119.)

[169] G. Ferrera, M. Grazzini and F. Tramontano, *Associated WH production at hadron colliders: a fully exclusive QCD calculation at NNLO*, *Phys. Rev. Lett.* **107** (2011) 152003 [1107.1164]. (Cited on page 119.)

[170] G. Ferrera, M. Grazzini and F. Tramontano, *Higher-order QCD effects for associated WH production and decay at the LHC*, *JHEP* **04** (2014) 039 [1312.1669]. (Cited on page 119.)

[171] G. Ferrera, M. Grazzini and F. Tramontano, *Associated ZH production at hadron colliders: the fully differential NNLO QCD calculation*, *Phys. Lett.* **B740** (2015) 51 [1407.4747]. (Cited on page 119.)

[172] G. Ferrera, G. Somogyi and F. Tramontano, *Associated production of a Higgs boson decaying into bottom quarks at the LHC in full NNLO QCD*, *Phys. Lett.* **B780** (2018) 346 [1705.10304]. (Cited on page 119.)

[173] O. Brein, R. V. Harlander and T. J. E. Zirke, *vh@nnlo - Higgs Strahlung at hadron colliders*, *Comput. Phys. Commun.* **184** (2013) 998 [1210.5347]. (Cited on page 119.)

[174] R. V. Harlander, S. Liebler and T. Zirke, *Higgs Strahlung at the Large Hadron Collider in the 2-Higgs-Doublet Model*, *JHEP* **02** (2014) 023 [1307.8122]. (Cited on page 119.)

[175] A. Denner, S. Dittmaier, S. Kallweit and A. Mück, *HAWK 2.0: A Monte Carlo program for Higgs production in vector-boson fusion and Higgs strahlung at hadron colliders*, *Comput. Phys. Commun.* **195** (2015) 161 [1412.5390]. (Cited on page 119.)

[176] R. Frederix and S. Frixione, *Merging meets matching in MC@NLO*, *JHEP* **12** (2012) 061 [1209.6215]. (Cited on page 120.)

[177] Y. Li and F. Petriello, *Combining QCD and electroweak corrections to dilepton production in FEWZ*, *Phys. Rev.* **D86** (2012) 094034 [1208.5967]. (Cited on page 120.)

[178] S. Kallweit, J. M. Lindert, P. Maierhofer, S. Pozzorini and M. Schönherr, *NLO QCD+EW predictions for V + jets including off-shell vector-boson decays and multijet merging*, *JHEP* **04** (2016) 021 [1511.08692]. (Cited on page 120.)

[179] S. Frixione, P. Nason and G. Ridolfi, *A Positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction*, *JHEP* **09** (2007) 126 [0707.3088]. (Cited on page 120.)

[180] M. Czakon and A. Mitov, *Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders*, *Comput. Phys. Commun.* **185** (2014) 2930 [1112.5675]. (Cited on page 120.)

[181] CMS collaboration, *Measurement of differential cross sections for top quark pair production using the lepton+jets final state in proton-proton collisions at 13 TeV*, *Phys. Rev.* **D95** (2017) 092001 [1610.04191]. (Cited on pages 120 and 154.)

[182] R. Frederix, E. Re and P. Torrielli, *Single-top t-channel hadroproduction in the four-flavour scheme with POWHEG and aMC@NLO*, *JHEP* **09** (2012) 130 [1207.5391]. (Cited on page 120.)

[183] S. Alioli, P. Nason, C. Oleari and E. Re, *NLO single-top production matched with shower in POWHEG: s- and t-channel contributions*, *JHEP* **09** (2009) 111 [0907.4076]. (Cited on page 120.)

[184] E. Re, *Single-top Wt-channel production matched with parton showers using the POWHEG method*, *Eur. Phys. J.* **C71** (2011) 1547 [1009.2450]. (Cited on page 120.)

[185] T. Melia, P. Nason, R. Rontsch and G. Zanderighi, *W+W-, WZ and ZZ production in the POWHEG BOX*, *JHEP* **11** (2011) 078 [1107.5051]. (Cited on page 120.)

[186] J. Alwall et al., *Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions*, *Eur. Phys. J.* **C53** (2008) 473 [0706.2569]. (Cited on page 120.)

[187] CMS collaboration, *Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at $\sqrt{s}$ = 8 TeV*, *JINST* **10** (2015) P06005 [1502.02701]. (Cited on page 121.)

[188] CMS collaboration, *Jet algorithms performance in 13 TeV data*, Tech. Rep. CMS-PAS-JME-16-003, 2017. (Cited on pages 123 and 124.)

[189] CMS collaboration, *Pileup Removal Algorithms*, Tech. Rep. CMS-PAS-JME-14-001, 2014. (Cited on page 123.)

[190] J. M. Butterworth, A. R. Davison, M. Rubin and G. P. Salam, *Jet substructure as a new Higgs search channel at the LHC*, *Phys. Rev. Lett.* **100** (2008) 242001 [0802.2470]. (Cited on page 125.)

[191] T. Plehn, G. P. Salam and M. Spannowsky, *Fat Jets for a Light Higgs*, *Phys. Rev. Lett.* **104** (2010) 111801 [0910.5472]. (Cited on page 125.)

[192] T. Chen and C. Guestrin, *XGBoost: A Scalable Tree Boosting System*, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pp. 785–794, 2016. (Cited on page 133.)

[193] J. Butterworth et al., *PDF4LHC recommendations for LHC Run II*, *J. Phys.* **G43** (2016) 023001 [1510.03865]. (Cited on page 155.)

[194] R. J. Barlow and C. Beeston, *Fitting using finite Monte Carlo samples*, *Comput. Phys. Commun.* **77** (1993) 219. (Cited on page 155.)

[195] CMS collaboration, *Measurement of the production cross section for single top quarks in association with W bosons in proton-proton collisions at* $\sqrt{s} = 13$ *TeV*, *JHEP* **10** (2018) 117 [1805.07399]. (Cited on page 155.)

[196] CMS collaboration, *Measurement of the single top quark and antiquark production cross sections in the t channel and their ratio in proton-proton collisions at* $\sqrt{s} = 13$ *TeV*, 1812.10514. (Cited on page 155.)

[197] CMS collaboration, *Measurement of the WW cross section pp collisions at sqrt(s)=13 TeV*, Tech. Rep. CMS-PAS-SMP-16-006, 2016. (Cited on page 155.)

[198] CMS collaboration, *Measurements of the* $\mathrm{pp} \to \mathrm{ZZ}$ *production cross section and the* $\mathrm{Z} \to 4\ell$ *branching fraction, and constraints on anomalous triple gauge couplings at* $\sqrt{s} = 13$ *TeV*, *Eur. Phys. J.* **C78** (2018) 165 [1709.08601]. (Cited on page 155.)

[199] CMS collaboration, *Measurements of the pp* $\to$ *WZ inclusive and differential production cross section and constraints on charged anomalous triple gauge couplings at* $\sqrt{s} = 13$ *TeV*, *JHEP* **04** (2019) 122 [1901.03428]. (Cited on page 155.)

[200] CMS collaboration, *CMS Luminosity Measurements for the 2016 Data Taking Period*, Tech. Rep. CMS-PAS-LUM-17-001, 2017. (Cited on page 156.)

[201] A. L. Read, *Presentation of search results: The CL(s) technique*, *J. Phys.* **G28** (2002) 2693. (Cited on page 165.)

[202] T. Junk, *Confidence level computation for combining searches with small statistics*, *Nucl. Instrum. Meth.* **A434** (1999) 435 [hep-ex/9902006]. (Cited on page 165.)

[203] G. Cowan, K. Cranmer, E. Gross and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, *Eur. Phys. J.* **C71** (2011) 1554 [1007.1727]. (Cited on page 165.)

[204] The ATLAS Collaboration, The CMS Collaboration, The LHC Higgs Combination Group collaboration, *Procedure for the LHC Higgs boson search combination in Summer 2011*, Tech. Rep. CMS-NOTE-2011-005, ATL-PHYS-PUB-2011-11, 2011. (Cited on page 165.)