PROBING NEW PHYSICS USING INITIAL STATE RADIATION JETS AT THE LARGE HADRON COLLIDER

by

Cristina Ana Mantilla Suarez

A dissertation submitted to The Johns Hopkins University

in conformity with the requirements for the degree of

Doctor of Philosophy

Baltimore, Maryland

July, 2020

Abstract

The existence of new particles and interactions could potentially address fundamental questions about our universe, for example, the nature of dark matter. If dark matter couples, even feebly, to the Standard Model, then new particles mediating this interaction could be produced in accelerator-based experiments. This dissertation describes the search for such mediators in a proton-proton collider, the LHC. The search is performed in a low-mass regime that has not been explored before, where these new mediator particles would couple weakly to the Standard Model quarks. Signal candidates will be recoiling against initial state radiation or ISR. The presence of ISR ensures that events in data will have enough energy to satisfy the trigger requirements that prevent saturation of the data bandwidth. The ISR also gives the resonance a large Lorentz boost, so that its decay products are highly collimated inside a single jet of hadrons. The distribution of the jet mass is probed for a potential narrow peaking signal over a smoothly falling background. No evidence for such dark matter mediator resonances is observed within the mass range of 50–450 GeV and the most stringent constraints to date are placed within 50–300 GeV.

Thesis Committee

Primary Readers

Petar Maksimovic (Advisor)
Professor
Department of Physics and Astronomy
Johns Hopkins Zanvyl Krieger School of Arts and Sciences

Morris Swartz
Professor
Department of Physics and Astronomy
Johns Hopkins Zanvyl Krieger School of Arts and Sciences

Nadia Zakamska
Professor
Department of Physics and Astronomy
Johns Hopkins Zanvyl Krieger School of Arts and Sciences

Philip Harris
Assistant Professor
Department of Physics
Massachusetts Institute of Technology

Nhan Tran
Wilson Fellow
Fermi National Laboratory

Acknowledgments

I thank my advisor Petar Maksimovic, as well as Phil Harris and Nhan Tran. I could have not asked for better advisors, their enthusiasm in physics has shaped my scientific career and I am sincerely and deeply grateful for their support and guidance.

I thank the group of postdocs and students that allowed for most of this work to be completed - the DAZSLE team. In particular, I thank Javier Duarte, Nick Smith, Dylan Rankin, David Yu, Jeff Krupa, Martin Kwok and Sang Eon Park. I am also thankful to other colleagues in HEP with whom I have collaborated: Morris Swartz, Oz Amram, Lucas Corcodilos, Tamas Vami, Marc Osherson, Caterina Vernieri, and in general the EXO, JMAR and Tracker DPG groups in CMS.

I thank members of my thesis and GBO committees, in particular Morris and Nadia for their helpful advice. I also want to thank my undergrad advisor Edgar Carrera for his efforts to guide Ecuadorian students in this field and in particular for getting me started in particle physics.

Finally, I thank my family and friends for their constant love and support. I thank my lifelong friends Pris and Pipo, my very loving friends with kick-ass physics aptitudes Kim and Tanvi, and my physics buddies Oz and Lucas. I

thank my cousins Paula and Roberto and the light and peace in their house: Lucia and Paz. I also want to thank Joe, for making my days so much brighter.

I dedicate this thesis to my parents whom I miss and love very much. I thank my always supporting siblings Rosita and Mundi, and keep my family close to my heart.

Table of Contents

List of Tables

List of Figures

Chapter 1

Motivation

The idea that gave rise to the study of particle physics is that all matter is composed of the most basic and *indivisible* particles. The successive discoveries of these elementary particles which began in the late 1890s prompted new questions about their properties and interactions. We now know that particles can be arranged in a manner that resembles the arrangement of chemical elements in the Periodic Table. Instead of listing the element symbol, atomic number and recurring chemical properties, one can list the spin, mass and fundamental quantum numbers of each particle. These properties are summarized in a theory framework called the Standard Model (SM) of Particle Physics. The Standard Model is capable of explaining how these particles interact and thereby of predicting particle phenomena that can be observed in experiments. However, as precise as it seems to be, the SM cannot explain all of the phenomena observed in our universe and still leaves questions unanswered.

One of the greatest outstanding questions is the nature of dark matter (DM), which makes up most of the amount of matter in our universe, is

approximately inert and, contrary to ordinary matter, non-luminous. Astronomical and cosmological measurements provide incontrovertible evidence of its existence, but little is known about its composition and physical properties. Modern particle physics hypothesizes that dark matter has a particle nature, like ordinary matter. This has motivated extension theories of the SM that provide possible particle DM candidates. Many of these theories are well-motivated because the existence of these candidates would solve other unanswered questions in particle physics. However, the range of masses of the particle DM candidate spans over almost 45 orders of magnitude. No single search technique can cover such a parameter range, so one should first select a plausible range of masses that facilitates detection.

The production mechanism of DM provides a good starting point for categorizing DM models. A popular hypothesis is that DM was produced through interactions with the bath of ordinary matter that filled the early Universe. In this scenario, DM particles and SM particles were in thermal equilibrium and the strength of their interactions determined the current abundance of dark matter, known as the thermal relic abundance. Using the thermal relic abundance, one can work backwards and discover a constrained mass range for possible DM candidates. The weakly interacting massive particle, or WIMP, is among the most promising candidates. A WIMP is an electrically neutral and stable particle, with a mass in the range from a few GeV to the electroweak scale, and with a thermal relic abundance set by weak-scale annihilations. The mass and coupling[1] ranges of WIMPs could allow their production at a TeV-scale collider, and are thus most relevant for the

[1] A coupling parameter denotes the strength of the interaction.

interpretation of collider results and the results of this thesis.

Particle colliders could produce DM particles in their collisions because they can reach a very high energy and high intensity scale. A commonly tested scenario in hadron colliders is the production of a DM mediator that has a measurable coupling to SM quarks [1, 2]. Its decay into a quark-antiquark pair leaves a signature that consists of two collimated sprays of particles, called *particle jets*. The mass of these jets can be measured and, for collision events where the DM mediator is produced, its value will be around the mass of the mediator. However, the occurrence of this signal would be rare. Most of the collision events that exhibit the same particle signature will be produced by spurious jets that did not originate from the decay of a massive particle. These jets constitute the background jets and the distribution of their mass values is continuous. The main task of this search is to identify the DM mediator signal as a resonance *peak* in the mass spectrum of the dijet pair.

Current and past searches for such resonance peaks in the dijet mass distribution have obtained null-results in most of the available mass range, recently reaching the TeV scale. Thus, current experiments look to extending the search to lower couplings. In the low-coupling parameter space, the DM mediator will couple very weakly with the SM quarks increasing the likelihood for the signal to be hidden under large background processes. One can take advantage of the unprecedented high-energy proton-proton collisions of the Large Hadron Collider (LHC) in Switzerland to search in this region.

Searches at the LHC have excluded resonances that couple to quarks with masses between 1.0 and 7.6 TeV [3, 4], in a region of $0.1 < g'_{\mathrm{q}} < 1.0$, where g'_{q}

represents the coupling of the resonance to quarks. However, below masses of 1 TeV their sensitivity is limited by the large background rate. The main experimental difficulties originate from the large increase in the cross section of multijet backgrounds at small resonance masses, and the more restrictive requirements in the hardware selection algorithm (trigger). The latter are needed to reduce the data recording rate because of limited resources for event processing and storage. To overcome this limitation, trigger-level analyses that record only partial event information and can set lower trigger thresholds, have placed limits on resonances with masses of 300–1000 GeV [4–8].

In this dissertation, we study a complimentary technique to access the low-mass regime. It concentrates on the mass region of 50–450 GeV, exploring for the first time masses below 140 GeV. We look for dijet resonances that would be produced with significant initial-state radiation (ISR) from quark/gluon radiation. The presence of ISR ensures that the events have enough energy to satisfy the trigger requirement, either by the ISR or by the resonance itself. The recoiling ISR also gives the resonance a high Lorentz boost, so that its decay results in a highly collimated particle jet signature. This strategy allows us to search for resonances in a coupling and mass regime to which previous searches were insensitive.

The ISR triggering technique is not only capable of probing low-mass DM mediators but also highly-energetic hadronic decays of known SM resonances in the 50–150 GeV range, such as the Higgs boson H. The SM Higgs boson, predicted independently in Refs. [9–11], solves the long-standing puzzle of the mechanism behind the mass generation of the massive mediators of the

weak force, the W and Z bosons, and the fermions. Following the discovery of a new Higgs-like particle in 2012, it has become a matter of precision to check that the properties of this particle are consistent with the SM predictions. A promising way of hunting for deviations from the SM, which may point to new physics, is by measuring the kinematic properties of the Higgs-like particle. For example, new heavy particles can modify the Higgs boson coupling to the top quark and consequently its kinematic spectrum at high energy. The second part of this thesis concerns the identification of low-mass hadronic decays of the Higgs boson when this is produced with high Lorentz boost.

To understand the challenges of accessing the low-mass regime, we must first develop the intuition that leads to the modern description of jet reconstruction and identification at a hadron collider. The remaining chapters in this dissertation are as follows:

- In **Chapter 2**, we provide an overview of the theoretical background behind the search for resonances that couple to quarks. This includes a description of the structure of the SM, the phenomenology of proton collisions as described by the SM and the limitations of the SM, in particular those related to our understanding of dark matter. We conclude with a discussion on how to test simplified models of dark matter at the LHC.

- In **Chapter 3**, we summarize the particle detection setup, the LHC design and its performance during the last proton run. We describe the Compact Muon Solenoid (CMS) detector and its sub-systems that identify outgoing particles from the collisions. Particle identification in any

of these sub-systems is based on the basic principle that when particles go through matter, they interact with it and leave a signal. The signal can be used to infer properties of the outgoing particles: its position, energy deposits and momentum. Finally, we describe the methods used to provide a description of the particles reconstructed in an event.

- In **Chapter 4** we focus on the reconstruction of hadronic signatures in the detector. We describe the algorithms that cluster hadronic energy deposits into particle-jets and how these are calibrated in data. The second part of this chapter focuses on the identification of the hadronic decays of highly-energetic resonances, that are reconstructed and merged into single large-radius jets. We detail the methods and identification observables that exploit the radiation pattern inside the jets. We derive a method that prevents any of the large-radius jet identification observables from distorting the jet mass distribution. Finally, we describe the calibration of these techniques in data.

- In **Chapter 5**, we present the search strategy for low-mass resonances that couple to quarks. We present an overview of the online and offline event selection, that aims to identify a highly energetic jet with two-prong substructure. We then describe how we estimate the contribution of background processes that mimic the event signature of our signal in the detector. The main background process in this search is the production of spurious jets in the detector and its contribution is estimated with a novel data-driven method. Finally, we present the results of this search and their interpretation within the framework of a DM mediator.

This search was performed using data collected by the CMS detector at the LHC in 2016-2017. It is described in an abbreviated form by the CMS Collaboration as two separate papers [12, 13].

- In **Chapter 6** we apply a similar analysis strategy to measure highly energetic decays of the 125.1 GeV Higgs boson. The H candidates are required to have a high Lorentz boost to meet the restrictive trigger criteria and are reconstructed as single large-radius jets. We explore methods of identification of Higgs-like jets that exploit the radiation pattern and color properties of the jets. This chapter is based on the work appearing in Ref. [14]

- Finally, in **Chapter 7**, we summarize the results presented here and give an outlook on how using ISR in the event signature can help uncover other hidden signatures in LHC data.

Chapter 2

Theoretical Background

2.1 The Standard Model

The SM provides a formal description of the particles and the dynamics of their interactions via a quantum field theory [15–19].

In this theory, particles are quantum matter fields: wave functions, $\psi(x)$, that take some value for every point x in the space-time. They are characterized by quantum numbers, such as their mass, m, and their spin, s. Fermions, with half-integer spin, follow Bose-Einstein statistics, while bosons, with integer spin, follow Fermi-Dirac statistics [20].

The structure of the SM is encoded in a Lagrangian density, $\mathcal{L}_{\text{SM}}$. The Lagrangian density depends on one or more fields and contains terms that describe free fields and others that describe interactions. The interaction terms in the SM Lagrangian describe the fundamental interactions as mediated by fields.

Fields in the SM respect Lorentz symmetries and thus must be invariant under translations, rotations and boosts on the space-time [21]. Furthermore,

they must be gauge invariant. Gauge invariance allows for the additional degrees of freedom, that correspond to the mediator gauge bosons, to be introduced in the SM Lagrangian. It implies that the observables derived from the Lagrangian will remain invariant under local transformations, $\psi(x) \rightarrow U(x)\psi(x)$, where the transformations $U(x)$ form Lie groups. Examples of the groups considered in the SM are $U(N)$, that describes N-dimensional rotations, and $SU(N)$, that imposes the additional condition of tracelessness.

In this section, we describe each of the gauge fields included in the SM Lagrangian. To show how interactions arise from the gauge principle, we first discuss the kinematics of the free fields and introduce the electromagnetic interaction as the simplest case.

2.1.1 The electromagnetic interaction

For a free field ψ of spin $1/2$ and mass m, the Dirac equation of motion can be written as:

$$(i\gamma^{\mu}\partial_{\mu} - m)\psi(x) = 0, \quad [1] \tag{2.1}$$

[1] In $h = c = 1$ units

where $\psi(x)$ is the four-component spinor [2] representing the field wave-function, γ^μ are the Dirac matrices corresponding to the space-time coordinates x^μ[3], and $\partial_\mu \equiv \frac{\partial}{\partial x^\mu}$. The solutions to this equation are plane waves propagating in the positive and negative time directions. The latter can be also interpreted as anti-particles propagating forwards in time.

The corresponding Lagrangian density for this free field is given by:

$$\mathcal{L}_{\text{free}} = \bar{\psi}(x)(i\gamma^\mu \partial_\mu - m)\psi(x). \tag{2.2}$$

If we introduce an *electromagnetic field*, $A_\mu(x)$, the equations of motion are described by Maxwell's equations. Under no external currents, these are:

$$\partial^\mu F_{\mu\nu} = 0, \tag{2.3}$$

where, $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is the antisymmetric field tensor for $A_\mu(x)$.

The Lagrangian corresponding to Eq. 2.3, $\mathcal{L}_{\text{EM}} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu}$, obeys local $U(1)$ symmetry. This means that both $\psi(x)$ and A_μ are invariant under the

[2] A spinor field describes a particle of spin s, where s is an integer or half-integer. This field has four components and is defined via a set of γ matrices. We can use the following representation for the γ matrices:

$$\gamma^0 = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \text{and,} \gamma^i = \begin{pmatrix} 0 & \sigma^i \\ -\sigma^i & 0 \end{pmatrix},$$

where $\mathbf{I}$ denotes a 2×2 identity matrix, $\mathbf{0}$ denotes a 2×2 null matrix, and the σ^i are the Pauli spin matrices.

[3] x^μ is a Lorentzian "four-vector": $x_0 = ct, x_1 = x, x_2 = y, x_3 = z$

unitary transformations:

$$\psi(x) \to e^{iQ_\psi \theta(x)} \psi(x), \tag{2.4}$$

$$A_\mu(x) \to A_\mu(x) + \frac{1}{q} \partial_\mu \theta(x). \tag{2.5}$$

In order for the Lagrangian to be invariant under this symmetry, the derivative operator, ∂_μ, needs to be replaced with the covariant derivative: $\partial_\mu \to D_\mu = \partial_\mu - iq_e A_\mu$. This replacement introduces the following interaction term in the Lagrangian:

$$\mathcal{L}_{\text{EM}} = q\overline{\psi}(x) A_\mu \gamma^\mu \psi(x). \tag{2.6}$$

Here, the field A_μ corresponds to a massless spin-1 gauge boson that is coupled to the fermion field via a coupling constant q. In quantum electrodynamics (QED), the gauge boson A_μ is the massless photon, γ, and q is the electric charge of the fermion. If a mass term, $\frac{1}{2}m^2 A_\mu A^\mu$, is added, this would spoil the gauge invariance. This situation will be discussed later in Sec. 2.1.4.

Within the QED model, all allowed processes occur due to photons and fermions propagating between space-time coordinates and photon-fermion interactions. Experimentally, these interactions (and any other interactions in quantum field theory) can be studied by measuring observable decay rates or scattering cross sections. For this, we need to consider the probability of interaction or transition amplitude $\langle \psi_f | H_{\text{int}} | \psi_i \rangle$, between the initial state $|\psi_i\rangle$ and the final state $|\psi_f\rangle$. Here, H_{int} is the interaction Hamiltonian that can be derived for any Lagrangian density $\mathcal{L}$, for example $\mathcal{L}_{\text{EM}}$. Measurable quantities such as scattering cross sections are proportional to the magnitude

squared of the transition amplitude, also called matrix element.

To evaluate the matrix element, we take the time-ordered series expansion in terms of the interaction strength or coupling g. If $g \ll 1$, the series converges quickly and it is sufficient to consider only the first few terms. Most of the calculations used in colliders are known to next-to-leading order (NLO), i.e. truncated at order g^2, or even next-to-next-to-leading order (NNLO), truncated at order g^3. Each term in the perturbation expansion is represented as a pictogram which is called a Feynman diagram. The basic vertex used in QED to construct fermion-photon interactions is shown in Fig. 2.1. The vertices and propagators are specified by the form of the interaction Lagrangian. In the following, we will derive similar basic vertices for the remaining fundamental interactions in the SM.

2.1.2 The strong interaction

The strong interaction acts on quarks, spin 1/2 particles that carry color charge. It is described by the theory of quantum chromodynamics (QCD). Gluons are the mediator particles in QCD, they are massless spin 1 particles that also carry color charge and can thus couple to each other.

QCD arises from the invariance of the Lagrangian under the $SU(3)_C$ gauge group, which leads to color charge conservation in the strong interactions. A quark spinor field has three components: $\Psi(x) = (\psi_r, \psi_g, \psi_b)$, where r, g, b are colors. It transforms under this group as:

$$\psi(x) \rightarrow \exp(ig_s\theta^a T^a)\psi(x). \tag{2.7}$$

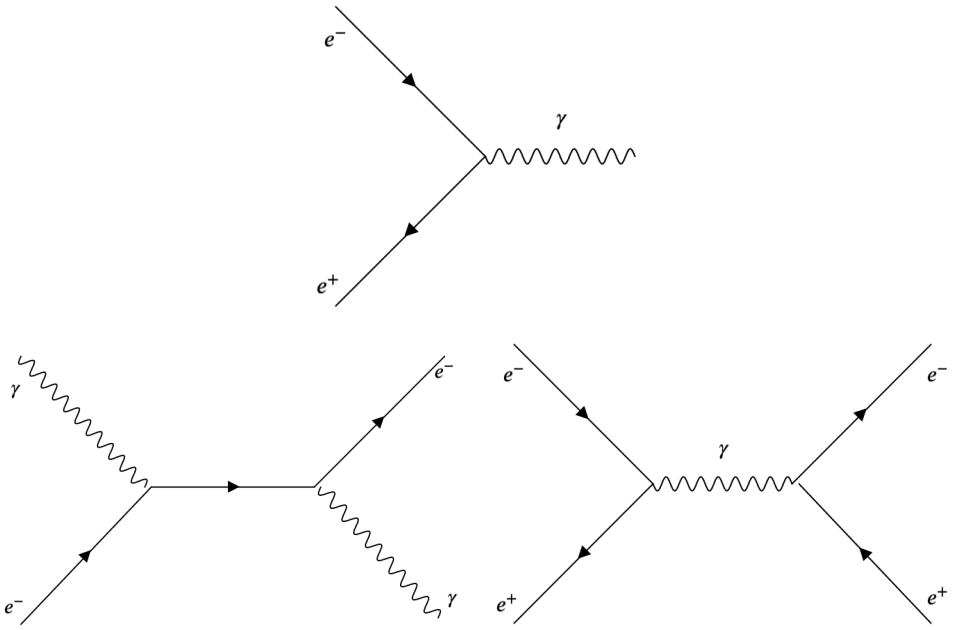

Figure 2.1: Basic QED diagrams, the top figure shows the basic vertex for QED while the bottom diagrams show process such as a fermion propagating between vertices and a fermion-photon interaction.

Here, T^a are the eight generators of the group represented by the matrices λ^a: $T^a = \frac{1}{2}\lambda^a$, θ^a are the associated rotation angles and g_s denotes the strong coupling strength.

To keep gauge invariance, the covariant derivative can be written as: $\partial_\mu \rightarrow D_\mu = \partial_\mu - ig_s G^a_\mu T^a$. G^a_μ are the eight gluon fields that must transform as: $G^a_\mu \rightarrow G^a_\mu - \partial_\mu \theta^a - g_s f_{bca} \theta^b G^b_\mu$. The last term arises from the non-abelian nature of QCD, and implies that the T^a generators do not commute but are instead related through the following relationship: $[T^a, T^b] = i f_{abc} T^c$. This expression can be simplified by writing the gluon octet as $G_\mu = G^a_\mu T^a$.

The Lagrangian density describing a quark field of mass m can be written as:

$$\mathcal{L}_{\text{QCD}} = -\frac{1}{4} G^{a,\mu\nu} G^a_{\mu\nu} + \bar{\psi}(i\gamma^\mu D_\mu + m)\psi, \tag{2.8}$$

where $G^a_{muv} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu + i g_s f_{abc} G^b_\mu G^c_\nu$ is the gluon field strength tensor. The presence of the structure constants, f_{abc}, implies that gluons carry color charge, as anticipated earlier, and thus interact with each other. The basic vertices of QCD are shown in Fig. 2.2 and include three- and four-point self-coupling gluon vertices.

QCD and renormalization In quantum field theory, calculations that involve high energy corrections to vertices and propagators may contain terms with ultraviolet divergences. These are divergences that arise from Feynman diagrams that are associated with unconstrained energy, or, equivalently with phenomena at very small distances, e.g. when considering the corrections to the quark self-energy. These can be addressed through a renormalization

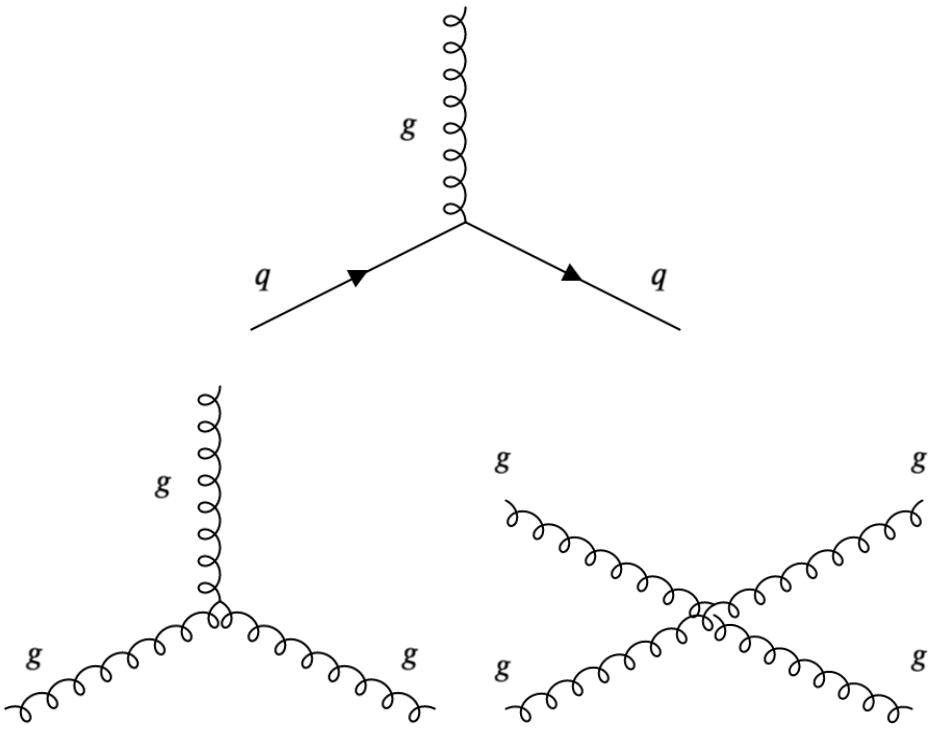

Figure 2.2: Basic QCD vertices for a quark-gluon interaction and gluon self-interactions.

scheme, by adding counter terms that absorb these infinities. Lagrangians in the SM are renormalizable since they only need a finite number of counter terms to keep all observables finite. Here, we will only focus on the renormalization of the strong coupling constant. This has important consequences on the behavior of quarks and gluons and its experimental detection.

The strong coupling constant, $\alpha_s = g_s^2/4\pi$, acquires a non-trivial energy scale dependence because of quantum corrections, such as corrections to the quark and gluon self-energies. After renormalization, the scale dependence of the coupling is described by:

$$\alpha_s = \frac{12\pi}{(32 - n_f)\ln(\mu_R^2/\lambda_{\mathrm{QCD}}^2)}, \tag{2.9}$$

where μ_R^2 is the energy renormalization scale equal to the momentum transfer Q^2, n_f is the number of fermions and the QCD scale parameter λ_{QCD} is an experimentally determined cutoff. When setting $n_f = 5$ and using the energy scale μ_R near the Z boson mass (90 GeV), λ_{QCD} takes a value of approximately 0.2 GeV. This conveys the general rule that perturbative QCD calculations only provide the evolution of an observable at an arbitrary scale, here μ_R.

From a quick look at the evolution of α_s one can already observe two important features of QCD. The first is that the coupling decreases with higher energy. As $\mu_R \to \infty$ and at very small distances, the coupling tends towards zero and quarks can be treated as free particles. This is the phenomenon of asymptotic freedom. The second is related to the coupling increasing at higher distance or lower energy. At very low energies, below λ_{QCD}, the coupling is too strong to allow perturbative calculations. In the regime, where $\alpha_S > 1$, the

theory becomes non-perturbative, necessitating different types of techniques to understand QCD in this regime. Fig. 2.3 shows the scaling behavior of α_s across a large range of momenta.

Colorless hadrons Since free quarks have not been observed in experiments and QCD has a long-range behavior [4] , the color charge is hypothesized to be confined. That is, quarks cannot be isolated and are only observed in bound colorless states known as hadrons. Hadrons are colorless and can be composed of a quark-antiquark pair (mesons) or three quarks (baryons). The theory description and experimental behavior of hadrons will be explained in Sec. 2.2.3 and the following two chapters.

2.1.3 The weak interactions

The third and final force described in the SM is the weak force. The weak interaction couples leptons with charged neutrinos, and allows a quark to change its flavor[5] (e.g. couples up and down quarks). Three massive gauge bosons mediate these interactions, the charged W^+ and W^- bosons and the neutral Z boson.

Since it was experimentally observed that weak interactions do not conserve parity (P) [23], the symmetry that flips the sign of one spatial coordinate,

[4] i.e. if one takes a heavy probe quark and an antiquark separated by a large distance, the force between them does not fall off with distance.

[5] There are six flavors of quarks: up (u), down(d), charm(c), strange(s), top(t), bottom(b).

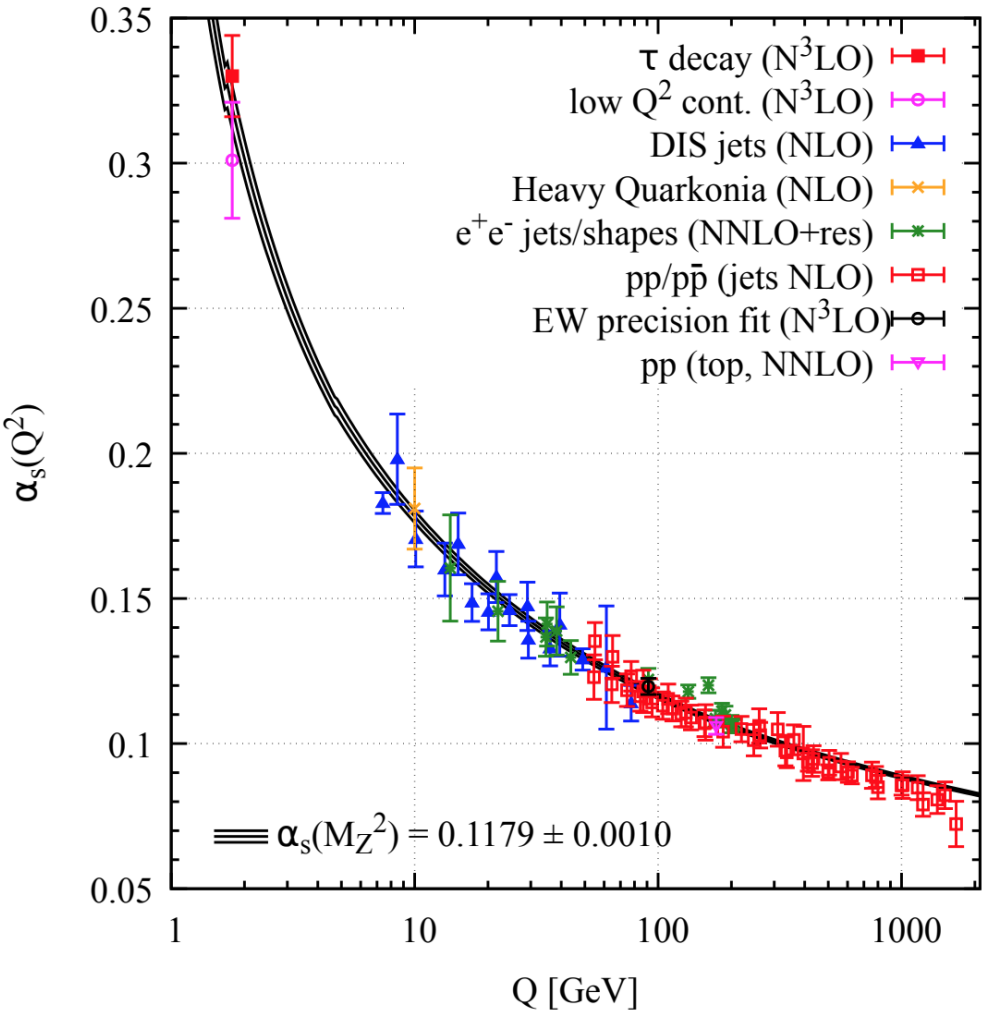

Figure 2.3: Summary of measurements of the strong coupling constant, α_s, as a function of the energy scale Q. The scaling predicted from perturbative QCD theory is compared to experimental measurements in different processes and energy regimes. The value of α_s decreases as the energy scale increases, reflecting the asymptotic freedom behavior of QCD. Figure extracted from [22].

we describe them by a chiral gauge theory. [6] In this theory, the weak interaction vertex includes not only a vector coupling (P=-1) but also an axial vector coupling (P=1). To distinguish the left- and right-chiral components of the Dirac spinors, we can introduce the γ^5 operator, $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$, with parity eigenvalues ± 1. The operators $\frac{1}{2}(1 \pm \gamma^5)$ project the left-handed, ψ_L, and right-handed ψ_R components of a spinor.

The weak interaction is symmetric under the $SU(2)_L$ non-abelian group. The $SU(2)_L$ transformations are described by three angles $\theta_j(x)$ so that left-handed fields transform as: $\psi^L \rightarrow \exp(\frac{i}{2}\theta_j\tau_j)\psi^L$. Right-handed fields ψ_R remain invariant. The $\tau_j = \frac{\sigma_j}{2}$ matrices are related to the three Pauli matrices and called the generators of the $SU(2)_L$ group.

We can group left-handed states into doublets: $\begin{pmatrix}\nu_L \\ \ell_L\end{pmatrix}$ for leptons and $\begin{pmatrix}q_L \\ q'_L\end{pmatrix}$ for quarks.

States that transform under $SU(2)_L$ are charged under a new quantum number: the weak isospin $I = \frac{1}{2}$. Its projection is $I_3 = +1/2$ for ν_L, q_L and $I_3 = -1/2$ for ℓ_L, q'_L. The right-handed fermions are uncharged under $SU(2)_L$, they have $I = I_3 = 0$ and form singlets. Thus, $\psi_l^R = e^R, \mu^R, \tau^R$ for leptons, and $\psi_q^R = u^R, d^R, c^R, s^R, t^R, b^R$ for quarks.

Imposing the local gauge invariance via the covariant derivative:

$$\partial_\mu \rightarrow \partial_\mu - i\frac{g}{2}\tau_j W^{j,\mu}, \tag{2.10}$$

[6] Here, the term chiral refers to how a particle's quantum mechanical wave function behaves with respect to its mirror image. The spin of a particle may be used to define a handedness, or helicity, for that particle, which, in the case of a massless particle, is the same as chirality. The helicity of a particle is positive ("right-handed") if the direction of its spin is the same as the direction of its motion. It is negative ("left-handed") if the directions of spin and motion are opposite.

requires a triplet of gauge boson fields: $W^{j,\mu}$. These can be associated to the two charged vector boson fields that mediate the weak force:

$$W^{\pm}\mu = \frac{1}{\sqrt{2}}(W^{1,\mu} \mp iW^{2,\mu}). \tag{2.11}$$

The third boson predicted by $SU(2)_L$ gauge invariance, $W^{3,\mu}$, is a neutral gauge boson. Since the electromagnetic force is also mediated by a neutral gauge boson, the photon, the presence of $W^{3,\mu}$ suggests the *unification of the electromagnetic and weak forces*.

The electroweak theory contains two interactions associated with the $SU(2)_L$ gauge field $W^{3,\mu}$ and a new $U(1)_Y$ gauge field B. Here, we have introduced a new quantum number, the weak hypercharge Y. Y is the same for both components of the $SU(2)_L$ doublets, so that the left handed doublets are invariant by both $SU(2)_L$ and $U(1)_Y$. The observed electric charges of the fermions are then related to the weak hypercharge Y and weak isospin by $Q = I_3 + Y/2$.

Invariance under $SU(2)_L \times U(1)_Y$ leads to the covariant derivative:

$$D^L_\mu = \partial_\mu - igW^{i,\mu}\tau^i - \frac{i}{2}g'YB_\mu, \tag{2.12}$$

for the left-handed terms and:

$$D^R_\mu = \partial_\mu - \frac{i}{2}g'YB_\mu, \tag{2.13}$$

for the right-handed terms. The coupling strengths of $SU(2)_L$ and $U(1)_Y$ are given by g and g', respectively.

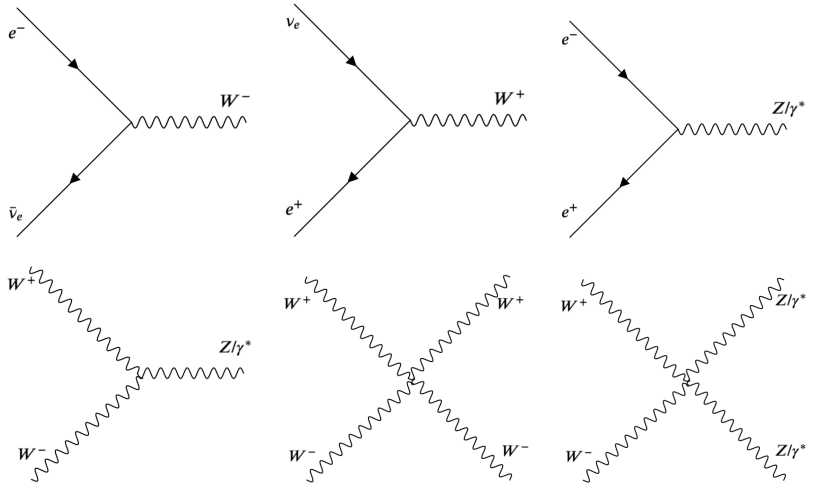

Figure 2.4: Basic vertices for massless terms in electroweak interactions. The interactions between the vector gauge bosons: photons, W and Z bosons are shown in the bottom panel.

One can change basis using a linear transformation given by:

$$\begin{pmatrix} W^{3,\mu} \\ B^\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} Z^\mu \\ A^\mu \end{pmatrix}, \tag{2.14}$$

such that Z^μ only couples to isospin and A^μ only couples to charge. Since $Q = g\sin\theta_W = g'\cos\theta_W$, the Weinberg angle $\theta_W = \tan^{-1}(g'/g)$ determines the relative coupling strengths of the electromagnetic and weak interactions.

Ignoring mass terms and the Higgs sector, the electroweak (EW) Lagrangian for electron and electron neutrino interactions is:

$$\mathcal{L}_{\text{EW}} = -\frac{1}{4}W^a_{\mu\nu}W^{a,\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} + i\bar{\psi}^L\not{\partial}\psi^L + i\bar{\nu}_e^R\not{\partial}\nu_e^R + i\bar{e}^R\not{\partial}e^R. \tag{2.15}$$

The field strength tensors are $W^a_{\mu\nu}$ and $B_{\mu\nu}$. The basic vertices for the massless terms in electroweak interactions are shown in Fig. 2.4.

2.1.4 The electroweak symmetry-breaking

So far we have not considered mass terms for the W and Z bosons in the electroweak Lagrangian. This is because these terms break the gauge symmetry.

For a boson V^μ, that transforms under U(1) such that $V^\mu \rightarrow V^\mu + \frac{1}{g}\partial^\mu\theta$, the mass term is described by: $\mathcal{L}_{\text{Vmass}} = -\frac{1}{2}m_V^2 V_\mu V^\mu$. This Lagrangian transforms under U(1) as:

$$\mathcal{L}_{\text{Vmass}} \rightarrow -\frac{1}{2}m_V^2 V_\mu V^\mu - \frac{1}{2g^2}m_V^2(\partial_\mu\theta)(\partial^\mu\theta) - \frac{1}{g^2}m_V^2 V_\mu(\partial^\mu\theta) \tag{2.16}$$

and *breaks the gauge symmetry*. In a similar way, the masses of W and Z bosons would break $\text{SU(2)}_L \times \text{U(1)}_Y$ symmetry.

Similarly, we have not yet introduced quadratic mass terms for fermions. The fermion mass terms of the form:

$$-m(\bar{\psi}^R\psi^L + \bar{\psi}^L\psi^R), \tag{2.17}$$

cannot be gauge invariant since only left-handed fields transform under SU(2)_L.

These problems can be treated by introducing two complex scalar fields arranged in a weak isospin complex doublet:

$$\phi = \frac{1}{\sqrt{2}}\begin{pmatrix}\phi_1^\dagger + i\phi_2^\dagger \\ \phi_3 + i\phi_4^\dagger\end{pmatrix}. \tag{2.18}$$

The Lagrangian density for this scalar field is:

$$\mathcal{L}_\phi = (\partial_\mu\phi)^\dagger(\partial^\mu\phi) - V(\phi). \tag{2.19}$$

The potential, $V(\phi)$, takes the form: $V(\phi) = \mu^2(\phi^\dagger\phi) + \lambda|\phi^\dagger\phi|^2$. This is the most general form of the potential that is invariant under $SU(2)_L \times U(1)_Y$ symmetry.

This potential has a degenerate global minimum at $\phi^\dagger\phi = \frac{v^2}{2} = -\mu^2/2\lambda$ with $\mu^2 < 0$. A continuum of ground states exist and we must pick an arbitrary vacuum state to fall into. If we choose the vacuum expectation value $v = \sqrt{-\mu^2/2\lambda}$ and, through gauge rotations, we fix $\langle\phi_{1,2,4}\rangle = 0$, we can rewrite the doublet as:

$$\phi = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v + H \end{pmatrix}. \tag{2.20}$$

Here we expanded $\phi_3 = v + H$, with H being the real Higgs field with mass $M_H = \mu\sqrt{2}$. This choice is called the unitary gauge and spontaneously breaks the symmetry.

The presence of a non-zero vacuum field leads to the existence of a mass term for W and Z bosons. By expanding $(\partial_\mu\phi)^\dagger(\partial^\mu\phi)$ in Eq 2.19 we introduce the following terms in the Lagrangian:

$$\frac{v^2}{8}\left(g^2(W^1_\mu W^{1,\mu} + W^2_\mu W^{2,\mu}) + (g'B_\mu - gW^3_\mu)^2\right). \tag{2.21}$$

With this mechanism, the charged vector W bosons obtain their masses from $m_W = \frac{vg}{2}$, the neutral Z boson obtains $m_Z = \frac{m_W}{\cos\theta_W}$ and the photon is massless $m_A = 0$.

The mass of the fermions is also introduced through spontaneous symmetry breaking [24, 25]. For each fermion field, the Yukawa interaction is then given by:

$$\mathcal{L}_{\text{Yukawa}} = -g_f v\bar{\psi}\psi - g_f\bar{\psi}H\psi. \tag{2.22}$$

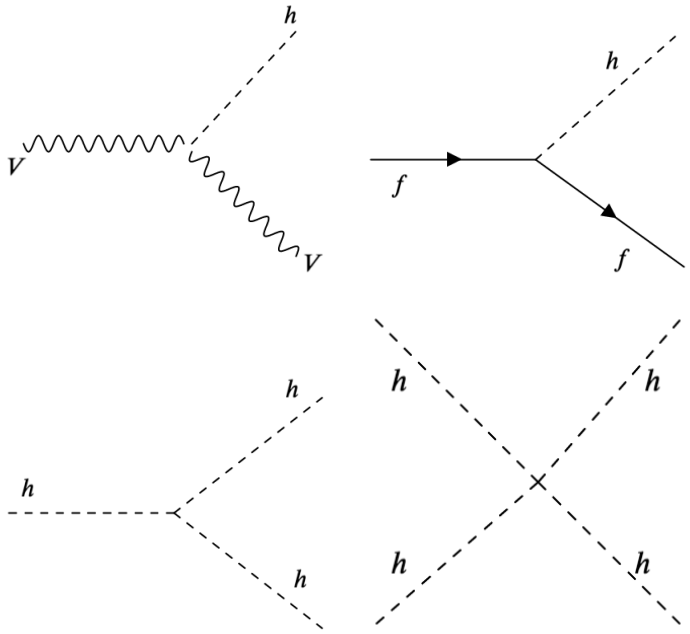

Figure 2.5: Basic vertices for the couplings of the Higgs boson field with fermions (f), vector bosons ($W^{\pm}$ or Z), or self-interactions.

Here, the first term is identified with the mass of the fermion $g_f v$ and the second term corresponds to the interaction of the fermion with the Higgs field. The coupling strength of the fermions to the Higgs doublet g_f is proportional to the mass of the particle and can be different for each fermion. The basic vertex interactions of the Higgs boson with fermions and massive gauge bosons are shown in Fig. 2.5. The mechanism of Electroweak Spontaneous Symmetry Breaking is further detailed in Refs. [9–11, 26, 27].

Summary The SM describes the dynamics and interactions of fundamental particles under the $SU(3)_C \times SU(2)_L \times U(1)_Y$ symmetry group. Electroweak interactions are invariant under $SU(2)_L \times U(1)_Y$ and mediated by $W/Z/\gamma$ bosons, while the strong interactions are invariant under $SU(3)_C$ and mediate

Table 2.1: Summary of the SM fermions and the action of the SM gauge symmetry group. The left- and right-handed chirality fields are represented by subscripts L and R. The electric charge is given by $Q = I_3 + Y/2$.

	$U(1)_Y$	$SU(2)_L$ rep.	$SU(3)_C$ rep.
Left-handed quark q_L	$1/6$	**2**	**3**
Right-handed up quark u_R	$2/3$	**1**	**3**
Right-handed down quark d_R	$-1/3$	**1**	**3**
Left-handed lepton ℓ_L	$-1/2$	**2**	**1**
Right-handed lepton (charged) ℓ_R	1	**1**	**1**
Right-handed lepton (neutrino) ν_R	0	**1**	**1**

processes that involve color charged particles. Table 2.1 summarizes the interactions of the particles, by showing the representations in which the matter fields (quarks and leptons) transform under the SM symmetry group. The particle content of the SM is summarized in the "Particle Table", which was referred to in Chapter 1 and shown in Fig 2.6. The table is color-coded by the three known generations of quarks and leptons, and shows the electric charge, relevant couplings and masses of each particle. It also shows each of the gauge mediators and the Higgs boson, which is associated with the mechanism that gives mass to the elementary particles.

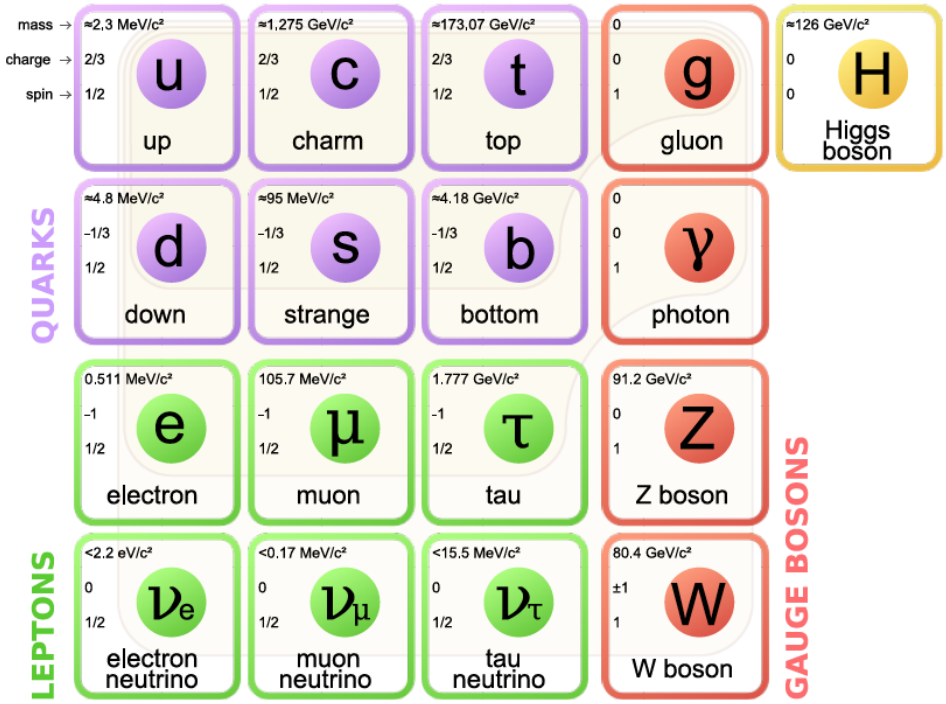

Figure 2.6: Particle content of the Standard Model. The table shows the SM gauge bosons and the fermions. For each particle species values of its mass, electrical charge and spin are also shown. Figure adapted from [28].

2.2 The Standard Model in Hadron Colliders

Although the structure of the SM can be concisely summarized in a Lagrangian, predicting the processes that arise from this structure is quite complex. The challenging nature of perturbative calculations, the high energy scale, and the complexity of composite hadron collisions make analytic predictions infeasible. Thus, to predict the behavior of scattering processes at a hadron collider, such as the LHC, we typically use Monte Carlo (MC) simulation. For a given scattering process, $pp \to XX$, we generate a large number N of events. Each event is normalized to the calculated cross section of the process at a given order. The simulated events are used in a variety of contexts, and their properties are often compared to collision data in frequency distributions (histograms). Given the finite nature of N, the MC distributions will have statistical fluctuations that are minimized by making N as large as needed.

In the following, we describe the elements needed to characterize a $pp \to X$ scattering process, and thus a Monte Carlo simulation at the LHC.

2.2.1 Parton distribution functions

The nature of the colliding protons complicates the calculation of the cross section of a scattering process. Protons are bound states of two up quarks and one down quark (uud). They are bound together by gluons, which produce short lived $q\bar{q}$ pairs of all flavors. Any of these particle constituents, referred to as partons, may be initial states for the hard scatter interaction in a pp collision. The relative fractions of a parton in the proton may vary, depending on the energy scale Q^2 with which the proton is probed. To parametrize the

probability of probing initial state partons with particular momenta when two protons collide, we use Parton Distribution Functions (PDFs), $f_i(x, Q^2)$. These specify the probability of finding a parton of species $i = g, u, d, ...$ with a fraction x of the proton's momentum.

The description of protons in terms of PDFs follows from the factorization theorem in QCD [29]. This theorem takes into account the breakdown of perturbative QCD, which we expect at low energy scales. It allows us to decouple the cross section calculation into a hard-scattering component, that describes the interaction of $ab \to X$ with perturbation theory, and a non-perturbative component, described by the PDFs of the partons a, b. This factorization depends on an arbitrary energy scale that defines the lower bound for interactions to be considered part of the hard-scattering, the factorization scale μ_F^2. This allows us to describe the cross section for a pp collision process by:

$$\sigma(pp \to XX) = \int_0^1 dx_a dx_b \sum_{a,b} f_a(x_a, \mu_F^2) f_b(x_b, \mu_F^2) \sigma(ab \to X)(Q^2, \mu_F^2). \quad (2.23)$$

In this expression, the sum is performed over all the incoming partons a, b and integrated over the allowed momenta.

One can quantitatively predict the dependence of the PDFs on the energy scale using the Dokshitser-Gribov-Lipatov-Altarelli-Parisi (DGLAP) evolution equations [30–32], which evolve the PDFs from one scale to another. But the variation of the PDF as a function of x, for a given scale, cannot be done analytically. Instead, it is constrained empirically from many experimental data. In this thesis we use the NNPDF3.0 PDF set [33], shown for example in Fig. 2.7.

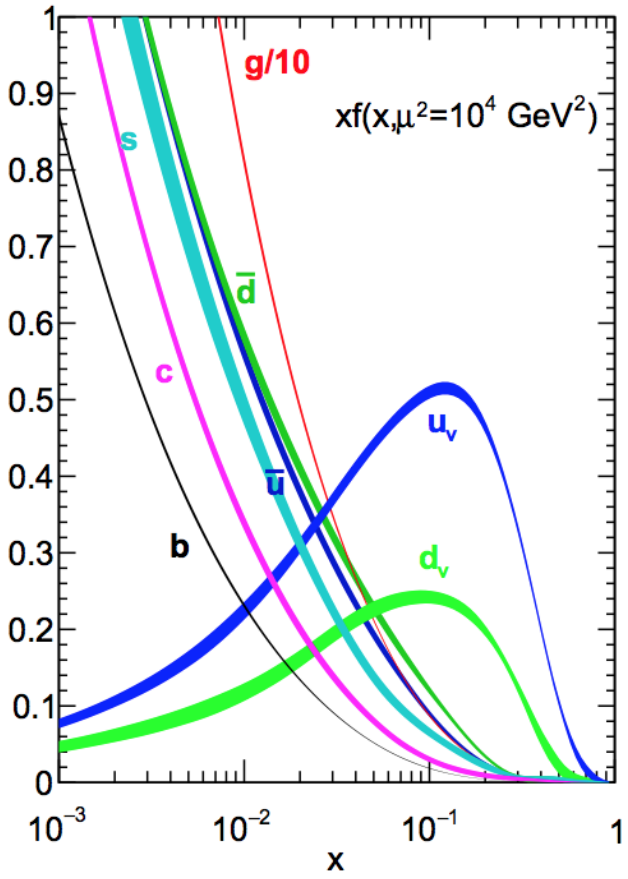

Figure 2.7: The parton distribution functions for protons at $\mu_F^2 = 10^4$ GeV^2. At higher momenta and low x, the gluon PDF is enhanced. Figure reproduced from [33].

2.2.2 Hard scattering

The hard-scattering component of the cross section is calculated at finite order in perturbation theory, for an energy scale above μ_F^2. The matrix element calculation uses all the relevant parameters of the field theory, including the renormalization scale μ_R, up to which the theory is defined. Although such calculations can be done analytically for the simplest processes, tools in MC generators such as MADGRAPH [34, 35] and Powheg [36–38] automatize the process. These generators have several stages but often start by importance sampling events, such that events occur with a probability proportional to the phase space and the matrix element. This step is simple at leading order (LO), but becomes complicated at higher orders, such as next-to-leading order (NLO), when cancellations between real and virtual corrections need to be accounted for.

An example of the different QCD NLO corrections is given in Fig. 2.8. The production of Z bosons via quark-antiquark fusion is defined by the hard process where no extra quark or gluons are produced. Higher-order corrections that include additional real emissions of gluons are shown at LO (NLO), these are proportional to $\alpha_s(\alpha_s^2)$. While these higher order QCD terms generally yield the most important corrections, higher order electroweak (EW) corrections are also important for precise measurements.

2.2.3 Parton shower and jets

At the scale of the hard interaction, there are only a few partons. Due to the confinement property of QCD, the partons in the initial and final state of

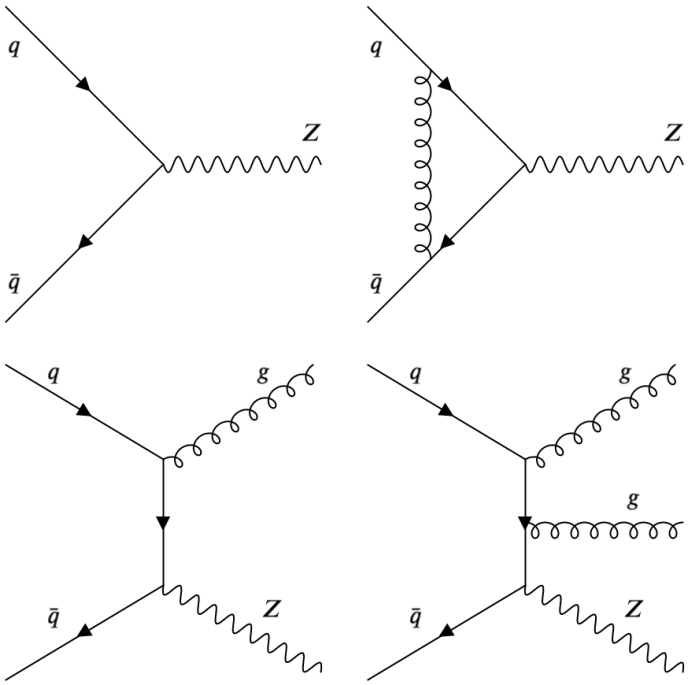

Figure 2.8: The production of Z bosons via quark-antiquark fusion (top-left) and its higher order QCD corrections for virtual (top-right) and real (bottom) contributions.

the hard scatter will split[7]. Such splittings include quark or gluon emission: $q \to qg, g \to qq, g \to gg$. These emissions occur iteratively with decreasing energy scale and develop the *parton shower* (PS).

We can relate the $(n+1)$-parton (post-emission) cross-section to the $n-$parton (pre-emission) cross-section for the initial (final) states:

$$d\sigma_{n+1} \sim d\sigma_n dP_c(z, Q^2), \tag{2.24}$$

where $dP_c(z, Q^2)$ is the probability that a parton c will split into two partons at a scale Q^2, with parton c_i carrying a fraction of momentum z of the original parton's momentum. In a similar way to PDFs, one can define are Altarelli-Parisi splitting functions [30], $P_{c\to c_i}$, that describe the possible parton branchings for a parton c [8] .

The PS evolution follows by the iterative implementation of Eq. 2.24. MC parton shower models, such as Pythia [39] or Herwig [40], provide a useful description of this regime via the construction of Sudakov form factors, which represent the probability that a parton does not undergo a splitting process between two energy scales. These take the form of:

$$\Delta_c(q^2, q'^2) = \exp\left\{ -\sum_{d\in q,g} \int_q^{q'} \frac{dq^2}{q^2} \int_{z_{\min}}^{z_{\max}} dz \frac{\alpha_s}{2\pi} \frac{1}{2} P_{cd}(z) \right\}, \tag{2.25}$$

for two scales q^2 and q'^2. The PS simulation of the final-state radiation (FSR) operates by following a forward evolution whereby partons initially at a

[7] As the hard-scatter partons exit the interaction the coupling will increase with their separation. This increases the probability to radiating gluons, which in turn radiate $q\bar{q}$ pairs and so on

[8] For example, the splitting function for $q \to qg$ is: $P_{q\to qg} = C_f \left(\frac{1+z^2}{1-z}\right)$, where $C_f = 4/3$. This implies that the cross section grows for $z \to 1$ and $\theta \to 0$, i.e. soft and collinear splittings.

scale Q^2 emit radiation at the scale q'^2, determined by sampling $\Delta_c(Q^2, q'^2)$ in Eq. 2.25. Initial state radiation (ISR) showers are those that develop on an incoming parton of the hard scatter and are modeled in the same way as FSR showers except that they evolve backwards from the hard scattering.

The parton splittings are repeated until all the partons have reached a certain energy scale near $Q_0^2 = \Lambda_{\text{QCD}}$. Below this scale (usually around 1 GeV), color confinement effects become important and partons evolve into color-neutral hadrons, preventing further splitting. This process is modeled non-perturbatively and is called *hadronization*.

Hadronization models in MC generators follow the Lund string model [41] to determine the overall multiplicity of the final state particles. The Lund string interprets the QCD field lines between quarks as massless strings that store potential energy. When a string is sufficiently stretched, it breaks and splits the system, leading to the production of a quark-antiquark pair. The implementation of this process is iterative and is tuned to match data observations. Results in this thesis follow the PS models by Pythia 8.230 [39].

When the endpoints of the hadronic shower reach the detector, they appear as collimated sprays of particles. These are clustered into particle **jets**. The algorithms used to cluster, reconstruct and identify jets in an event are detailed in Chapter 4.

Other objects are also produced in association with the final states that result from the hard-scatter. These are called the underlying event (UE). They can result from the fragmentation of beam remnants or the interaction between

the remaining partons of the protons that accompany the hard scattering. Modeling the UE is complicated because of the combination of perturbative and non-perturbative processes. In MC generators, it is simulated using perturbative models that are tuned to match the particle multiplicities observed in data. In this thesis, the Pythia parameters for the underlying event description are set with the CP5 tune as described in Ref. [42]. Since it contributes to extra low-energy particles in the event, the UE also impacts jet reconstruction as described in Chapter 4.

As a summary, the simulation process of a pp collision event in MC is illustrated in Fig. 2.9.

2.3 Beyond the Standard Model

The Standard Model was formulated fifty years ago and its predictive power is immense. Collider measurements of the cross-sections for various known production processes, spanning over 10 orders of magnitude, have been shown to be consistent with the theoretical predictions [43]. These have significantly improved our understanding of physics at the TeV energy scale. Moreover, the discovery of the Higgs boson in 2012 confirmed the role of this particle as a fundamental scalar boson and opened an experimental program to characterize its properties [44, 45].

In spite of this success, the SM conceptual picture of the Universe is not complete and leaves open questions, such as:

- The failure to provide a description of gravity.

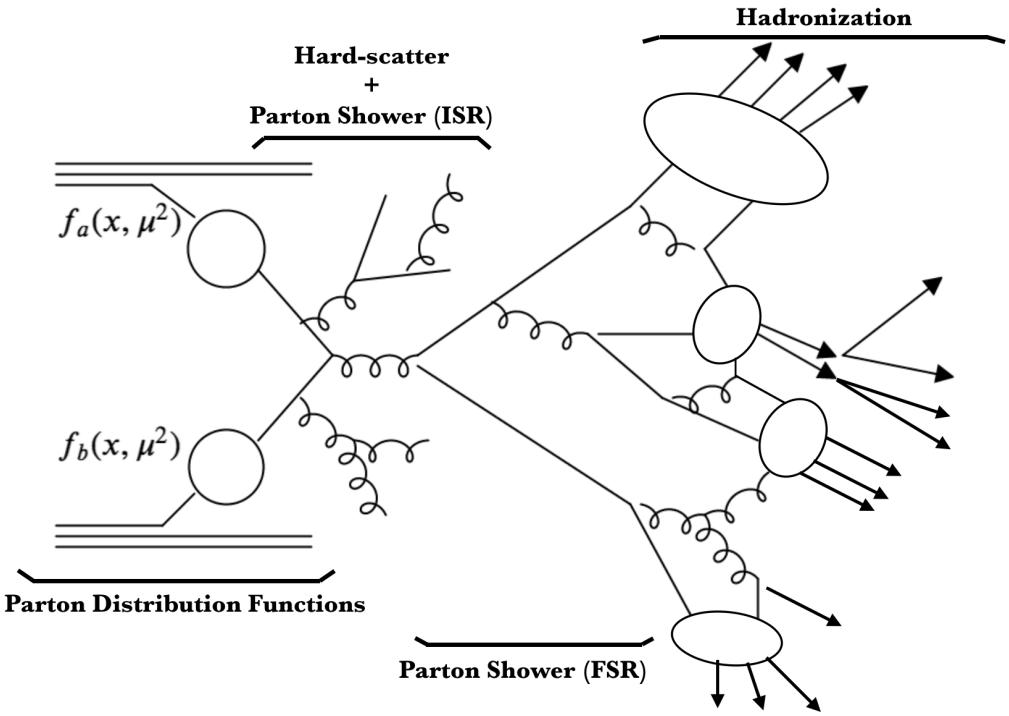

Figure 2.9: Illustration of the Monte Carlo simulation process of a pp collision event. The end result of the hard-scatter + fragmentation + hadronization (ovals) process is shown in the top, indicated by arrows. The gluons emitted from the incident partons constitute the starting point of the initial state radiation (ISR). The ISR-emitted gluons also hadronize but this process is omitted to avoid cluttering.

- The nature of non-luminous dark matter, whose existence has manifested in astrophysical and cosmological observations.
- The nature of dark energy, that accounts for 68% of the energy budget of the universe and has been suggested by CMB measurements, galaxy clusters and other measurements of the universe's expansion rate [46, 47].
- The observed matter-antimatter symmetry in the universe that cannot be explained by the SM CP violation and baryon number violation [48].
- The masses of neutrinos, which are hypothesized to be massless and only left-handed in the SM, but have been shown to be non-zero from neutrino oscillation observations [49].
- The smallness of the electroweak scale v (about 240 GeV), which is much lower than the Planck mass around 10^{19} GeV. This fine-tuning is also called the hierarchy problem.

There have been a multitude of proposals of beyond the Standard Model (BSM) theories of particles and fields that attempt to solve some of these remaining mysteries. In this thesis, we explore a generic signature in proton-proton collisions that could probe certain extensions of the SM. In particular, we focus on models that could provide a leptophobic candidate for a vector mediator that would couple to quarks [50]. This candidate falls within the description of a WIMP dark matter mediator candidate between SM particles and dark matter. In the next sections, we give a short overview of our current

understanding of dark matter with an emphasis on WIMPs and its current constraints from colliders.

2.3.1 Dark matter

The visible universe consists of baryons (or normal matter) and radiation (photons, or light). The invisible universe consists of neutrinos but, most importantly, of dark matter. Dark matter (DM) is a non-baryonic type of matter that represents a quarter of the energy budget of the universe. Understanding the nature of dark matter is one of the most important quests in modern physics.

So far what we know about that dark matter is the following:

- *DM has gravitational mass*. One of the first pieces of evidence of DM came from the discovery that the movement of luminous matter (gas, stars and galaxies) was inconsistent with the motion calculated from their brightness. In 1933, measurements of relative velocities of galaxies in the Coma cluster pointed to the a large amount of internal kinetic energy in the cluster. For the structure to be dynamically stable, the gravitational potential energy required was 20 times what would have been inferred from the luminous matter alone [51]. This indicated the presence of non-luminous matter.

- *DM is important for structure formation of the universe* Measurements of the galactic rotational curves also gave us hints of the presence of dark matter. In 1970, Vera Rubin et.al. measured the rotation velocities of stars and gas clouds with respect to their distance to the center of the

host spiral galaxies, $v(r)$. To a fair approximation, assuming Newtonian gravity and a two-body system in equilibrium, we expect a behavior of the type: $v(r) = \sqrt{GM(r)/r}$. In particular, for outer-radius stars, that would perceive a mass density much lower than the central mass density, we expect $v(r) \propto 1/\sqrt{r}$. Instead, it was observed that $v(r)$ increases in r and eventually plateaus, as shown in Fig. 2.10 [52]. The observed rotational curves $v(r)$ are well-described by a 3 component fit: the visible disk, a gas cloud, and a dark halo. A non-zero component of the dark halo, DM, is needed to support this fast rotation and to, in short, keep galaxies and galaxy clusters from flying apart.

- *DM is abundant* The abundance of dark matter Ω_c can be obtained from cosmological measurements of the cosmic microwave background (CMB) temperature anisotropies. The CMB measures the remnant of photons after their decoupling from matter in the early universe, i.e. at the time when free electrons and protons started forming electrically neutral hydrogen atoms, allowing photons to travel freely [53, 54].

  The power spectrum of the CMB is mostly isotropic but it has anisotropies at the level of about 1 part in 100,000, these are driven by matter anisotropies at the time of decoupling. The spectrum is modified when there are two matter populations (SM and DM) as opposed to one (SM), and one (DM) feels the pull of gravity but not the electromagnetic push from light rays. Different modes, each with their characteristic length scales, oscillate at frequencies that depend on the gravitational potential set by the dark matter abundance; the amplitude of each mode at

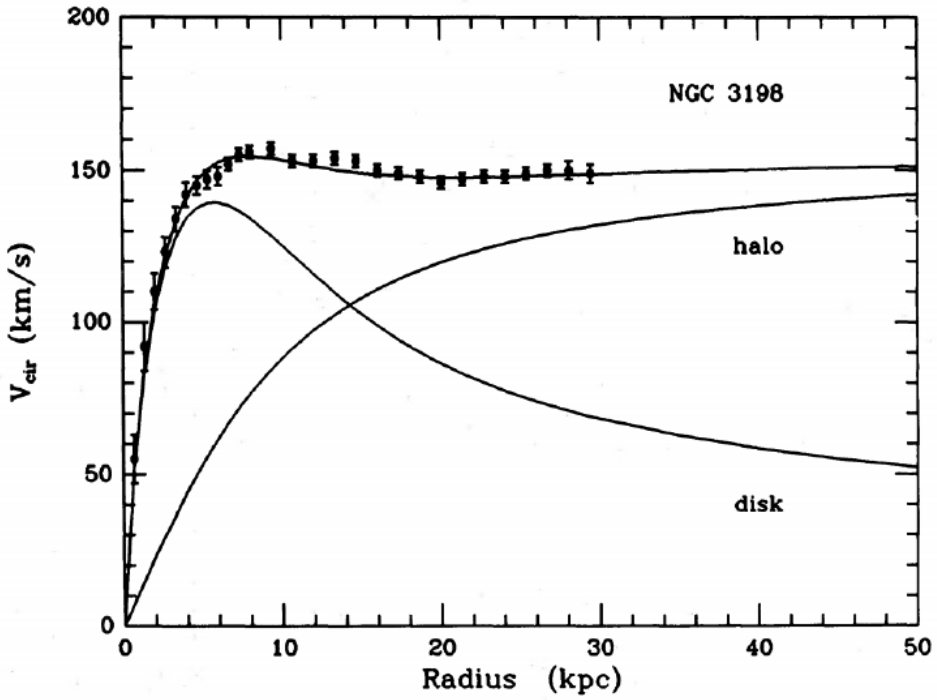

Figure 2.10: Rotation curve for the spiral galaxy NGC 3198. The curve labeled disk shows the expected rotation curve if the surface density distribution followed the surface brightness distribution of the galaxy. The curve labeled gas is the contribution to the rotation curve from the observed gas. Together, the gas and disk cannot reproduce the observed flat rotation curve at large radii. Figure reproduced from [52].

recombination can therefore be turned into a measurement of Ω_c.

The latest results show that $\Omega_c h^2 = 0.1200 \pm 0.0012$, where h is the Hubble parameter in units of 100km s^{-1}Mpc [22, 47]. This represents approximately 25% of all matter in the universe, with a relative uncertainty of about 2%.

Given the incontrovertible evidence in favor of the existence of DM, there is a spirited hunt to discover its nature and there is no shortage of ideas on what this might be. Serious candidates have been proposed with a mass range that covers up to 75 orders of magnitude, from 10^{-5}eV (10^{-71} solar masses) to 10^4 solar masses. Among these, some of the most popular are non-baryonic candidates, which are either undetected elementary particles or new particles with non-standard properties. They can be characterized by how they came to exist in large quantity since the Early Universe, and also by how easy they are to detect.

A very plausible hypothesis for the production of dark matter is that it consists of thermal *relics* of the Big Bang (much like the photons of the CMB). The term relics refers to particles left over when their annihilation reactions were no longer efficient. This implies that when the universe expanded, DM fell out of thermodynamic equilibrium with other fields. The decoupling of the DM field from the hot and dense thermal bath in the universe is called freeze-out. If DM decoupling occurred by freeze-out, then DM was in thermal equilibrium with the SM and must have interacted with SM particles at some energy scale.

One popular candidate of thermal-DM is the Weakly Interactive Massive

Particle (WIMP). WIMP candidates have the following properties:

- *weakly interacting*: they must have a small coupling with the SM. The WIMP interaction is assumed to occur at or around the weak force scale ($\sim 0.1 - 1$ TeV).
- *stable*: or at least long-lived to still be abundant.
- *massive*: or massive enough to be *cold* or non-relativistic at the time of structure formation. The WIMP mass is assumed to be in the GeV range.

The SM neutrino could satisfy some of the requirements for a WIMP, since at least two types of neutrinos are known to be massive. However, constraints on the neutrino mass restrict $\Omega_c h^2 \leq 0.0062$, at most 2% of the energy associated with DM.

We can calculate the expected abundance of thermal relics with two simple assumptions on the interaction strength and mass for a WIMP. It turns out that a stable particle of mass near 100 GeV and interacting with the strength of the weak force will leave a relic density of $\Omega_c h^2 0.1$. That is, it will leave just about the right amount of "leftovers" to account for the observed dark matter density. This coincidence is sometimes popularized as the "WIMP Miracle".

An additional feature of WIMPs is that they appear naturally in many model frameworks designed to understand the weak force, including supersymmetric theories, theories with extra spatial dimensions, and others. In particular, an exhaustive search for theories in supersymmetry (SUSY) has been and still is being pursued in colliders. These introduce a bosonic (fermionic) partner for every fermion (boson) and along with providing a

WIMP candidate [9] , they yield a solution to the hierarchy problem by alleviating the fine-tuning of fundamental parameters [55].

Other models suggest various types of particle DM that also solve other problems in particle physics. Such is the case of sterile neutrinos, neutrinos that do not couple to the Z boson and provide a mechanism for neutrino mass generation; or axions, new light scalar fields that solve the question of why there is no evidence of CP violation in QCD. This is not an exhaustive list and detailed notes can be found in [56].

2.3.2 Dark matter constraints in colliders

If one assumes a weak coupling of particle DM with the SM, there are three mechanisms for its detection:

- via indirect detection of DM annihilation via resonances in cosmic ray energy distributions (protons, electrons, photons...);
- via direct detection of DM scattering against atomic nuclei;
- or via pair production from SM particle annihilation.

These are summarized in Fig. 2.11. The latter is the detection mode that particle colliders follow.

The literature of the DM models probed in colliders can be divided into two extremes. Fully specified, self-consistent models, such as SUSY, that provide specific features that can be exploited in narrowly targeted searches, or,

[9] A popular candidate is the lightest neutralino, a supersymmetric particle that behaves similar to a neutrino. It tends to have masses at the weak scale (100 GeV - 1 TeV) and is stable.

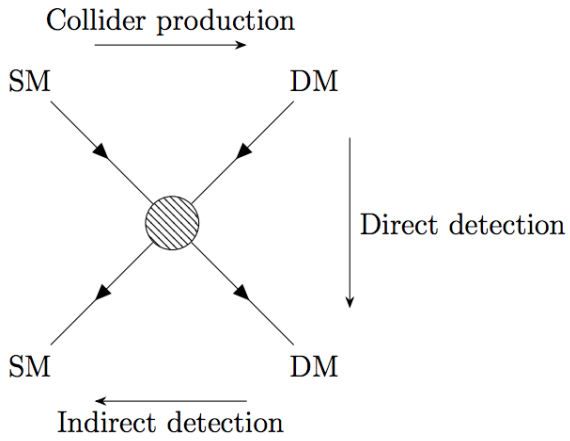

Figure 2.11: Schematic of particle dark matter detection channels. From left-to-right DM is produced from SM particle annihilation, top-to-bottom DM scatters against atomic nuclei and from right-to-left DM annihilates into SM particles.

simplified models, that make as few assumptions as possible on the underlying theory and depend only on a few components to capture broad collider signatures. A third class of models, called portal-models, considers the direct decay of Z or Higgs bosons into DM particles [57]. However, the probability of the decay of the Z boson into invisible particles, i.e. its partial width of $Z \to \nu\nu$, is already constrained at the 10^{-3} level. Similarly, couplings of DM particles to the SM Higgs boson are constrained by measurements of the branching ratio of $h \to \chi\bar{\chi} < 0.34$ [58].

In this thesis, we explore a very generic collider signature: a SM quark-antiquark pair in the final state. Thus, here we explore constraints on a class of simplified models where the DM is a particle annihilating to SM fermions via a new gauge boson mediator. This mediator, called Z', is under U(1) gauge symmetry. A representative Feynman diagram for this process is shown in Fig. 2.12.

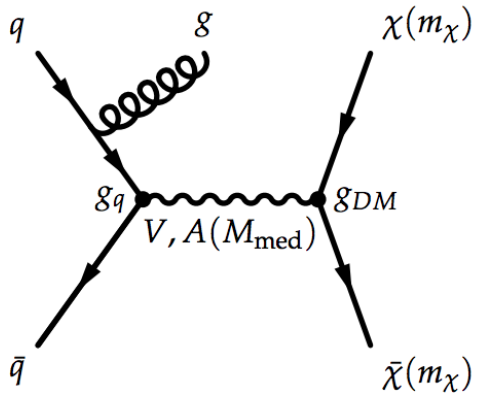

Figure 2.12: Feynman diagram showing the pair production of DM particles (χ) with mass m_χ from the anhilation of SM particles. In this simplified model, the interaction is mediated by a U(1) gauge boson called Z′ with mass m'_Z. The cross section and kinematics depend upon the mediator and DM masses, and the mediator couplings $g_{q'}$ and g_{DM}. Figure reproduced from [1].

We focus on a spin-1 Z′ mediator with universal coupling to all quarks, $g_{q'}$, and vector or axial-vector couplings to DM, g_{DM}. These massive spin-1 bosons are nearly ubiquitous in extension theories of the SM, so Z′ bosons as the mediators connect a wide class of models. The corresponding interaction Lagrangian, for a vector coupling, is:

$$\mathcal{L}_{\text{vector}} = g_{q'} \sum_{q=u,d,s,c,b,t} Z'_\mu \bar{q}\gamma^\mu q + g_{DM} Z'_\mu \bar{\chi}\gamma^\mu \chi \tag{2.26}$$

Just as neutrinos do, DM particles would pass invisibly through the detector. To detect this invisible signature an additional SM interaction is required, the most common addition is ISR radiated from the incoming quarks. The experimental signature of this process usually consists on large missing momentum $p_{\text{T}}^{\text{miss}}$ (from the invisible particles) and the ISR radiation deposited in the detector (typically from a gluon, a photon or Z boson). This collider

search is commonly known as $p_{\mathrm{T}}^{\mathrm{miss}} + X$.

Because invisible particles have feeble interactions with the colliding partons, and thus low production cross sections, the $p_{\mathrm{T}}^{\mathrm{miss}} + X$ searches need precise estimates of the $p_{\mathrm{T}}^{\mathrm{miss}}$ shapes of the background processes, especially in the low-$p_{\mathrm{T}}^{\mathrm{miss}}$ regions. With no excesses observed, these searches set constraints on the production cross section of invisible particles, typically ranging from 0.5 pb to 2 fb, depending on the $p_{\mathrm{T}}^{\mathrm{miss}}$ threshold [59]. These constraints can be used to set exclusion limits on the interactions between the mediator and the SM coupling (e.g. $g_{\mathrm{q}'}$), under specific sets of model assumptions. For example, for vector mediators, masses of up to 1.5âĂŞ1.9 TeV are ruled out for an invisible coupling of $g_{\mathrm{DM}} = 1$ and $g_{\mathrm{q}'} = 0.25$.

In this thesis, we do not explore this signature but, instead, we probe these interactions without actually producing invisible particles. For example, if the mediator particle can be produced via interactions with quarks, it may also decay into quarks. Thus, we focus on the decay of the Z′ boson back into a quark-antiquark pair as shown in Fig. 2.13. These two quarks hadronize into jets in the detector. Further details of this search will be given in Chapter 5.

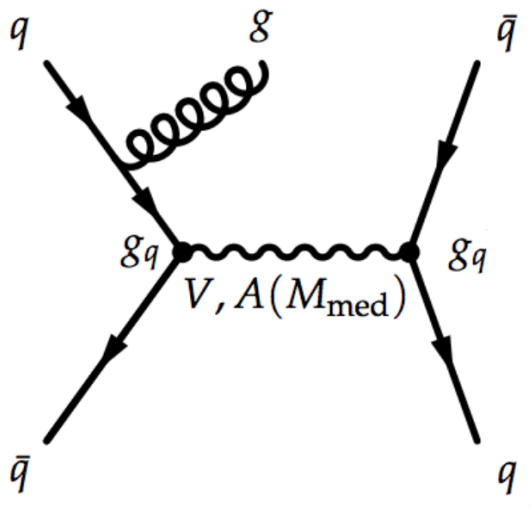

Figure 2.13: Feynman diagram showing the decay of the Z′ boson back into a pair of SM quarks. The coupling of this mediator to quarks is assumed to be universal and denoted as $g_{q'}$.

Chapter 3

Experimental setup

There are two common methods to study particle interactions at high energy and high intensity. Both require the use of a particle accelerator so that a particle beam reaches a given energy. Then, we can either point the beam into a thin target of nuclei or collide two particle beams together. In this study we focus on collisions of two proton beams in a circular accelerator. Protons are stable, abundant, and most importantly they are heavy enough to lose kinetic energy to synchrotron radiation far more slowly than electrons.

In this chapter, we describe in detail the phenomenology of proton-proton collisions at the LHC. We first detail the design of the accelerator and of the CMS experiment that records these collisions and provides the data used in this thesis. We discuss the sub-detector systems of the experiment that are carefully layered and designed to identify different types of particles. Part of the discussion is focused on the inner tracker detector system, that precisely identifies the position of outgoing charged particles from proton collisions. Finally, we detail how the full event reconstruction is performed.

3.1 The design of the Large Hadron Collider

The CERN Large Hadron Collider (LHC) is a circular particle collider, with 27 km in circumference and placed approximately 45 m to 170 m underneath the Switzerland/France border. It was designed to accelerate two proton beams to an energy of 7 TeV and collide them at a center of mass energy $\sqrt{s} = 14$ TeV. During its first run (Run I: 2010-2012), the LHC was able to deliver collisions at $\sqrt{s} = 7$ TeV and during its second run (Run II: 2015-2018) it achieved $\sqrt{s} = 13$ TeV collisions. The target energy will finally be achieved in the next upcoming run (Run 3: 2020-2023). In addition to protons, the LHC also collides heavy ions (Pb) at a lower energy of approximately 2.7 TeV. This thesis only uses proton-proton collision data taken in a sub-period of Run 2 (2016-2017).

The discovery power of the LHC lies on the unprecedented energy and intensity of its proton collisions: the large number of high-energy events collected do not only allow for precise SM measurements but also open a window to probe new physics coupled very weakly to the SM.

To reach a high energy regime, the protons in the LHC are accelerated step by step through an injection chain. This chain is illustrated in Fig. 3.1. The proton source that initiates the chain simply comes from hydrogen molecules which are ionized and stripped of electrons with an electric field. These protons are accelerated to a kinetic energy of 50 MeV by a linear accelerator, the Linac2. The beams of protons are then injected into a series of synchrotron accelerators, the Proton Synchrotron Booster (PSB), the Proton Synchrotron (PS) and the Super Proton Synchrotron (SPS) that increase their energy to 1.4 GeV, 25 GeV, and 450 GeV. Finally two beams enter the LHC at two different

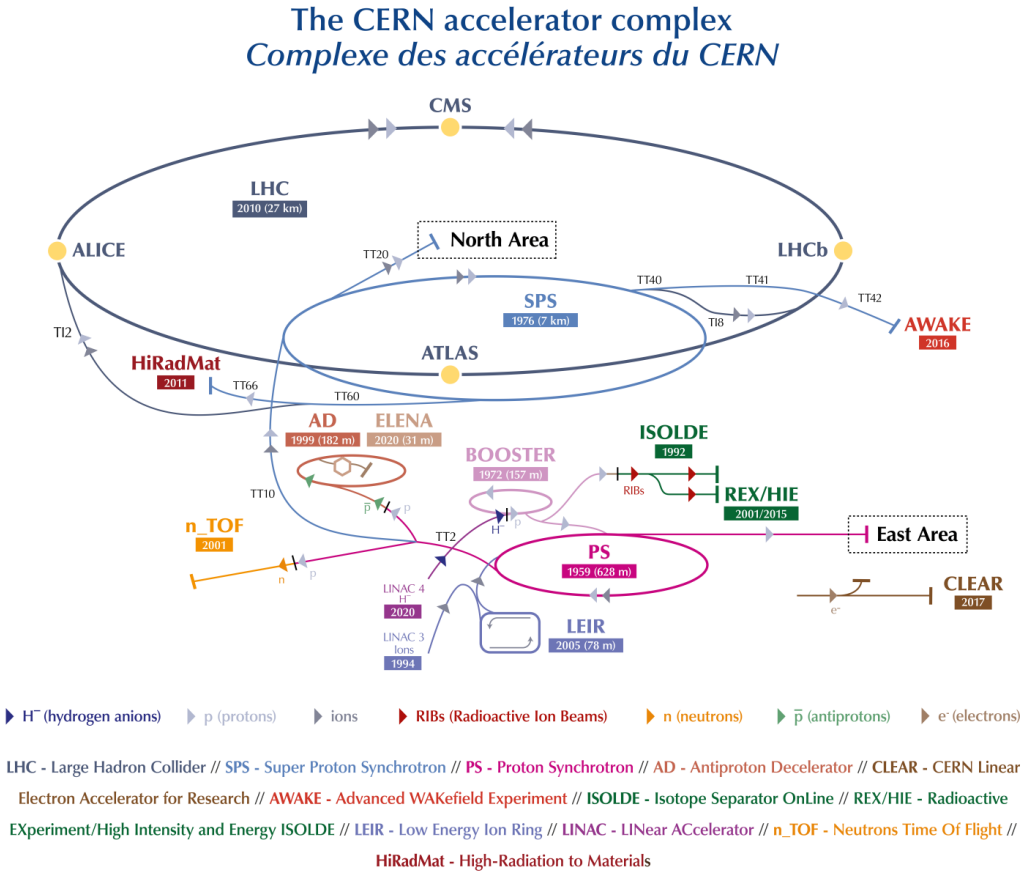

Figure 3.1: The CERN accelerator complex. Protons are injected into the LHC accelerator ring (dark blue) by a chain of intermediate accelerators that begins in LINAC2 (violet). Reprinted from Reference [60].

places circulating in opposite directions.

Inside the LHC, each beam is accelerated by eight super-conducting radio-frequency (RF) cavities. These exert 400 MHz oscillating electric fields parallel to the beam line. The physical and temporal design of the RF system creates bunches of protons, corresponding to nodes of the oscillating field. The proton bunches are effectively shaped by the oscillating RF field: protons in a bunch behind (ahead) of those particles at the center of the bunches will be

accelerated (decelerated) so as to be pushed back into the center of the bunch. Each bunch is approximately 7.5 cm in length and has around 120 billion protons; the bunches are moving at nearly the speed of light, and their leading edges are separated by 25 ns. The bunch-spacing is not only determined by the conditions at the LHC but also by what the detectors surrounding the collision points can tolerate. A 25 ns bunch spacing translates to a maximal pp collision rate of 40 MHz. A different bunch structure may push the detectors at the interaction points beyond their detection and data acquisition limits.

The LHC uses super-conducting NbTi dipoles to generate a magnetic field that bends the proton beams as they travel around the ring. The magnets are cooled down to 2 K using super-fluid helium, and their generated magnetic field is between 0.54 and 8.33 T. Besides the 1238 dipole magnets, there are also a number of quadrupole magnets that are used to focus the beams into a tiny spot as they approach collision.

The proton beams intersect in eight places along the LHC, four of which are instrumented by detectors: CMS, ATLAS, LHCb and ALICE. In the interaction region, two beams share a common beam pipe of approximately 130 m in length. To avoid undesired parasitic collisions from the protons in the bunch, there is a dedicated crossing angle between proton beams. The proton beams are finally dumped at Point 6 using a deflecting magnet complex in both horizontal and vertical directions.

The performance of the collider can be measured in terms of the center of mass energy and the number of collision events recorded per unit time. For a

given process with cross section $\sigma(pp \to X)$, the latter is given by:

$$N = \sigma \int L dt = \sigma L_{\text{int}}, \tag{3.1}$$

where L is the instantaneous luminosity and L_{int} the integrated luminosity. Since σ is fixed, varying the luminosity allows to increase N.

The luminosity depends on:

$$L = \frac{N_b^2 n_b f_{\text{rev}} \gamma_r}{4\pi\epsilon\beta^*} F, \tag{3.2}$$

where N_b is the number of particles per bunch, n_b is the number of colliding bunches per beam, f_{rev} is the frequency of revolution, γ_r is the gamma factor E/m of the beam, ϵ_n is the emittance of the beam, β^* is the beta function at the collision point, and F is the factor that accounts for the cross angle at the point of interaction. The instantaneous luminosity, of units $\text{cms}^2\text{s}^{-1}$, can be thought of as the outgoing flux of particles per unit area and time, after a bunch crossing with successful pp collisions. Since N_b are modified in collisions[1], L varies over time.

At the LHC, the design collision energy is 14 TeV and the designed instantaneous luminosity is 10^{34} cm^2s^{-1}. Its delivered luminosity during Run 2, at 13 TeV of center-of-mass energy, is shown in Fig. 3.2. A common unit for luminosity is the barn (b), one barn is 10^{-24} cm^{-2}. The size of the datasets collected by the LHC experiments are such that the femtobarn (fb), 10^{-39} cm^{-2}. The inverse femtobarn (fb^{-1}) is a measurement of particle-collision events per femtobarn. The integrated luminosity during the period of 2016-2017 is

[1]For example, a dominant source of loss of protons in the beam is due to the burn-off in collision.

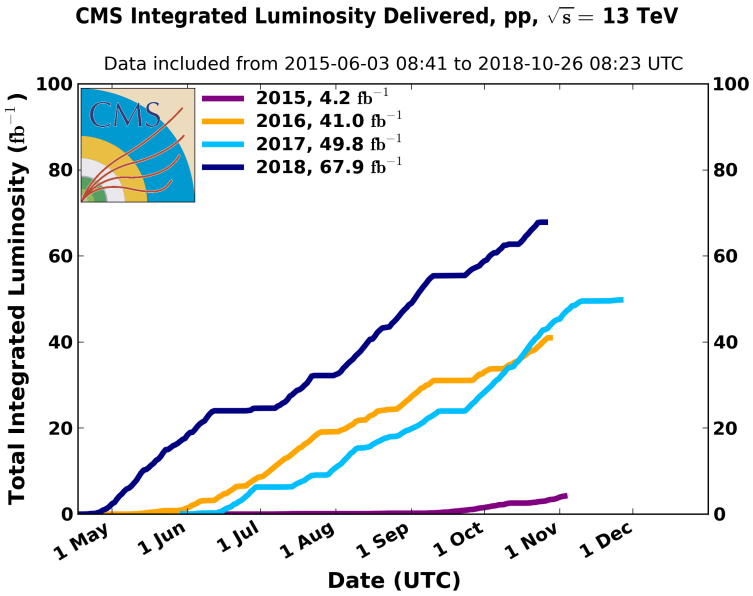

Figure 3.2: Cumulative delivered luminosity in the CMS experiment versus time for 2015-2018 (pp data only). These plots use the final approved physics calibrations for all years. The results in this thesis use data from 2016(orange) and 2017 (light blue) data-taking. Reprinted from Reference [61].

77 $fb^{-1}$and from 2016-2018 is 132 fb^{-1}.

To maximize the probability of having a hard scatter proton-proton interaction the bunches are filled with a large number of protons. This results in more than one interaction per bunch crossing, known as *in-time pileup*. Furthermore, since the bunch spacing is *only 25 ns*, fractions of energy from neighboring bunch crossings, before or after the current one, can still be left-out leading to *out-of-time pileup*. That is, at one instant several parts of the detectors may be looking at particles from different bunch crossings. This represents an experimental challenge in the event reconstruction. A summary of the average number of interactions per bunch crossing can be found in Fig. 3.3.

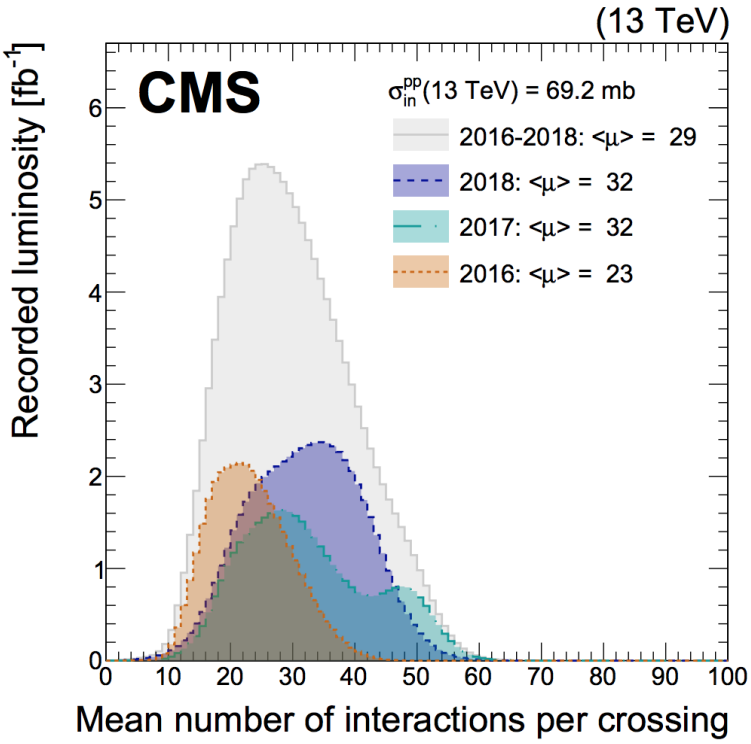

Figure 3.3: Mean number of inelastic interactions per bunch crossing for the 2016 (orange) and 2017 (teal) pp run at 13 TeV. The mean number of interactions per crossing corresponds to the mean of the Poisson distribution of the number of interactions per crossing calculated for each bunch. It is calculated from the instantaneous per bunch luminosity as $\mu = L_{\text{bunch}} x \sigma_{\text{inel}} / f_r$, where L_{bunch} is the per bunch instantaneous luminosity, σ_{inel} is the inelastic cross section which we take to be 69.2 mb for 13 TeV collisions, and f_r is the LHC revolution frequency. The mean number per bunch crossing and year of inelastic interactions is provided in the legend. Figure reprinted from Reference [62].

3.2 The Compact Muon Solenoid detector

The Compact Muon Solenoid (CMS) is one of two multi-purpose particle detectors built at the LHC, the other is named ATLAS. Both have the broad physics goals of discovering and studying the properties of the Higgs boson, and searching for new phenomena at high energies. Both apparati have a cylindrical shape that encapsulates the interaction point and follows a layered design of sub-detectors. These sub-detectors systems are specialized to identify different kind of particles that result from the hard-scatter proton interaction. The particles that can be easily identified because of their mass, charge, or decay of length properties are: muons, electrons, photons, and charged and neutral hadrons. Other particles can be reconstructed by identifying the products of its decay in the detector, e.g. the leptonic decay of τ leptons. Neutrinos escape the detector but its presence can be inferred by accounting for the total energy in the collision, as explained below.

A cut-away view of the experiment and its sub-detectors is shown in Fig. 3.4. There are four main sub-systems: the silicon tracker, that measures the tracks of charged particles, the electromagnetic calorimeter, that measures the energy of electrons and photons, the hadron calorimeter, that measures the energy of charged and neutral hadrons, and the muon detectors, that identify and measure the momentum of muons.

The central feature of CMS is a super-conducting solenoid of 6 m internal diameter, that provides a magnetic field of 3.8 T inside its volume. The magnetic field is essential for bending the trajectory of charged particles, and thus allowing the precise measurement of their momenta and charge. The

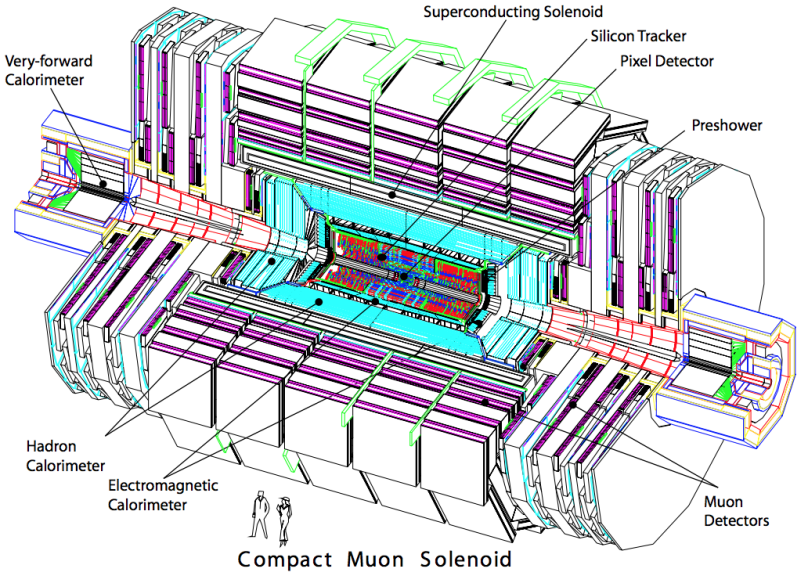

Figure 3.4: A perspective view of the CMS detector. At the center are the silicon pixel and strip trackers. Surrounding the tracker components are the electromagnetic hadronic calorimeters. The barrel hadronic calorimeter is separated into inner and outer regions by the super-conducting solenoid. The hadronic calorimeter endcaps are organized as near and far components to provide better coverage of radiation deposited at larger pseudorapidity (as defined in the text). The muon chambers and the steel yoke of the magnetic field are located at the outermost layers of the detector. Figure reprinted from Reference [63].

space inside the coil is large enough to accommodate for the tracker and calorimetry sub-systems, thereby reducing the amount of material in front of the calorimeters. This is what makes CMS a "compact" apparatus. The return field outside the solenoid is also large enough to saturate 1.5 m of iron, allowing the integration of muon sub-detectors in between.

The (r, θ, ϕ) coordinate system adopted by CMS is centered at the interaction point and measured with respect to the z axis. It is shown in Fig. 3.5. Here, z is along the beam axis, x is pointing inwards towards the LHC ring and y is pointing upward. ϕ is the azimuthal angle measured from x in the plane transverse to z.

The polar angle, θ, is measured from the z axis and defines the pseudo-rapidity $\eta = -ln\tan(\theta/2) = \frac{1}{2}\ln\frac{|\vec{v}|+v_z}{|\vec{v}|-v_z}$. This quantity is invariant under $z-$boosts. Another boost-invariant quantity largely used in the kinematic description of particles, is the transverse momentum vector $\vec{p_T} = (p_x, p_y)$ and, in particular, its magnitude p_T. The CMS detector has full coverage in the azimuthal direction ($0 < \phi < 2\pi$) and partial but good coverage in the polar direction (up to $-5 < \eta < 5$).

3.2.1 Tracker

The tracker is the first layer of the detector that outgoing particles from the collisions encounter and is also the busiest. Because a large number of charged particles emerge from the interaction region, the particle hit density per unit area in each sub-layer is the greatest[2]. The purpose of the tracker is to precisely

[2]For a luminosity of 10^{34} cm^{2}s^{-1}, an average of 1000 particles go through the tracker in each bunch crossing.

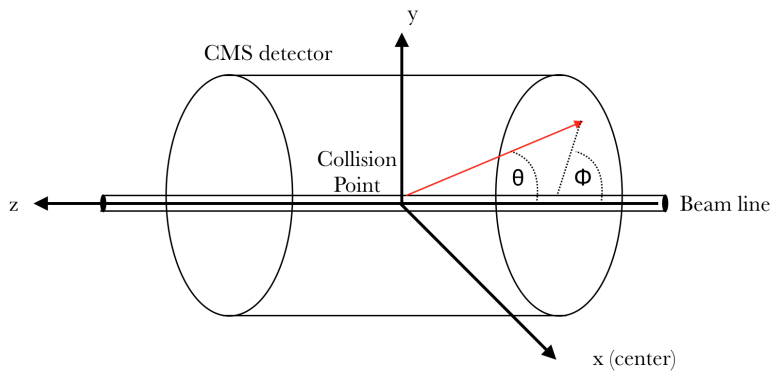

Figure 3.5: Coordinate system adopted by the CMS detector. Figure reprinted from Reference [63].

reconstruct their trajectories or tracks. It exploits the fact that charged particles lose energy by ionization when moving through matter.

The CMS tracker design features high granularity, high precision and fast response in identifying tracks and attributing them to the correct bunch crossing. The tracking volume is composed by a small silicon pixel detector and a surrounding large silicon strip tracker. Both occupy a cylinder of 5.8 m in length and 2.5 m in diameter. They also operate in a similar way: when charged particles pass through the detector, they ionize the silicon. The deposited charge drifts through the sensor to an electrode, then the analog signal recorded by the electrode is digitized, buffered and read out.

The pixel detector is closest to the interaction point and is composed of three barrel layers and two forward disks, located at each of the cylinder, that extend the acceptance to $|\eta| < 2.5$. It provides three-dimensional position measurements of the hits based on the interaction of charged particles with the sensors. In 2017, the pixel detector was replaced with four layers in the barrel region and three disks in the forward region. The surrounding strip

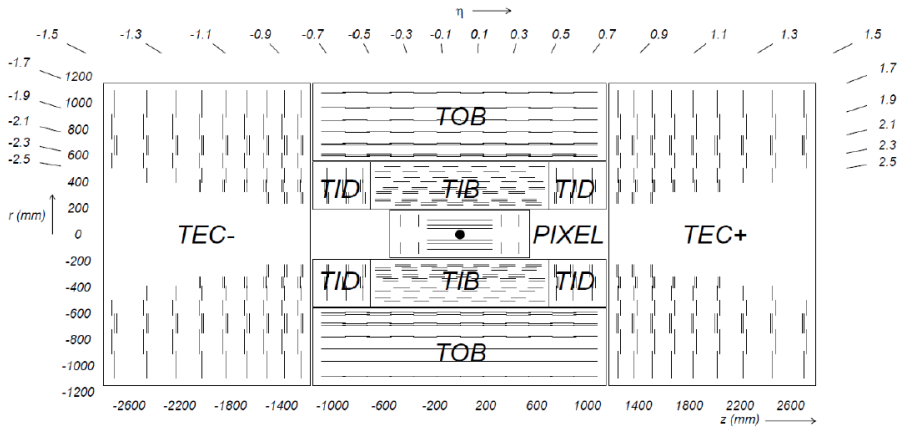

Figure 3.6: Detailed drawing of the locations of the subsystems of the Phase-0 CMS tracker. Each line represents a detector module. Double lines indicate double-sided modules (explained later in the text). Both the Pixel and the Strip modules are represented here, the latter labeled as TID,TIB,TEC and TOB. The Pixel detector is the innermost part and is surrounded by the outer tracker layers. Figure reproduced from [63].

tracker is composed of ten layers of micro-strip detectors in the barrel region and twelve in the forward region and provides a two-dimensional position measurement. We discuss details of these sub-detectors below. The design shown in Fig. 3.6 refers to the Phase-0 version of the tracker.

Pixel detector The design of the pixel detector is strongly driven by figures of merit in tracking: a good momentum and impact parameter resolution. The latter needs a precise single-point estimate in $r - \phi$ and is important for the reconstruction of secondary vertices, i.e. those not coming from the primary interaction point. The pixel layers in CMS are able to provide a single-point position resolution of approximately 10 μm in the $r - \phi$ plane and $20 - 40\mu$m in the longitudinal coordinate $z(r)$.

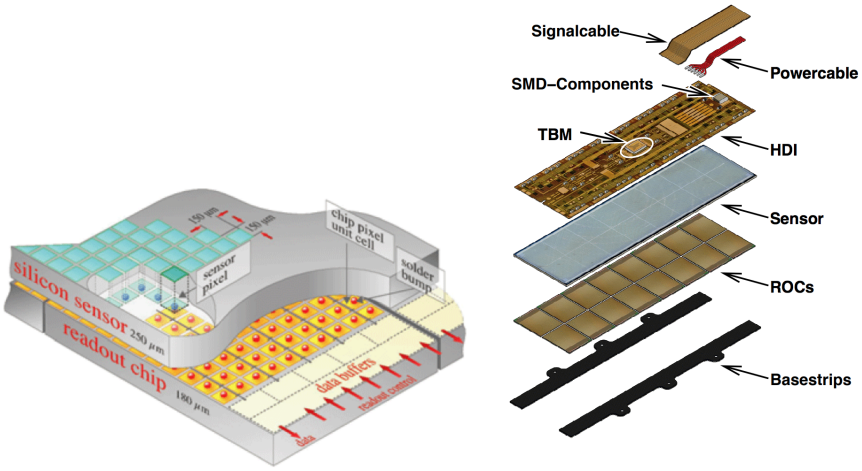

Figure 3.7: Structure of the pixel detector modules (a) and exploded view of a barrel module (b) [64]. Each module consists of a detector diode array (sensor) bump-bonded to readout chips that amplifies the signal.

The CMS pixel detector contains modules that are designed in the following way. Rectangular arrays of highly segmented silicon diodes of approximately 285 μm of thickness, are bump-bonded to corresponding readout chips (ROCs). There are 8 or 16 ROCs, each of which have 52×80 pixel sensors. The pixel sensors themselves measure 100×150 (μm^2). The upper layer of each module consists of a High Density Interconnect (HDI) printed circuit board that distributes the signal and power to the chips. The HDI is also equipped with a Token Bit Manager (TBM) chip that controls the readout of the ROCs. This arrangement is shown in Fig. 3.7.

A voltage is applied across the module so that when a charged particle passes through and creates electron-hole pairs in the silicon bulk, the charges drift to the front of the module and are collected by the ROC. Although the charge is collected by individual pixels, the ROC measures the distribution of charge across neighboring pixels. From this shape, we can determine the

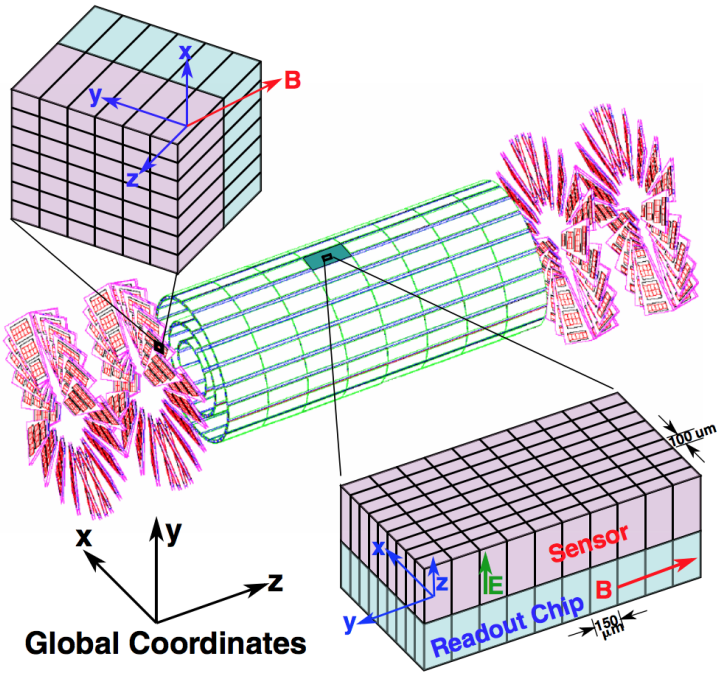

Figure 3.8: Sketch of the CMS pixel detector and examples of the local coordinate systems with respect to the global coordinate system. It is generally arranged into barrel layers (green) and endcap disks (pink). Figure reproduced from [65].

position of the hit to a precision of tens of microns (one order of magnitude smaller than the pixel size).

Sensor modules are arranged in a barrel region (BPIX) and forward disks (FPIX), as shown in Fig. 3.8. The BPIX is constructed as two 53 cm-long half-barrels, that contain carbon fiber ladders of thickness 0.24 mm on which the sensor modules are mounted. The orientation of the modules on each ladder alternates so that modules are facing towards or away from the interaction point. The FPIX detector sections at each end of the BPIX are split vertically down the middle into two half-cylinders so that the detector can be installed around the beam-pipe.

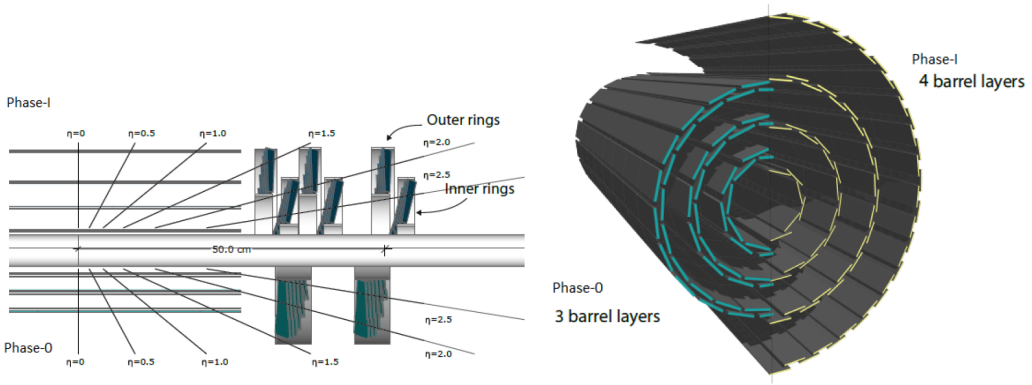

Figure 3.9: Comparison of the Phase-0 and Phase-1 pixel detectors. The left panel shows the conceptual layout of the different layers and disks in the Phase-0 and Phase-1 configurations, and the right panel compares the pixel barrel layers in a transverse-oblique view. Figure reproduced from [66].

In 2017, the detector was upgraded by moving the inner layer closer to the interaction point, from $r = 4.4$ cm to $r = 2.9$ cm, and adding a new outer layer at $r = 16$ cm and new disks in the forward region. This new design is referred as the Phase-1 version and its differences with respect to Phase-0 are shown in Fig. 3.9. Furthermore, since the increase in luminosity can cause inefficiencies in the internal ROC buffers, the capabilities of the readout chip were also improved to cope with a luminosity of 2×10^{34} cm^2 s^{-1}.

Strip detector After traversing the pixel layers, the charged particles enter the silicon strip system. Strip detectors measure the position of a traversing particle in one direction. They are formed by segmenting the pn-junction in one direction, effectively dividing the detector in several diodes. The silicon strips are used at larger radii from the interaction point, where there is a reduction in number of hits per area. The strips used in the tracker have a typical cell size of 10cm $\times$ 80μm to ensure that the single strip occupancy is

low. In the outer tracking region the cell size increases by making the strips longer and allowing for a wider strip separation.

The detector consists of several regions: the tracker inner barrel and discs (TIB/TID), the outer barrel (TOB) and the tracker endcaps (TEC). The TIB/TID deliver four measurements of a particle trajectory in the $r-\phi$ direction and cover $r < 55$ cm and $|z| < 118$ cm. They provide single position measurements with a resolution of approximately $13-38$ μm. The TOB (TEC) consists of six (nine) layers providing measurements also in the $r-\phi$ direction, both with a resolution of approximately $18-48\mu$m. To provide the measurement of a second coordinate a second strip detector module is mounted back-to-back in some cases. These are called stereo modules and are rotated on a stereo angle of 100 mrad with respect to regular modules. The hits from these two modules provide a measurement of z in the barrel and r on the disks.

The tracking system has to maximize the number of measurement points for each particle trajectory while keeping the material budget at a minimum. The amount of interacting material can be measured in terms of radiation length X_0, which corresponds to the distance over which the energy of a charged particle is reduced by a factor of e. Fig. 3.10 shows that the total thickness of the tracker is between 0.4 and 1.8 radiation lengths X_0.

3.2.2 Calorimeters

Calorimeters are detectors that stop most of the particles of a particular kind, and measure the energy that these deposit. There are two type of calorimeters installed in CMS: the electromagnetic calorimeter (ECAL), that is used to

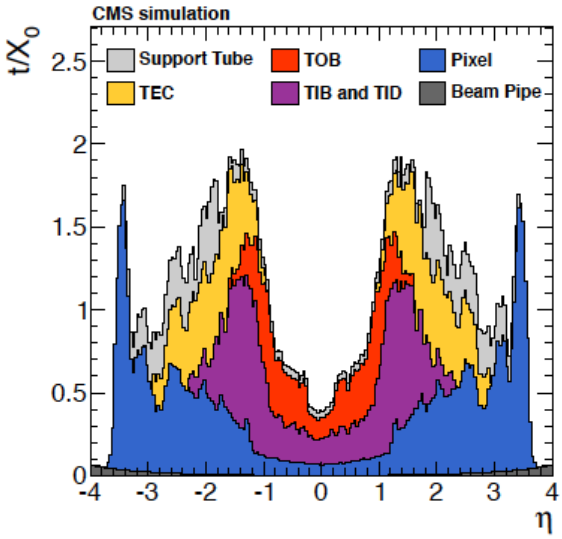

Figure 3.10: Total thickness t of the tracker material traversed by a particle produced at the nominal interaction point, as a function of pseudorapidity η, expressed in units of radiation length X_0. Figure reproduced from [67].

measure the energy of electromagnetic particles such as electrons and photons and the hadronic calorimeter (HCAL) that is used to detect energy deposits from charged and neutral hadrons.

Electromagnetic calorimeter The electromagnetic calorimeter (ECAL) is a calorimeter that is just outside the tracker and within the super-conducting solenoid volume. Its main function is to measure the energy of light particles through the production of scintillation light from electromagnetic cascades. These cascades can be produced directly by charged electrons or by photons, which have no charge but can produce electron-positron pairs or directly interact with an electron in the calorimeter.

The ECAL is composed of lead-tungstate ($PbWO_4$) crystals that are mounted in the barrel and endcap and provide coverage up to $|\eta| < 3.0$. These crystals

are very dense ($8.28 g/cm^3$) and have a short radiation length (0.89 cm). The decay time of the scintillation light is also short, so that most of the light is emitted within the 25 ns bunch spacing of the LHC. The blue-green light (420-430 nm) that the crystals emit is proportional to the energy of the particle. It is converted to electric current by avalanche photo-diodes (vacuum photo-triodes) in the barrel (endcaps), which provide an amplification factor of 50 (10). The efficiency of the photo-detectors is such that a particle depositing 1 MeV of of energy yields around 4.5 photoelectrons.

The crystal material also causes about two thirds of the hadrons to start showering in the ECAL. In particular, the decay of highly energetic neutral pions into photons, $\pi^0 \to \gamma\gamma$, may merge into a single crystal. To differentiate between one- and two-photon deposits, a two-layer lead absorber and silicon strip sensor detector is located between the endcaps and the interaction point. This finer-grained detector is called pre-shower and enhances the capabilities to identify a photon-initiated shower in the lead. This is possible by resolving the shower of incident photons in the silicon strips, that have a resolution of $1 - 10$ mm. The layout of the detector is shown in Fig. 3.11.

The ECAL barrel energy resolution for electrons is measured in an electron test beam to be [68]:

$$\frac{\sigma_E}{E} = \frac{2.8\%}{\sqrt{E/\text{GeV}}} \oplus \frac{12\%}{E/\text{GeV}} \oplus 0.3\%, \tag{3.3}$$

where the three contributions are the stochastic, noise, and constant terms. The actual energy resolution for electrons and photons is measured in CMS using data from decays of known particles into electrons and photons, such as

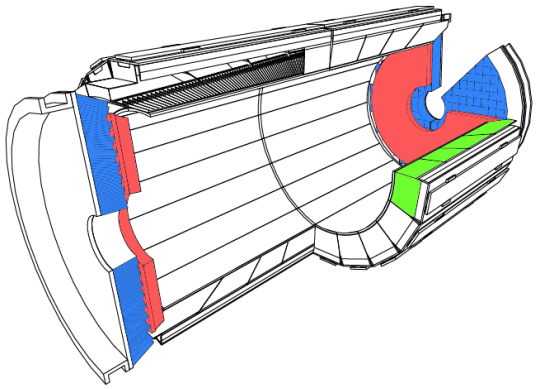

Figure 3.11: Layout of the CMS ECAL, showing the barrel modules (green), the two endcaps (blue), and the pre-shower detectors (red). The ECAL barrel coverage is up to $|eta| < 1.479$ and the endcaps extend the coverage up to $1.479 < |eta| < 3.0$. The pre-shower fiducial area is approximately $1.65 < |eta| < 2.6$. Figure reproduced from [63].

$Z \to e^+e^-$ and $h \to \gamma\gamma$. These measurements result in an energy resolution of $1 - 3\%$ similar for electrons and photons [69, 70].

Hadronic calorimeter Heavier particles, such as hadrons, pass through the ECAL depositing little energy. Its energy is measured in the hadronic calorimeter that surrounds the ECAL. The hadronic calorimeter (HCAL) is designed as a sampling calorimeter, not all of the energy is measured. It consists of several alternating layers of brass absorber and plastic scintillator.

When hadrons pass through the HCAL, they interact with brass nuclei, lose their energy and produce more and lower energy hadrons. The energy of the particles in the hadronic shower is measured by the scintillator layers. The scintillation light is converted by wavelength-shifting fibers embedded in the scintillator tiles and channeled to photo-detectors via clear fibers. The photo-detectors (hybrid photo-diodes) are designed to operate in high axial

magnetic fields.

The barrel (HB) and endcap (HE) components, inside the solenoid, provide coverage up to $|\eta| < 3.0$. The dimensions of the absorber layers, whose thickness ranges from 40 to 75 mm, are limited by the constraint that the HB and HE be inside the solenoid 1.77m $< r <$ 2.95m. To augment the number of interaction lengths, additional layers of scintillator are placed outside the solenoid. This tail-catcher in the outer barrel region, labeled HO, ensures that hadronic showers are sampled with nearly eleven hadronic interaction lengths. Finally, to capture radiation that travels almost parallel to the beam line, a forward component (HF) is situated at $\pm$11m of the interaction point and provides coverage up to $|\eta| < 5.0$. The HO uses the steel return yoke as the absorber, instead of brass, while the HF consists of a steel absorber plates, read out by photo-multiplier tubes. The layout of the detector is shown in Fig. 3.12.

The signals from the HCAL subsystems are grouped so as to define calorimeter towers. These correspond to a small segment in $\Delta\eta \times \Delta\phi$. The HCAL energy resolution has been determined in test beams using single pions and found to be [71]:

$$\frac{\sigma_E}{E} = \frac{110\%}{\sqrt{E/\text{GeV}}} \oplus 9\%, \tag{3.4}$$

with a typical readout noise of 200 MeV per tower.

3.2.3 Muon chambers

The final and outermost detector layers are the muon chambers. Due to the large amount of material preceding the muon system ($\sim$ 16 hadronic

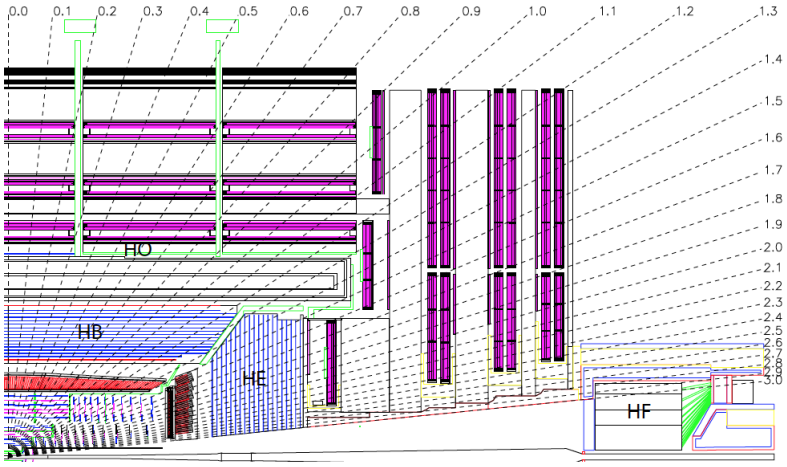

Figure 3.12: Layout of one quadrant of the CMS HCAL, showing the barrel component (HB), the endcap (HB), the forward (HF) and the outer calorimeters (HO). The slight overlap of the detectors in *eta* ensures hermeticity. Figure reproduced from [63].

radiation lengths) the punch-through of hadronic particles in the muon system is negligible. Muons, in the other hand, are too heavy to be stopped by the ECAL but not heavy enough to be stopped by the HCAL.

The muon system in CMS is composed of gaseous detectors sandwiched among the layers of steel flux-return yoke that allow a traversing muon to be detected at multiple points along the track path. They all operate the same principle: as the muons transverse the gas they knock electrons of gas atoms. The electrons are collected by wires or strips and provide a measurement of the energy of the muon.

Drift tube (DT) chambers and cathode strip chambers (CSC) detect muons in the regions $|\eta| < 1.2$ and $0.9 < |\eta| < 2.4$, respectively. These are complemented by a system of resistive plate chambers that cover a range of $|\eta| < 1.6$. The layout of the detectors is shown in Fig. 3.13.

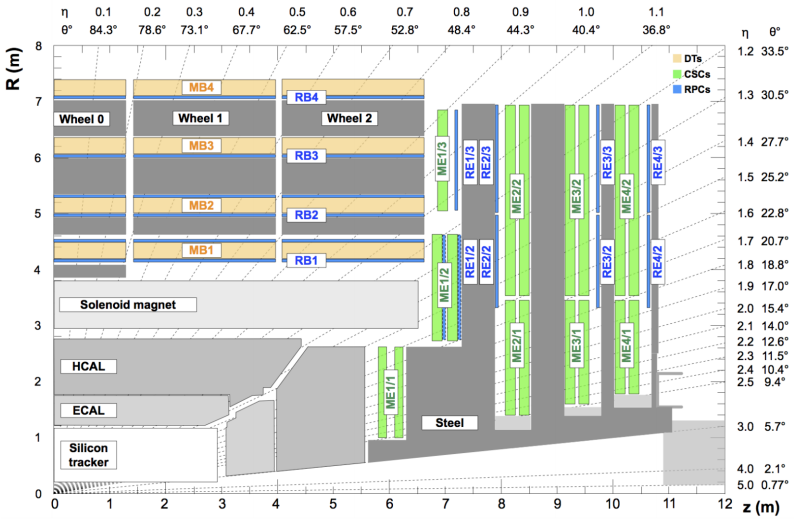

Figure 3.13: Layout of one quadrant of the CMS muon system. The drift tube stations (DTs) are labeled MB ("Muon Barrel"), the cathode strip chambers (CSCs) are labeled ME ("Muon Endcap"). The resistive plate chambers (RPCs) are mounted in both the barrel and endcaps of CMS where they are labeled RB and RE respectively. Figure reproduced from [72].

The drift chambers provide a timing resolution of a few nanoseconds that allows the muon system readout to be assigned to a bunch crossing. They also provide measurements of the r-ϕ (r-z) position with a resolution of 78-120 μm (140-390 μm) [72]. The CSCs can operate at the high rates and non-uniform magnetic field in the endcaps and their resolution varies from 40 to 152 μm. The spatial resolution of the RPC hits is worse than the DTs and CSCs, but the timing resolution improves the efficiency to record interesting muon events.

3.2.4 Trigger and data acquisition

The cross sections of interesting SM processes are several orders of magnitude below the inclusive pp cross section at the LHC, as shown in Fig 3.14. To produce a significant number of rare and interesting collision events, the LHC

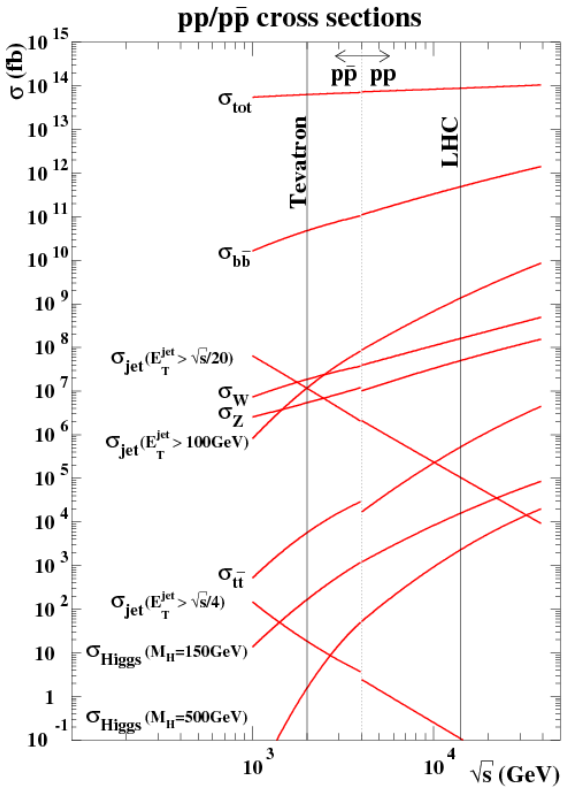

Figure 3.14: Standard Model cross sections at hadron colliders as a function of the machine center of mass energy, $\sqrt{s}$. Interesting SM processes such as the pair production of top quarks, or the production of the Higgs boson have cross sections several orders of magnitude below the inclusive pp cross section. Figure reproduced from [73].

collides proton beams every 25 ns (equivalent to 40 MHz). Only a small fraction of these events will be of interest, and only a small fraction of them can be recorded. The maximum rate of the data acquisition and storage system is of the order of 1 kHz, so the data needs to be reduced by a factor of 10^6. This reduction needs to be achieved while maintaining the largest possible acceptance of interesting physics signals. This is the role of the hardware selection algorithm, called trigger.

The CMS experiment uses a two-tiered trigger system. The first tier, known as the Level 1 (L1) trigger, is composed of custom hardware processors. It

uses incomplete detector information from the calorimeters and muon system to select events in a time interval of less than 4 μs. The second tier, known as the high-level trigger (HLT), consists of a farm of processors running a version of the full event reconstruction software optimized for fast processing. It further reduces the event rate from around 100 kHz to less than 1 kHz before data storage. The final selected data rate is 400 Hz and corresponds to a total reduction of 99.9975%.

Level 1 Trigger The L1 trigger makes fast decisions using field programmable gate arrays (FPGAs) and application specific integrated circuits (ASICs). The latter are typically used when limitations like high radiation doses prevent the use of FPGAs. The L1 selection algorithm is composed of *trigger primitives* (TP). These are simple objects reconstructed from signals of individual detector systems: calorimeter towers of energy deposits and hits on the muon stations. The signals from the inner tracker are not included in the L1 decision because their readout and reconstruction algorithms are too slow to fit into the 4μs L1 time window.

The TP information is combined using regional algorithms that operate in limited spatial regions of the detector. These place quality selections on calorimeter towers, and aggregate them into clusters of energy deposits. For muon primitives, a simple segment-finding and tracking algorithm produces muon tracks. These objects are compared by the global calorimeter trigger and the muon trigger. They determine whether there are sufficient good-quality calorimeter or muon objects to accept the event. In case the event passes the L1 trigger algorithms, also called seeds, the full detector is read out producing

an output at a rate of 100 kHz.

High Level Trigger The HLT software is responsible for reducing the data rate to 400 MHz. It consists of a streamlined version of the offline reconstruction algorithms that run on a CPU farm. It exploits the same software used for offline event reconstruction but optimizes it in order to comply with the strict time requirements of around 300 ms. The total amount of data for a bunch-crossing is on the order of 1 Megabyte (MB) per event [63].

The HLT also consists of a series of paths, or algorithms, that must quickly make a decision about whether or not to accept or reject an event for storage. But, unlike the L1 trigger, the HLT paths make use of the full detector information including the inner tracker. Each path is a sequence of reconstruction modules (producers) and filtering modules (filters). Producers usually perform a given calculation and generate new information, while filters typically select events based on the properties of a given physics object (e.g. high-p_{T} jets). The offline reconstruction of these objects will be detailed below.

Data Storage If an event passes the HLT, it is transferred to the storage manager. It is stored locally on disk and eventually transferred to the Tier-0 computing center, which performs a first pass of offline event reconstruction. The processed data is transferred to several Tier-1 centers for storage, where data can be re-processed with improved calibrations at a later stage. Data analysis and generation of MC simulated events happens primarily at Tier-2 centers, which provide limited disk space, and no tape archiving.

Processing of data almost always occurs promptly but it can be delayed

and sent to other data streams. The scouting data stream reduces the event size from the default of 1 MB to 10-1.5 kilobyte (kB). This allows to increase the recorded event rate and thus increase physics signal acceptance. This data stream is particularly useful to explore new physics channels that need very low trigger thresholds. The parking data stream sends the full raw events from the scouting stream directly to tape without performing reconstruction. This provides a complementary set of collision events to perform new physics analyses or improve the existing ones.

3.3 Detector response simulation

The detector response to particles produced in collisions is simulated with the GEANT4 software package [74, 75]. GEANT4 is interfaced with the particle-level events, simulated by MC event generators that include the matrix element, hard-scattering and parton shower, as described in Sec 2.2.

GEANT4 implements an accurate simulation of the passage of particles through matter. It can therefore simulate the passage of particles through the magnetic field, the energy deposited in the detector when the particles interact with the detector material, and the evolution of any showers of particles produced in these interactions. The simulation contains an accurate description of the detector geometry configuration, the behavior of the sensitive elements in the sub-detectors, hit collections and the modeling of the readout electronics response. In order to account for the effect of pileup, multiple simulated proton-proton collisions are overlaid into a single event. The algorithms used

to reconstruct the event signatures, described in Sec. 3.6, are the same in simulation and data. Finally, a full emulation of the trigger system (L1+HLT) is added to mimic the event selection process.

3.4 Reconstruction in the Inner tracker

Before providing a description of the full event reconstruction in CMS, we focus on describing the reconstruction in the inner tracker. Identifying the position hits in the silicon sensors is the first step in processing the signals from the inner tracker prior to track reconstruction. We begin with a thorough description of the sensors and the radiation damage that they sustain, being the detectors closest to the interaction point. The simulation of this radiation damage is crucial to maintain accurate hit position estimates and good resolution throughout the lifetime of the detector. We describe its implementation and usage in the CMS pixel hit reconstruction. Finally, we briefly describe the trajectory and vertex reconstruction.

3.4.1 Pixel sensors

The pixel silicon sensor design in CMS consists of an n-type bulk silicon, where electrons are loosely bound and the majority are carriers. A p-n junction is formed at the back side of the n-bulk by implanting high positive (n+) dose regions. The other side is implanted with acceptor impurities (p+). In the p-n junction, electrons from the n-type side and holes from the p-type side diffuse across the interface until thermal equilibrium is reached. By applying an external potential, from the p-side (-) to the n-side (+), called "reverse bias", the free charge is swept and the depletion zone grows, with only a small leakage current flowing. That depleted n-type zone is the tracking sensor. A sketch of this design is shown in Fig. 3.15.

When a charged particle traverses the depletion zone and ionizes atoms

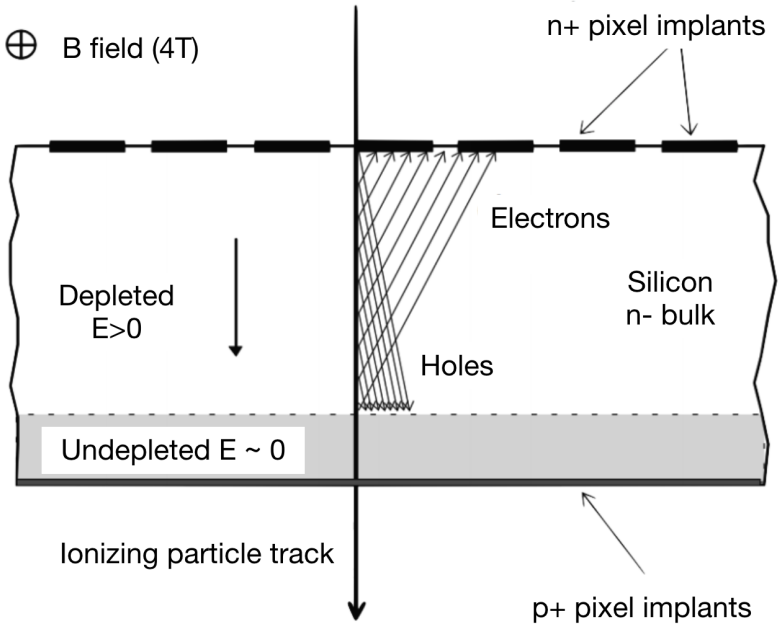

Figure 3.15: Illustration of the sensor charge collection and sharing in the CMS pixel barrel sensor array. The sensor array is made from an n-doped bulk silicon wafer of approximately 300 μm thickness. One side of the wafer is implanted with acceptor impurities at much larger density (labeled p+). The other side is implanted with an array of donor implants of high density (labeled n+). The n+ implants are metalized and held at a high enough voltage so that all free charge is swept out of the sensor leaving a non-zero electric field across detector (fully depleted). If the applied voltage is too small, the field will vanish in part of the bulk (partly depleted). When a charge particle traverses the detector, electron hole pairs are created. The electrons drift under influence of the field to the n+ implants and the holes drift to the p+ implant. To improve the position resolution the charge should be collected by at least two pixels. This is possible when the charge drifts induced by the Lorentz force in the presence of the magnetic field.

along its track, there is no free charge present to extinguish the liberated electrons and holes. They drift along the external electric field and deposit charge in the electrodes on the opposing surfaces. The analog pulse height is read by the electronics. Pixels with a signal greater than a tunable readout threshold (usually around 1500 Q_e) are read out. These pixels are then aggregated with adjacent signals to form pixel clusters, which also obey readout thresholds (approx. 4000 Q_e).

Tracks that enter the sensor at an angle (as seen in Fig. 3.15) deposit charge in several pixels. This is because the drift of the electrons and holes in the sensors gets influenced by the magnetic field B of CMS in the direction $E \times B$. The resulting Lorentz drift leads to the spreading of the collected signal charge over more than one pixel. Charge sharing between neighboring pixels, due to Lorentz drift or tilted sensor position, is a favorable process. This is because single-pixel clusters can achieve a fixed resolution of pitch$/\sqrt{12}$, [3] while two-pixel clusters or larger can use charge interpolation to estimate the position. Charge interpolation leads to an improvement of the position resolution.

3.4.2 Radiation damage simulation

The operation of silicon detectors so close to the interaction point results in the degradation of the sensor properties. Radiation damage in silicon can be

[3] This assumption can be derived from the average difference between the "real" impact position x_r and the measured impact position $x_m = 0$, in a pixel with pitch size p, hit by a uniform density of particles $D(x) = 1$:

$$\sigma^2_{\text{position}} = \frac{\int_{-p/2} p/2 (x_r - x_m)^2 D(x_r) \mathrm{d}x_r}{\int_{-p/2} p/2 D(x_r) \mathrm{d}x_r} = \frac{p^2}{12}.$$

caused by non-ionizing interactions from heavy particles and nuclei. These modify the sensor bulk and can: (1) alter the collected charge, (2) increase the leakage current and (3) change the operational voltage [76, 77].

The main modification to the sensor bulk comes from the displacement of a silicon atom out of its lattice site that results in a silicon interstitial state and a leftover vacancy. Both can migrate through the sensor and form clusters and point defects in the silicon lattice, that have energy levels in the middle of the forbidden gap. When activated and occupied, these states act as trapping sites and *reduce the collected charge*. They further act as recombination/generation centers and lead to an *increase in the sensor leakage current*. This increase is proportional to the fluence received, $I_{\text{leak}} \propto \Phi$, and translates to an increase in noise.

Finally, the effective doping concentration, the difference of all donor-like states and all acceptor-like states, can change. Before irradiation, the depletion region grows from the back side of the sensor towards the pixel n+ implant. After irradiation, the effective doping concentration decreases with increasing fluence until the sensor bulk undergoes space-charge sign inversion (or type inversion) from n-type to p-type. The depletion region behavior is now p-material like and grows from the n+ implant towards the back side of the sensor. This is shown in Fig. 3.16. Further irradiation leads to a *gradual increase of the depletion voltage*. These effects can be further complicated by the temperature history, since the thermal motion in the silicon lattice leads to an annealing current that causes new defects to be formed or existing defects to dissociate, canceling the damage into the lattice.

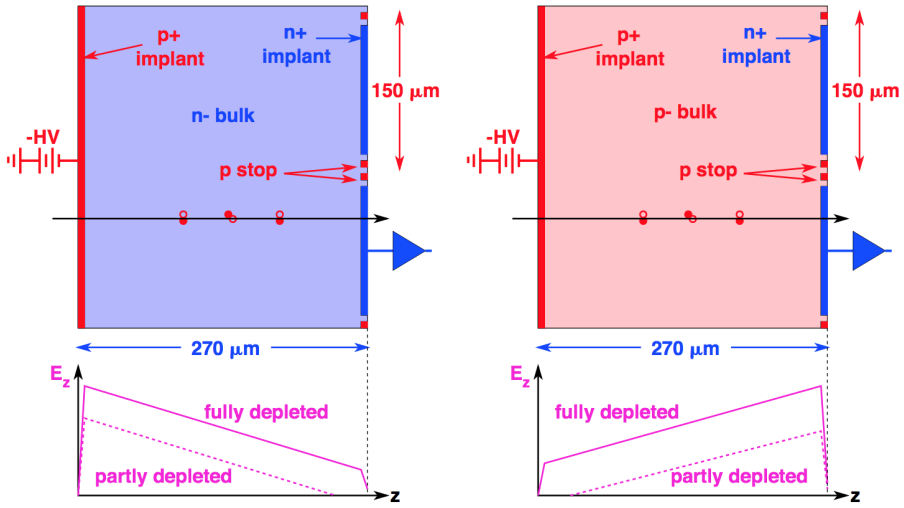

Figure 3.16: A simplified cross sectional view of a sensor array before (left) and after a radiation exposure of more than 10^{13} charged hadrons per cm^2 (right). The radiation exposure produces lattice defects which, on balance, create acceptor states. The doping density of the bulk material changes from n- to p-. This process is called type inversion. The profile of the electric field across the diode is also shown in the bottom panels. Figure reproduced from Reference [78].

In order to improve the simulation of charged-particle interactions and the track reconstruction, it is crucial to model the impact of radiation damage. A general parameterization should model the change of the electric field distribution in the silicon bulk and the signal loss with the accumulated luminosity. CMS makes use of a stand-alone simulation `PixelAV` [78, 79] that is independent from the full CMS simulation and reconstruction framework `CMSSW` [80].

`PixelAV` simulates the passage of a pion (π) through the sensor and incorporates the following elements:

- **Charge deposition:** An accurate model of charge deposition by primary

hadronic tracks uses the "exact" $\pi - e$ elastic cross sections of Bichsel [81], that depend on the electron energy, to determine the π mean free path. This model takes into account the number of electron-hole pairs produced when the scattered electrons or "delta rays" lose energy, assuming that it takes 3.68 eV in energy to produce a pair.

- **Electric field:** A realistic three-dimensional electric field profile resulting from the simultaneous solution of Poisson's Equation, carrier continuity equations, and various charge transport models is generated with the TCAD package [79]. By taking as input the pixel cell geometry description and material properties, TCAD predicts a non-uniform spatial distribution of space-charge density for computing charge propagation inside the sensor bulk.

- **Charge transport:** The electrons and holes produced by the primary hadron drift to the sensor implants under the influence of the internal electric field and the external magnetic field. This drift depends on the electric field (E) and temperature.

- **Charge trapping:** When charge carriers are trapped they are captured for periods of time that are long as compared with the integrating time of the pre-amplifiers and are not detected with full efficiency. This trapping time is incorporated in the simulation by halting the propagation of that charge carrier according to the effective trapping times measured in [82]. The trapping constants used for electrons and holes in CMS simulation are tuned to measurements from data of the charge collection vs depth.

For each event, the simulation outputs the coordinates of the pion entry and direction, the generated number of electron-hole pairs, and a set of collected electrons and the induced signals from trapped charge. The final step, which is performed by a separate analysis code, includes a simplified simulation of the analog response of the ROC after adding noise to the signal.

The results of the PixelAV simulation are mainly used to predict cluster shapes. These cluster shapes are produced for different incident tracks with angles α and β with respect to the local-x and y-axes (as shown in Fig. 3.17). In particular, these shapes can be projected into the $x-$ and $y-$local coordinates and these projection shapes are called "1D-templates" [65]. The 1D-templates take into account the sensor geometry and are produced under certain conditions of radiation fluence, temperature, bias voltage and magnetic field. They are produced about every few fb^{-1}just after installation, when type inversion occurs, and later about every 10 fb^{-1}following periods of annealing or changes in the detector parameters after calibration. For example, for the Run-2 2017 run period this amounted to a database of 12 1D-templates.

The a-priori cluster shape information stored in the PixelAV templates is used to improve the resolution of the hit reconstruction, as detailed below. However, they can be also used to re-weight the digitized cluster charge profile generated by the standard CMS simulation so that they agree with the observed clusters after irradiation [65].

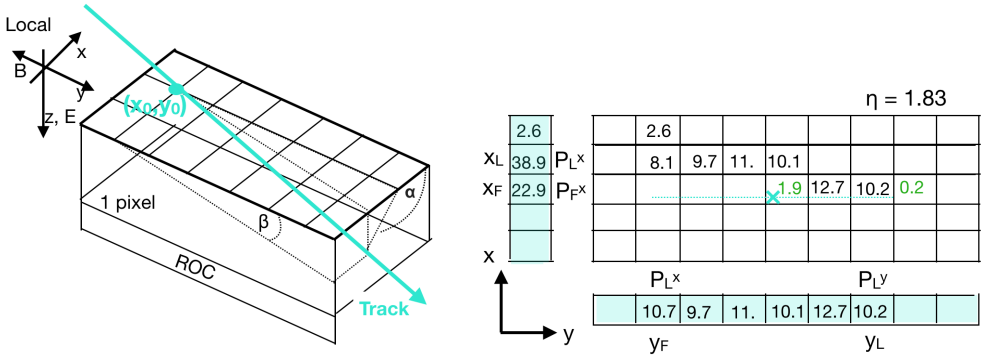

Figure 3.17: (a) Deposition of charge into a barrel module by a track having angles α and β with respect to the local $x-$ and $y-$axes of a barrel module. (b) A pixel cluster example at $\eta = 1.83$ for a barrel hit. The charge deposition in each pixel is shown in thousands of electrons. Numbers shown in green are below the readout threshold and not included in the cluster. The dotted teal line indicates the track projection in the module plane and the red cross shows the true hit position. The $x-$ and $y-$projections are also shown as 1D-arrays. The coordinates of the boundaries between the first and second pixels (x_F/y_F) and next-to-last and last pixels (x_L/y_L) and the charges of the first and last pixels $P^{x/y}_{F/L}$ are also shown. Figure reprinted from Reference [78].

3.4.3 Hit position estimates

If there is no radiation damage, all of the information about a cluster's position can be inferred using the pitch size of the pixel, the signals of the head and tail pixels of the cluster, and the track angles. In the local $x-$direction, for example, the reconstructed hit coordinates are given by [83]:

$$x_{\text{rec}} = \frac{x_F + x_L}{2} + \frac{P^x_L - P^x_F}{P^x_L + P^x_F} \times \frac{W^x_{\text{eff}}(\cot\alpha)}{2} - \frac{\Delta_x}{2}. \tag{3.5}$$

Here, $x_F(x_L)$ are the coordinates of the boundaries between the first and second (next-to-last and last) pixels, W^x_{eff} are the total charge widths in the end pixels which depends on the track angles, $P^{x/y}_{F/L}$ are the signals in the first and last pixels of the $x-$projection and Δ_x is the maximum Lorentz-drift. A

similar expression, that depends on cot β, is applied in the y-direction. This technique, called the "standard reconstruction", is valid for a first-pass of the hit reconstruction algorithm but is prone to biases and worse resolutions after radiation exposure.

Instead, the hit position is inferred by fitting the charge distribution of the pixels in the cluster to the pre-determined templates from `PixelAV` [78, 84]. The so-called "template technique" implicitly incorporates all of the relevant detector physics, such as Lorentz drift and charge loss and trapping, into the templates. For a given track with incident angles α and β on a pixel module, it compares the cluster projections produced by the track to the pre-determined cluster 1D-templates. The hit position is given by the x and y coordinates which minimize the χ^2 comparison. The pixel hit resolution for irradiated sensors improves with the template technique when compared to the standard technique [84]. Since this technique requires knowledge of the track angles it is only used in the second-pass of the hit reconstruction algorithm.

Figure 3.18 shows the resolution for pixel hits in layer 2 (disk 2) of the barrel (forward) inner detector. The residuals are extracted from a triplet method where the measured hit position in a layer is compared with the extrapolated position from neighboring layers [85]. A helix based on the hits on the neighboring layers and the momentum of the track is extrapolated to that layer. The distance between the track extrapolation and the measured hit position is called the residual and is given by $\Delta x = x_{\text{track}} - x_{\text{meas}}$. The residuals shown in Fig. 3.18 show a resolution of the order of 10μm in $r - \phi$ and 20-30μm in z.

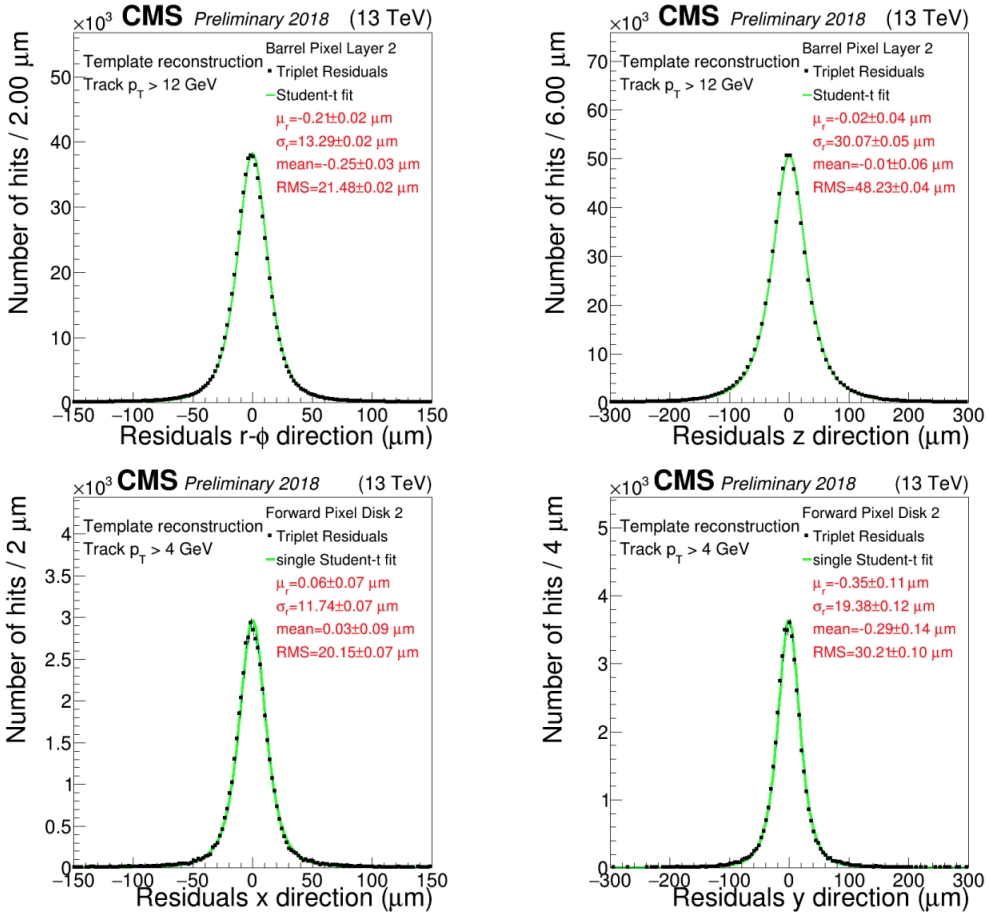

Figure 3.18: Hit residuals for layer 2 and disk 2 of the inner tracker. The residuals are computed using the triplet method: tracks with pixel hits in layers 1, 2 and 3 are re-fitted excluding the hit on layer 2, and the distribution of residuals between the measured and interpolated hit position from the re-fitted track is obtained. The standard deviation obtained from fitting the residual distribution provides an estimate of the hit resolution. Here, "template reconstruction" refers to the method to reconstruct hit positions taking into account a detailed cluster shape simulation predicted by `PixelAV`. Figures extracted from [85].

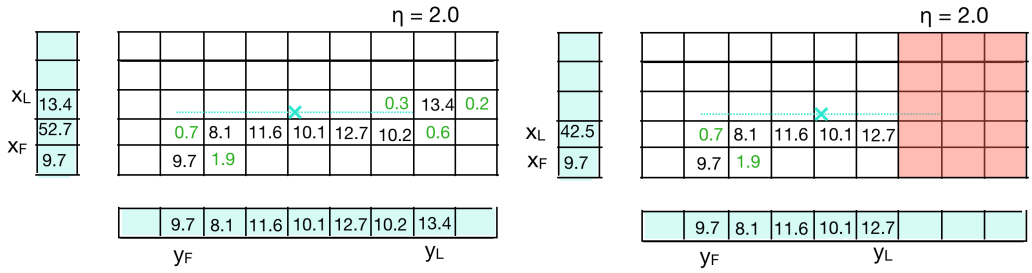

Figure 3.19: Illustrations of a pixel cluster example without loss of hits (left) and truncated (right). The $x-$ and $y-$ projections that contain the information of the charge on the head and tail pixels loose information about the cluster shape when the cluster is truncated.

Truncated clusters Both the standard and template techniques rely on the complete measurement of the pixel clusters. However, the cluster shape can be mismeasured due to presence of dead pixels or bad detector elements. This results in a biased measurement of the hit position, as the $x-$ and $y-$projections get truncated and mis-measure the edge charge. This is illustrated in Figure 3.19. The same bias effect is seen in clusters containing physical-edge pixels where the side of the "truncation" of the cluster is known.

In order to improve the resolution and reduce the bias on the hit position measurement, a 2D-template approach can be used. This approach attempts to use the full 2D information of the deposited charges in the cluster, instead of only the $x-$ and $y-$projections. Since it is time and memory consuming, this algorithm should only be applied in the last step of track reconstruction and to truncated clusters. In order for it to be applied it requires to know whether a cluster is on the physical-edge of if it is a poorly reconstructed cluster. For the latter, one can use the information from the templates to apply requirements on the expected length of the cluster and on the ratio of the measured cluster charge over the expected total charge. If a cluster is shorter or contains less

charge than expected, one can try to use the 2D information to infer the hit position.

The 2D-template algorithm also uses the pre-determined cluster shapes from `PixelAV` to perform a χ^2 fit, given a track with angles α and β. However, since the cluster is only partially reconstructed, one needs to take a guess at which physical-edge an odd cluster is likely truncated, or, for even clusters, one can rely on the side that provides the best χ^2 fit. This guessing is of course not needed for edge clusters, for which the physical-edge of truncation is known.

We tested the 2D-algorithm in a sample of events with edge pixel clusters, for both data and MC, as shown in Fig. 3.20. It was found that improves the resolution of edge-y hits by 20% for layers 2-4 of the barrel. The algorithm was implemented in the reconstruction chain for the re-processing champaign of 2017 data and MC. However, since it is difficult to find a sample of events in data and MC that fully validates the resolution of broken or damaged clusters, the algorithm now runs only on edge clusters and not broken clusters. Other developments that include the simulation of bad detector components will be able to provide a sample of events to test these conditions in the future.

3.4.4 Tracking

Tracks in the inner tracker are found using an iterative *inside-out* approach, with a combinatorial track finder based on Kalman filtering [86]. The process starts with at least three hits, which can be pixel hits or double-strip hits (two 3D-hits), and an estimate of the origin of the trajectory i.e. of the collision

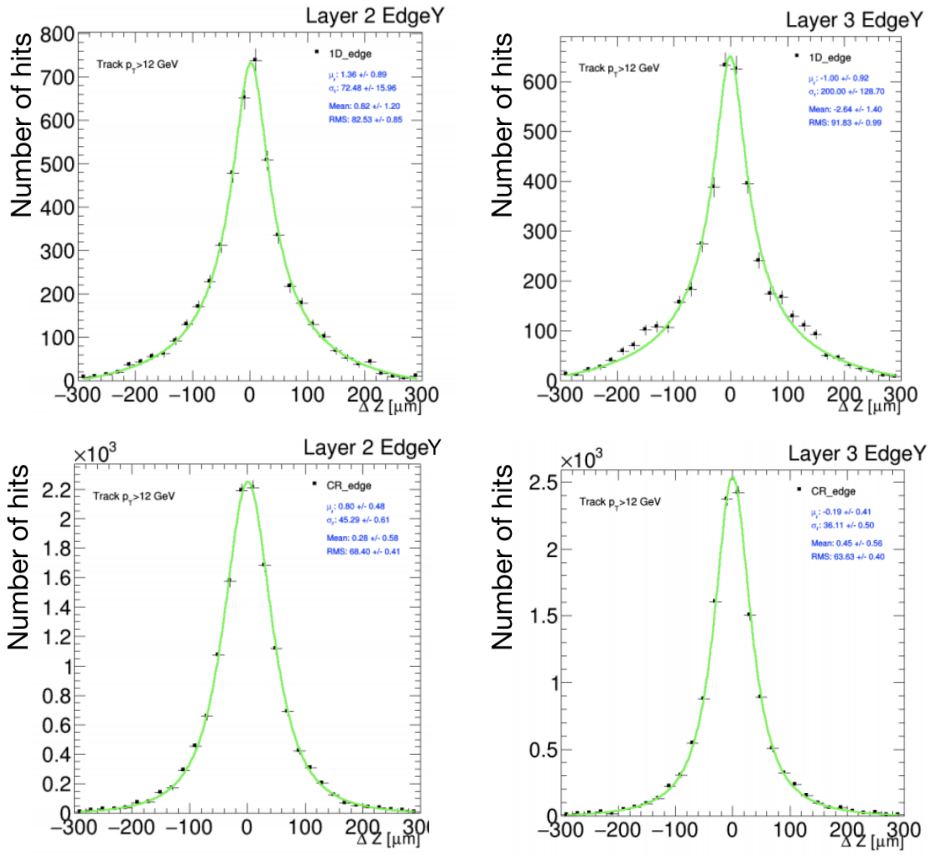

Figure 3.20: Hit residuals for physical-edge-y clusters in the layers 2 and 3 of the barrel inner tracker. The top panel shows residuals using the 1D-template reconstruction, which uses $x-$ and $y-$projections of pre-determined cluster shapes to determine the hit position. The bottom panel shows residuals that use the full 2D information to fit the pre-determined cluster shapes to predict the hit position. These residuals in the $z-$direction were computed using a track re-fitting method. The method drops hits from a given layer, re-fits the track, and computes the residual by comparing the hit position from the re-fitted track with respect to the measured hit position. The resolution is improved by at least 20% in both layers, although edge clusters only represent in average 3% of the total number of clusters in an event.

point. These are used to define seeds that are needed to later identify an initial trajectory to extrapolate. A Kalman filter algorithm [86, 87] evolves the track seeds through the rest of the tracker, and the parameters of each trajectory are estimated by an iterative Kalman filter and smoother fit. Charged particles follow helical paths inside the magnetic field so its trajectories are described by five parameters. These are: the curvature, the azimuthal angle (ϕ_0), $\lambda = \cot\theta$, the impact parameter or minimum r of the track (d_0), and the minimum $|z|$ of track (z_0). Finally, track quality flags determined based on various selection criteria are used to keep or reject found tracks. The average track-reconstruction efficiency for promptly-produced charged particles with $p_{\mathrm{T}} > 0.9$ GeV is 94% for $|\eta| < 0.9$ and 85% for $0.9 < |\eta| < 2.5$ [67].

3.4.5 Vertexing

The precise identification of primary event vertices is needed to assign tracks to collisions and determine the event kinematics. Secondary vertices are instead needed to identify long-lived particles like heavy flavor hadrons and τ leptons that will decay inside the detector.

The primary vertex (PV) reconstruction makes use of the available tracks that are consistent with being produced promptly in the primary interaction region. These tracks are clustered taking into account their z-coordinates at their point of closest approach to the center of the beam spot. That is, the clustering in z is done on the basis that tracks in a cluster are most likely to arise from a single PV [67]. An adaptive fit algorithm [88] is used to determine the vertex for each cluster, where the parameters of the fit are the three spatial

coordinates of the vertex. The PV resolution depends strongly on the number of tracks used to fit the vertex and the p_{T} of those tracks. For a jet-enriched sample, the resolution is around 10μm in x and 12μm in z for primary vertices that use at least 50 tracks [67].

Secondary vertices (SV) are reconstructed with the inclusive vertex fitter (IVF) [89] using as input all the tracks in the event with $p_{\mathrm{T}} > 0.8$ GeV and $d_0 < 0.3cm$. After a first fit to the complete set of tracks the ones compatible with the vertex candidate are removed. Then, a track is discarded from the secondary vertex if it is more compatible with the primary vertex, e.g. based on the ΔR between the track and the SV flight direction. Finally, the secondary vertices can be re-fitted with the adaptive fit algorithm.

The precise determination of SVs is crucial for heavy flavor tagging such as b-jet tagging. The SV reconstruction efficiency for jets is defined as the number of jets containing a reconstructed secondary vertex divided by the total number of jets. For jets with $p_{\mathrm{T}} > 20$ GeV in events with pair production of top quarks (which almost always decay to b quarks), the efficiency for reconstructing a secondary vertex for b (udsg) jets is about 75% (12%) [89].

3.5 Alignment and Calibration

The detector alignment is a key ingredient in translating the local information recorded by the detectors to the global description of the event. CMS has a powerful framework for alignment and calibration, which is based on dedicated "skims", or subsets of data samples [90, 91]. Depending on the needs of the specific workflow, these samples can be selected offline, at reconstruction level, or directly online, at the HLT event. Some important offline alignment workflows are: the energy calibration of the ECAL response (single channel and overall energy scale calibration), the measurement and correction of the tracker orientation with respect to the magnetic field, and the tracker module alignment. Beyond the data workflows, the stability of the position of the large structures over time can be controlled with relatively small amounts of data or via a system of infrared lasers [91].

As an example, we describe the alignment of tracker modules, which is able to provide a complete set of parameters describing its location and geometrical properties [92]. The alignment step is crucial for track reconstruction because if the assumed positions of the silicon modules differ from the true positions, as seen in Fig. 3.21, the track-hit residual distributions will be broadened. Standard alignment algorithms perform a χ^2 minimization fit of the measured position of the hit and the expected position; the latter has as free parameters the position and rotation of the tracker modules. The tracker alignment task is to perform a unique fit to $\sim$ 24000 sensors, which results in more than 200000 free parameters. The target precision is $\sim 10\mu$m, which can be achieved with less than 1 fb^{-1}of data. In order to validate the aligned geometry, muonic

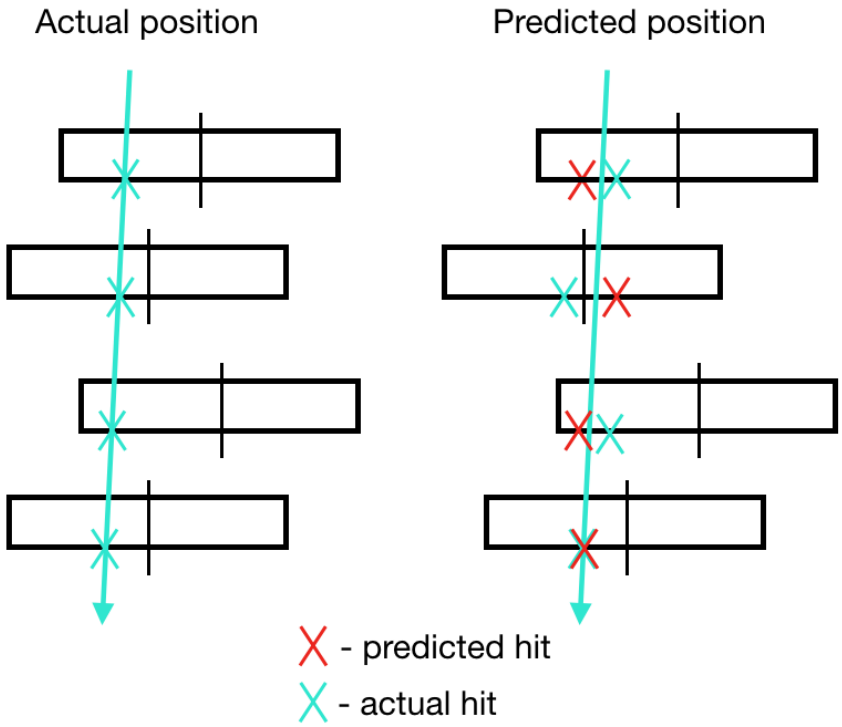

Figure 3.21: A simplified illustration of tracker alignment. The left panel shows the actual position of the detector during data taking, with a track traversing the detector and leaving 4 hits in the modules. The right panel shows the assumed position of the detector, with the second module from the top assumed to be to the left of its actual position. In this case, the position of the predicted hits shown in red is biased. Figure taken from [92].

decays of Z bosons can be measured both in data and simulation. The invariant mass distributions from aligned data are expected to align to the known value of the mass of the Z boson. [91]

3.6 Reconstruction of particles in an event

3.6.1 Particle flow algorithm

Once an event passes the HLT selection, signals from the different subsystems are recorded. These signals can be particles hits in the tracker or muon systems, or deposited energies in the calorimeters. Since a collision event will contain a high particle multiplicity, a dedicated algorithm is used to build the global description of an event. This event algorithm, the CMS particle-flow (PF) algorithm [93], reconstructs and identifies individual particles with an optimized combination of information from the various CMS sub-detectors.

An individual PF particle candidate can be identified as either: an electron, a photon, a charged or neutral hadron or a muon. Each of these particles yields a specific signature in the detector: electrons and photons leave a cluster in the ECAL but not the HCAL, hadrons may initiate a shower in the ECAL but they are fully absorbed in the HCAL, while muons traverse the calorimeters with little or no interactions and produce hits in the muon detectors. Furthermore, the presence of the magnetic field bends the trajectories of all charged particles and allows the electric charges and momenta of these particles to be measured. This simplified view is summarized in Fig. 3.22, which displays a sketch of a transverse slice of the CMS detector.

The PF algorithm links multiple detector signals together into a single PF candidate. It associates inner charged tracks from the silicon tracker to calorimeter clusters and muon tracks, based on their proximity in the $r - \phi$ plane. It proceeds by "subtracting" objects from the event in order of

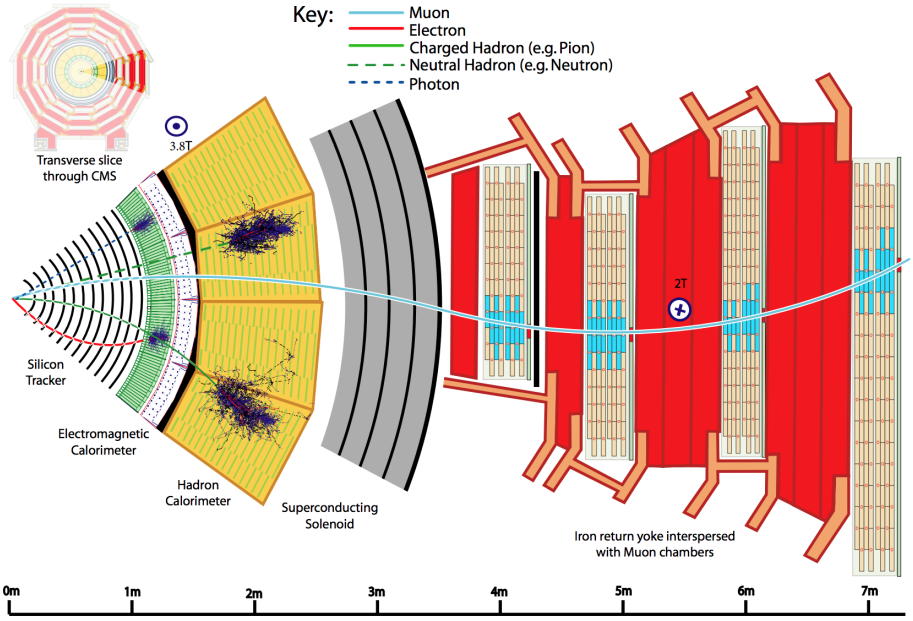

Figure 3.22: A sketch of the specific particle interactions in a transverse slice of the CMS detector, from the beam interaction region to the muon detector. The muon and the charged pion are positively charged, and the electron is negatively charged. Figure reproduced from [93].

decreasing reconstruction accuracy. The subtraction starts from muons and is followed by electrons, isolated photons and charged hadrons, such that neutral hadrons and non-isolated photons are built from calorimeter clusters that are not compatible with any tracks. In this section, we describe the reconstruction of each of the PF particles following the subtraction sequence present in the PF algorithm. The PF elements that make up each these identified particles are masked against further processing in the corresponding PF block, i.e. are not used as building elements for other particles.

3.6.2 Muons

Muon reconstruction links inner tracks and muon spectrometer hits. The latter allows muons to be identified with high efficiency over the full detector acceptance. Three different muon types can be reconstructed:

- *Stand-alone muons*: they rely solely on muon chamber hits, which are clustered to form muon-track segments.
- *Global muons*: they match stand-alone muons with tracks in the inner tracker. The hits from the inner track and from the stand-alone-muon track are combined and fit to form a global-muon track.
- *Tracker muons*: they are reconstructed inside-out and rely on the extrapolation of an inner track to the muon system.

Given the high efficiency of the inner track and muon segment reconstruction, about 99% of the muons produced within the geometrical acceptance of the

muon system are reconstructed either as a global muon or a tracker muon and very often as both [93].

The identification of muons proceeds by a set of selections based on the global and tracker muon properties. In this thesis, we use a *tight-selection* to identify muons from Z and W boson decays. This selection requires a global-muon track with a $\chi^2/N_{\text{dof}} < 10$, to ensure a good track fit, and at least one hit in the muon detectors. In addition, the candidate should be a tracker muon with at least two matched muon segments in different muon stations and an inner track reconstructed from at least five inner-tracking layers, including one pixel detector layer [93].

To adequately reject hadrons that would be misidentified as muons an isolation criteria is applied. The muon isolation is computed as the sum of the transverse energy of the particles inside a cone of radius $\Delta R = 0.3$ around the muon direction divided by the muon transverse momentum. The isolation is required to be lower than 0.1, i.e. the sum of the p_{T} tracks and calorimeter energy deposits should not exceed 10% of the muon p_{T}. Muons inside jets, for example those from semileptonic heavy-flavor decays, would be non-isolated. To reject charged hadrons misidentified as non-isolated muons e.g. because of punch-through, it is required either that at least three matching track segments be found in the muon detectors, or that the calorimeter deposits associated with the track be compatible with the muon hypothesis.

An event sample containing decays of $Z \rightarrow \mu^+\mu^-$ is used to evaluate the prompt muon identification/isolation efficiency. From this data sample, corrections (known as scale factors) are derived: $SF = \frac{\epsilon_{\text{Data}}}{\epsilon_{\text{MC}}}$. Scale factors

account for differences in the efficiency ϵ in data and simulation, where the efficiency is defined from the fraction of correctly identified/isolated muons. To find the efficiency, the *tag-and-probe* method is used. This method applies strict selection criteria are applied on one of the two decay muons, called *tag*, and the second muon, the *probe*, is used for the efficiency measurements. Both muons are required to form a system with invariant mass consistent with a di-muon resonance. The number of $Z \to \mu\mu$ events in each category is measured by fitting the invariant mass distribution. The efficiency is defined as the fraction of probe electrons satisfying the tested criteria: $\epsilon = \frac{N_{\text{pass}}}{N_{\text{pass+fail}}}$. The muon reconstruction and identification efficiency is greater than 96% and measured as a function of the muon p_{T} and η as described in [72].

3.6.3 Electrons and isolated photons

Both the electron and photon reconstruction are based on energy deposits on the ECAL. These can be spread laterally over several crystals and are thus clustered together. To suppress noise in the calorimeters, the PF algorithm only considers cells with energies above a given threshold, this procedure is referred to as "zero suppression". The energy of photons is obtained directly from the ECAL measurement, corrected for zero-suppression effects. Electron reconstruction combines the corresponding ECAL cluster with an associated track. Due to the large amount of material in the tracker, electrons will often lose a significant amount of energy through bremsstrahlung photons before reaching the ECAL. To gather all the radiated energy, the energy of the electron and of possible bremsstrahlung photons in a window in η and ϕ is grouped

into a *super-cluster*. For electron tracks the Gaussian Sum Filter (GSF) [94] is used to fit the track, instead of the Kalman Filter (KF). This method allows for sudden and substantial energy losses along the trajectory.

The ratio between the HCAL and ECAL energy deposits is used as a discrimination variable to ensure optimal energy containment for electron and photon showers. For ECAL-based electron and photon candidates, the sum of the energies measured in the HCAL cells with a distance $\Delta R < 0.15$ to the super-cluster must not exceed 10% of the super-cluster energy [72]. Photon candidates are retained if they are isolated from other tracks and calorimeter clusters in the event. Electron candidates must satisfy extra identification (ID) criteria in order to reject hadron or photon backgrounds. These may arise from neutral hadrons or photons whose energy is overlapped with a charged hadron, or from electrons that are originated from a photon conversion. The electron ID criteria includes requirements on the amount of energy radiated off the GSF track, the GSF χ^2 and number of hits.

Similarly to the muons, scale factors that correct for differences between data and MC efficiencies for the electron or photon IDs are derived using a data sample with $Z \rightarrow e^+e^-$ decays [95]. In this thesis, we only identify electrons and photons for veto-purposes in the event selection.

3.6.4 Hadrons and non-isolated photons

Once muons, and isolated electrons and photons are identified, the remaining particles are used to identify hadrons. These may appear as charged hadrons (e.g. $\pi^{\pm}$, $\mathrm{K}^{\pm}$ or protons), neutral hadrons (e.g. K^0 or neutrons), non-isolated

photons (e.g. from π^0 decays), or more rarely as muons from heavy hadron decays.

Photons and neutral hadrons are associated to track-less ECAL and HCAL clusters respectively, i.e. these clusters are not linked to any tracks within the tracker acceptance ($|\eta| < 2.5$). The precedence of assigning ECAL clusters to photons is due to the observation that, in hadronic jets, neutral hadrons only leave 3% of the jet energy in the ECAL. Beyond the tracker acceptance this statement is no longer justified, so linked ECAL and HCAL clusters are assigned to neutral hadrons and ECAL clusters without such a link are assigned to photons. Charged hadrons are associated to the remaining HCAL clusters that can be linked to tracks and these may in turn be linked to the remaining ECAL clusters. The energy of charged hadrons is determined from a combination of their momentum measured in the tracker and the matching ECAL and HCAL energy deposits, corrected for zero-suppression effects and for the response function of the calorimeters to hadronic showers. This response function is initially obtained from calibration measurements in test beam data but ultimately refined by simulating a large sample of single hadrons in the detector.

The experimental signature of hadrons are particle-jets, as explained in Sec. 2.2.3. The methods for clustering, calibration and identification of these hadronic jets will be described in the next chapter, as this is one of the crucial steps in the search for hadronic decaying resonances.

3.6.5 Missing momentum

The PF missing transverse momentum vector is defined as the negative vectorial sum of the transverse momenta of all the PF particles identified in the event:

$$\vec{E_T^{miss}} = -\sum_p \vec{p_T} \tag{3.6}$$

and its magnitude is referred to as p_T^{miss}. Because of conservation of momentum, in a perfectly reconstructed event, non-zero p_T^{miss} implies the presence particles that escape the detector, such as neutrinos or DM candidates. However, the p_T^{miss} reconstruction is sensitive to detector malfunctions and various reconstruction effects. These can result in mismeasurements of the particle momentum and misidentifications of particles which in turn result in not genuine p_T^{miss}. The performance of p_T^{miss} reconstruction can be assessed for example using events where an identified Z boson decay to leptons is present. While there is no genuine p_T^{miss} in these events, it can be induced by removing the vector boson from the event reconstruction. Since the energy and momentum of the Z boson is well understood, the scale and resolution of p_T^{miss} can be measured as a function of the energy scale of the event [96].

Chapter 4

Reconstruction and Identification of hadronic particles

When studying high-energy proton collisions one often has to consider processes where quarks are gluons are present. These can be originated either from the decay of heavy resonances, or other decay chains emitting partons or, finally, from a high energy QCD parton radiated in the initial or final state. However, these partons are not directly observed in the final state. The successive splittings associated with parton shower and hadronization transform them into a cluster of collimated particles. These particles are recombined into a *hadronic jet*, with the aim of reconstructing the original parton.

Jets can be seen as proxies to the high-energy quarks and gluons produced in a collision. However, this picture is over-simplified since there is no way to tell whether two particles are part of the same jet or belong to the same jet. Furthermore, due to higher order corrections in QCD the definition of partons may include real or virtual contributions. Thus, a robust and unambiguous definition of a jet is needed, i.e. a well-defined procedure that tells how to cluster and reconstruct the jets from the set of hadrons in the final state of the

collision.

In this chapter, we describe the algorithms for clustering, reconstruction and calibration of these jets in the CMS detector. We also describe the pileup removal algorithm that is designed to identify and subtract contributions from a different interaction vertex, by eliminating uncorrelated radiation from jets. We then describe methods used for heavy flavor jet identification and, finally, we venture into a description of various aspects of jet substructure. Jet substructure consists of a set of tools that exploit information from the radiation pattern inside the jets. They can be used, for example, to identify the hadronic decays of highly boosted particles which can be merged and reconstructed into a single large jet. In preparation for Chapter 5, we will emphasize in the tools for the identification of an energetic low mass Z' mediator decaying into a quark-antiquark pair.

4.1 Jet clustering and reconstruction

The precise definition of a jet is dependent on the algorithm used to cluster particles. For a clustering algorithm to be robust, it should be infrared and collinear (IRC) safe, where "infrared" refers to the low-energy limit. That is, the resulting jets should be insensitive to arbitrarily low energy particles and small-angle splittings. If an algorithm is collinear unsafe, the splitting of a higher energy (hard) particle will result in the altering of the number and contents of the jets. A similar problem arises for an infrared unsafe algorithm when a lower energy (soft) gluon is added to the system. IRC safety is a useful theoretical requirement for making calculations in perturbative-QCD.

Cone-type algorithms, that sum the momenta of all particles within a cone with fixed size, are not IRC safe. Instead, LHC experiments use *sequential recombination* algorithms to which the *generalized-k_{T}* family belongs [97–99]. These assume that particles inside jets will have small differences in transverse momenta, and therefore group particles based on momentum space. The result are jets that have fluctuating areas in (η, ϕ) space.

The family of *generalized-k_{T}* algorithms is built under the following criteria. Given a set of particle candidates in an event, we can define an inter-particle distance measure d_{ij} and a beam distance d_{iB}:

$$d_{ij} = \min\{p_{\mathrm{T},i}^{2k}, p_{\mathrm{T},j}^{2k}\}\frac{\Delta R^2}{R}, \tag{4.1}$$

$$d_{iB} = p_{\mathrm{T},i}^{2k}, \tag{4.2}$$

where $\Delta R^2 = \Delta\phi^2 + \Delta y^2$, k and R are tunable parameters, and B refers to the beam axis. Iteratively, we can find the smallest distance among all d_{ij} and d_{iB}:

- If the smallest distance is a d_{iB}, then the particle i is removed and added to the set of candidate jets.
- If the smallest is a d_{ij}, then the particles i and j are removed from the list and recombined into a new object k.

This process is repeated until all particles are exhausted. When the inter-particle distances are such that $\Delta R_{ij} > R$, the beam distance becomes smaller than the inter-particle distance and objects are no longer recombined, making R a typical measure of the size of the jet, called the *jet radius*.

The jet size determines the susceptibility of a jet to soft radiation. A larger

jet radius is important as it allows the jet to capture enough of the hadronized particles for the accurate calculation of the mass and energy of the jet. However, a smaller jet radius is useful in reducing the amount of the underlying event and pileup captured by the jet, preventing the overestimation of the jets mass and energy.

In Equations. 4.2, the parameter k controls the dependence of the distance measure d_{ij} with the momentum of the final state particles and thereby the sequence of clustering. The original k_{T}algorithm, uses $k = 1$ and clusters soft and collinear particles first. The Cambridge-Aachen algorithm (CA) [99], with $k = 0$, prioritizes particles in the clustering solely by their angular proximity. Finally, the anti-k_{T}algorithm (AK) [97], with $k = 1$, combines the hardest particles first. The latter also results in nearly perfect conical jets, as seen in Fig. 4.1.

From Equations. 4.2, it can also be seen that, for any value of k, collinear splittings ($\Delta R \to 0$) lead to $d_{ij} \to 0$. Similarly, additional soft final state particles may be recombined as part of a jet at any stage in the clustering, but will have no impact on the output of the remainder of the clustering history, since their momentum contribution will be small. Thus, Equations 4.2 parameterize a class of IRC-safe jet algorithms.

Jet clustering algorithms can be applied to any set of topological objects. These include simulated particles before and after the parton shower, energy measurements and reconstructed tracks. In CMS, the inputs to jet clustering algorithms are PF candidates and the algorithms are implemented using the Fast-Jet library [100]. In this thesis, we use three types of jet collections: jets

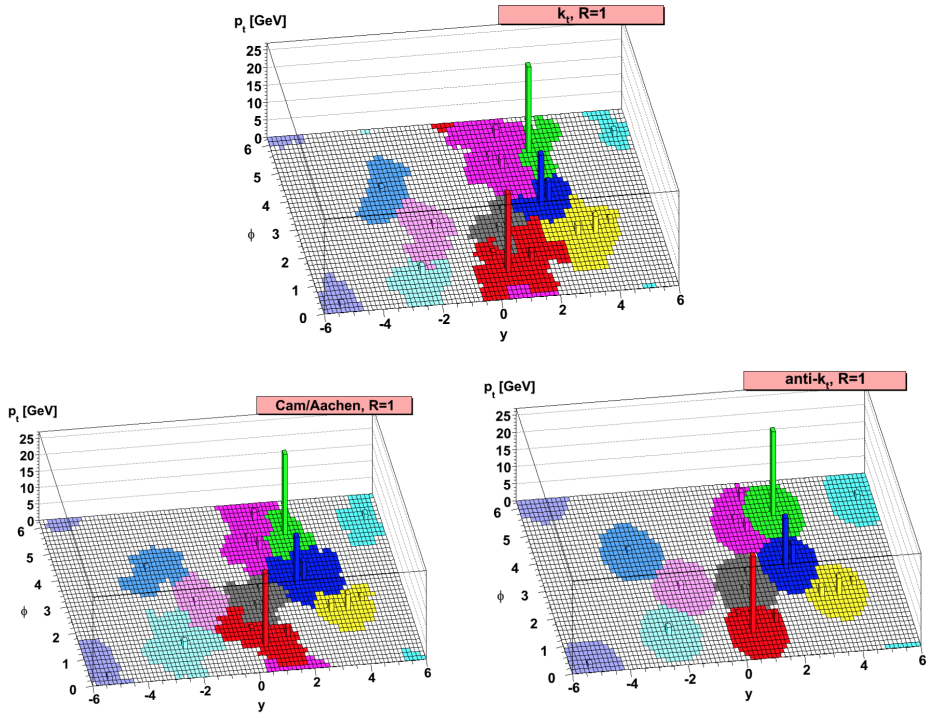

Figure 4.1: Examples of the effect of different sequential algorithms applied to the same event: the k_{T}algorithm (top), the Cambridge-Aachen (bottom-left) and the anti-k_{T}algorithm (bottom-right). The colored regions correspond to the area in (η, ϕ) of the different reconstructed jets. Figures reprinted from [97].

clustered with the anti-$k_T$algorithm with distance parameters of $R = 0.4, 0.8$, and referred to AK4 and AK8 jets, respectively; and jets clustered with the Cambridge-Aachen algorithm with $R = 1.5$, referred to as CA15 jets.

4.1.1 Jet calibration

Like all experimentally-reconstructed objects, jets need to be calibrated in order to have the correct energy scale that matches the detector response: this is the aim of the jet energy corrections (JEC) [101]. The jet energy corrections are calculated using MC simulation, and are then adjusted for data using a combination of several channels and data-driven methods. This calibration is performed in a series of steps

1. A first correction is derived to account for the additional contributions to the jet energy and momentum due to *pileup*.

   It is parameterized as a function of the jet p_T and η and two pileup observables: the event wide pileup density ρ and the jet area $A \sim \pi R^2$. For the last two calculations, a uniform distribution of non-physical particles (ghosts) with infinitesimal momenta is added. The active area of a jet is defined as the number of ghosts clustered in the jet. Physically, this definition mimics the effect of pileup in the sense that pileup particles are uniformly distributed, in the limit where their p_T becomes infinitely small. Similarly, the event quantity ρ can be computed with ghosts and is a measure of the median jet p_T per area.

   The correction is derived from simulation with and without pileup overlay. Residual differences between data and detector simulation are

also derived using data events collected with a random, or zero-bias, trigger. As these events are not triggered by any specific energy deposit, their main sources of energy deposits are detector noise and pileup.

2. After jets have been corrected for the pileup offset, the simulated response corrections are applied. These account for biases in the detector response by using the GEANT4 simulation of the detector.

   The corrections are derived from a simulated sample of multijet events produced by quantum chromodynamics, also called QCD multijet events throughout the text. In this sample of events, two jet collections are computed: PF jets, that use reconstructed PF candidates, and "truth" jets, that use the particle collection produced after the hard scattering and parton shower simulation. The energy scale of the reconstructed PF jets is corrected to match that of truth jets, as a function of the jet p_{T} and η.

3. The last two corrections are applied to both data and simulation. A third $p_{\mathrm{T}}-$dependent scale factor is applied only to data to correct for residual differences between the real detector and the CMS detector simulation and event reconstruction.

   These corrections are determined using X+jet events: $Z(\rightarrow \ell\ell)$+jet, γ+jet, and dijet events. The basic idea, in all the considered topologies, is to exploit the transverse momentum balance, at hard-scattering level, between the jet to be calibrated and a reference well-measured object X (the Z boson, the photon or the jet).

The final uncertainties on the jet energy scale are below 3% across the

phase space considered by most analyses ($p_{\mathrm{T}} > 30$ GeV and $|\eta| < 5.0$), owing to the excellent performance of the particle flow reconstruction.

Finally, these corrections can improve the estimation of $p_{\mathrm{T}}^{\mathrm{miss}}$, by propagating them in the following way:

$$\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}} \mapsto \vec{p}_{\mathrm{T}}^{\,\mathrm{miss}} + \sum_{j \in \text{ jets}} \left(\vec{p}_{\mathrm{T}}^{\,j,\mathrm{corr.}} - \vec{p}_{\mathrm{T}}^{\,j,\mathrm{raw}} \right), \tag{4.3}$$

where $\vec{p}_{\mathrm{T}}^{\,j,\mathrm{raw}}$ refers to the uncorrected $p_{\mathrm{T}}^{\mathrm{miss}}$as defined in Eq. 3.6, and the sum is over jets with $p_{\mathrm{T}} > 15$ GeV.

4.1.2 Heavy flavor jet ID

A crucial part of the LHC physics program is the identification of the parton that gives rise to a jet. In particular, jets from bottom quarks or charm quarks (heavy-flavor jets) are usually present in the decay of heavy SM particles such as top quarks ($t \to Wb$) and Higgs boson decays ($h \to b\bar{b}$). In the context of this thesis, the identification of $b-$jets is useful to select events that contain a top quark; these events are used for the calibration of jet substructure tools as it will be detailed below.

Since we cannot directly access the quantum numbers of the heavy-flavor parton that induces a jet, we instead measure a number of jet-related observables that are connected to the hadronization properties of a b or c quark. For instance, the hadronization of the bottom quark involves the production of b hadron, that has a relatively large lifetime of 10^{-12} seconds and a large Lorentz factor. This results in the production of a jet with a secondary vertex that is displaced several millimeters ~ 5 mm with respect to the interaction

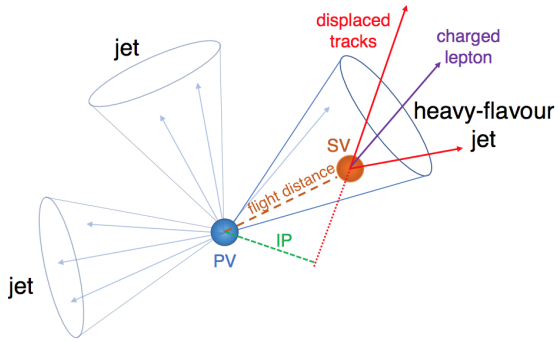

Figure 4.2: An illustration of the decay of a b quark, along with the definition of the secondary vertex (SV) and the impact parameter (IP). The presence of a SV is a distinct signature of B hadron decays, which have a long lifetime of about 1.5 ps. Algorithms used for $b-$ tagging use various input parameters related to the SV or charged particle tracks that originate from the B hadron decay. Figure reprinted from [89].

point. This topology is illustrated in Fig. 4.2.

The idea of b/c-tagging is to use a combination of discriminating variables to build an discriminator that distinguishes b/c jets from $u/d/s/g$ jets. Those jets that are associated to discriminator values above certain thresholds are taken to be tagged as a certain flavor. The discriminating variables are related to the existence and properties of tracks from charged particles and vertices associated to primary or the secondary vertices. Examples of these variables are the impact parameter of the SV, presence of soft leptons (to account for semi-leptonic decays of the b hadron $b \to \mu^- \nu_\mu c$), the average track multiplicity or the invariant mass of the tracks. CMS uses the Combined Secondary Vertex (CSVv2) tagger [89], which combines nineteen characteristics of the jet with a multivariate technique. In this thesis, we consider a jet as $b-$tagged if the discriminator value of the CSVv2 > 0.54. The efficiency of this threshold,

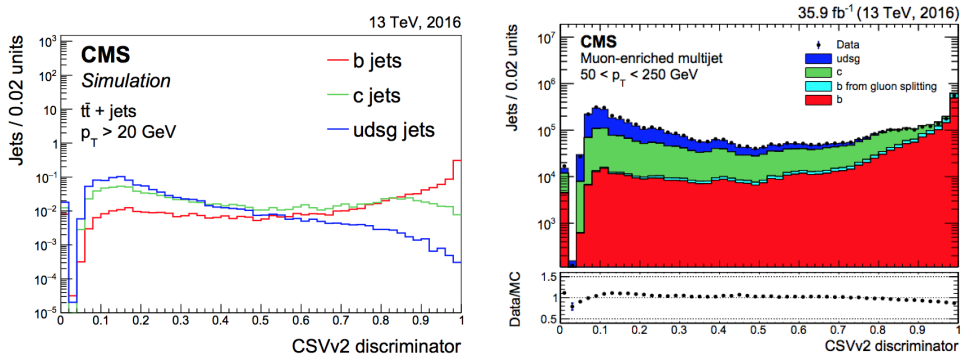

Figure 4.3: Distribution of the CSVv2 (left) discriminator values for jets of different flavors in $t\bar{t}$ events. The distribution is normalized to unit area. In the right panel, the CSVv2 discriminator distribution is shown in data for a dijet sample enriched with $g \to b\bar{b}$ events. The simulated contributions of each jet flavor are shown with different colors. The total number of entries in the simulation is normalized to the number of observed entries in data. Figures reprinted from Ref. [89].

or "working point", is around 65% for $b-$jet identification, with a misidentification rate of 1% for light quark and gluon jets. Figure 4.3 compares the CSVv2 response in data and simulation. The scale factors to correct this distribution are derived in $b-$enriched samples, such as $t\bar{t}$ events, where the $b-$quarks from the top quark decays are used, or gluon splitting $g \to b\bar{b}$ events, that are selected by triggering on a non-isolated soft-muon coming from the decay of one of the $b-$hadrons.

4.1.3 Pileup mitigation

To mitigate the impact of particles arising from pileup, a pileup mitigation algorithm is used. In general, these type of algorithms utilize three basic pieces of information to identify pileup: the event-wide pileup density, vertex information from charged tracks, and the local distribution of pileup with respect to particles from the primary or leading vertex (PV or LV). That is, they

can exploit either global information from the event or local topological information from the jet. In this thesis, we use the pileup per-particle identification algorithm (PUPPI) [62, 102], which combines both pieces of information.

The PUPPI algorithm operates at the PF candidate level, before any clustering is performed. It calculates a weight in a range from 0 to 1 for each particle. The weight is applied to the four-momentum of each particle, such that $p_i^\mu \rightarrow w_i \times p_i^\mu$. The ideal weight is 1 for leading vertex particles and 0 for pileup particles. This weighting method allows for experimental information, such as tracking, vertexing and timing information, to be included.

For charged particles, the weight is based on tracking and vertexing information, that is available in the central region $|\eta| < 2.5$. All charged particles used in the fit to the LV are assigned a weight $w_i = 1$ and charged particles associated with other reconstructed collision vertices, referred to as pileup (PU) vertices, are assigned a weight $w_i = 0$. If charged particles are not associated with any vertex, a weight $w_i = 1$ is assigned if d_z, the distance of closest approach to the LV along the z axis, is smaller than 0.3 cm.; a weight of 0 is assigned in other scenarios. [1]

Neutral particles, require the definition of a shape that attempts to distinguish parton shower-like radiation from pile-up-like radiation. This shape, α_i, is defined for every particle i in the event. The shape is calculated from the p_T and the angular distance to nearby particles, and, in CMS, is defined

[1] The threshold of 0.3 cm is in accordance with 15 standard deviations of the measured vertex reconstruction resolution in z direction.

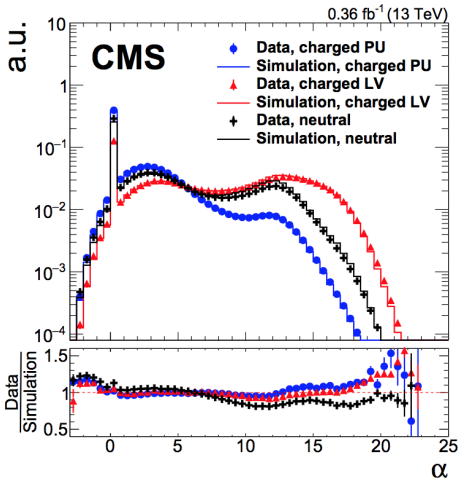

Figure 4.4: The distribution of the PUPPI shape α_i used in CMS for particles i: for charged particles associated with the LV (red triangles), charged particles associated with PU vertices (blue circles), and neutral particles (black crosses) for $|\eta| < 2.5$. The distribution is shown for both data and simulation and normalized to unity. Reprinted from Reference [62].

as [62]:

$$\alpha_i = \log \sum_{j \neq i, \Delta R_{ij} < R_0} \left(\frac{p_{\mathrm{T},j}}{\Delta R_{ij}} \right)^2 \begin{cases} \text{for } |\eta_i| < 2.5, & j \text{ are charged particles from LV.} \\ \text{for } |\eta_i| > 2.5, & j \text{ are all kinds of reconstructed particles.} \end{cases} \tag{4.4}$$

where j are other particles, R_{min}, R_0 are tunable parameters. The distribution is expected to be larger for LV particles than PU particles, as PU radiation is uniformly distributed, whereas LV radiation is centered around hard partons. The distribution of α_i, determined for PF particles in CMS, is shown in Fig. 4.4.

To translate α_i of each particle into a probability, charged particles assigned to PU vertices are used to generate the expected PU distribution in an event. From this expected distribution a median and root-mean-square (RMS) of the α values are computed, and compared to α_i using the following approximation:

$$\text{signed } \chi_i^2 = \frac{(\alpha_i - \bar{\alpha}_{\text{PU}})|\alpha_i - \bar{\alpha}_{\text{PU}}|}{\sigma^2}, \tag{4.5}$$

where $\bar{\alpha}_{\text{PU}}$ and σ are the median and RMS of the charged PU α_i distribution. This expression measures how far α_i fluctuates from the pileup median. If the signed χ_i^2 is large, the particle most likely originates from the LV.

To assign a weight, it is noted that, since the PU distribution of α looks Gaussian-like, χ_i^2 should follow a χ^2 distribution with 1 degree of freedom. Then, we can define the PUPPI weight as:

$$w_i = F_{\chi^2, N_{\text{dof}}=1}(\chi_i^2), \tag{4.6}$$

where F is the cumulative distribution function of the χ^2 distribution. Almost all pile-up particles have values within a few standard deviations of the median and are assigned small weights. Values that deviate far from the charged pile-up are indicative of a hard scatter, and these particles are assigned large weights.

The performance of pileup removal algorithms is evaluated in the context of the reconstruction of jets and missing transverse momentum, and the calculation of jet substructure observables, which will be detailed below. The PUPPI algorithm is extensively used in CMS because it shows the best performance for jet mass and substructure observables, i.e. the least dependence on PU interactions among different pileup removal algorithms [62].

4.2 Jet substructure

The two-body hadronic decay of heavy resonances at rest usually results in an event topology where *two small-radius jets* are recoiling back to back. However, when these particles are produced with high energy and sufficiently large transverse momentum in the laboratory frame, the momenta of their decay daughters increases and their angular separation ΔR becomes smaller. This is illustrated in Fig. 4.5.

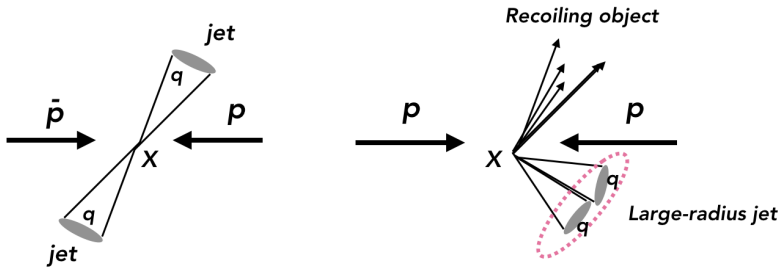

Figure 4.5: Illustration of the hadronic decay of a resonance X in the laboratory frame. Their two-body decay results in two hadronic small-radius jets that are recoiling back to back (left). When the transverse momentum of the resonance X is large, its decays are very collimated and the outgoing quarks are not sufficiently separated relative to each other to be resolved into individual jets (right). The small opening angle between the decay products leads to fully-merged particle decays.

In this boosted regime, the opening angle between the daughter quarks depends on the momentum of the parent particle p_{T} and its mass M:

$$\Delta R \sim \frac{2M}{p_{\mathrm{T}}}. \tag{4.7}$$

[2]

[2]This rule of thumb can be derived from the invariant mass m of the products of the resonance decay. For a two-body quark decay we have:

$$M^2 = (p_1 + p_2)^2 \approx 2E_1E_2(1 - \cos\theta_{12}) \approx E^2 z(1-z)\theta_{12}^2, \text{for} \theta_{12} \ll 1. \tag{4.8}$$

This means, that for $p_{\mathrm{T}} > 2M/R$, the decay products may become sufficiently collimated to be reconstructed into a single jet with radius R. Small-radius jets, such as AK4 jets, will begin to overlap and this may reduce the efficiency of reconstructing each of the decay products. Thus, larger-radius jets must be used. In this thesis, AK8 jets with parameter radius $R = 0.8$ and CA15 jets with parameter radius $R = 1.5$ are used. CA15 jets are used to reconstruct higher mass boosted resonances, e.g. with mass above 300 GeV. This is because a heavier resonance with the same transverse momentum has a lower Lorentz boost and a larger radius jet is required to contain its hadronization products.

Having reconstructed these jets, an important task is then to distinguish these hadronic two-body decays from the dominant non-resonant QCD multi-jet production, i.e. from the hadronization of light quarks and gluons, which may also be reconstructed in this large-R jet topology. There are two main handles for its identification: the jet mass and the jet substructure observables, that exploit differences in the radiation patterns of the jets. In the following, we explore these handles, as well as the tools used for their calibration in experimental data.

Where, z and $(1-z)$ are the energy fractions carried by the two almost massless quarks, $p_{1,2}$ and $E_{1,2}$ are their four-momenta and energies; and, θ_{12} is the angle between the quarks with respect to the direction of motion of the resonance that corresponds roughly to the distance ΔR between the quarks in the $(\eta - \phi)$ plane. The second approximation holds for symmetric decays ($z = 1/2$) in the boosted kinematic regime ($p_{\mathrm{T}} \ll m$). Then we have:

$$\Delta R = \frac{M}{E} \frac{1}{\sqrt{z(1-z)}} \approx \frac{2M}{p_{\mathrm{T}}} \tag{4.9}$$

4.2.1 Jet grooming

The drawback of using large-radius jets are unwanted contributions from the underlying event and pileup. Their net effect are UE/PU hadrons which are spatially overlapped with hadrons from the final state. This complicates the jet finding and worsens the resolution in substructure quantities, such as the jet mass. While pileup mitigation techniques, such as PUPPI, can help to eliminate uncorrelated radiation from the jets by subtracting non-PV contributions, grooming techniques can help to remove unwanted soft-radiation, thereby reducing the contamination effects. A combination of these techniques leads to the best overall performance.

Jet grooming is an additional "post-processing" treatment of large radius jets, where radiation is systematically removed from the jet. Grooming algorithms often target soft and wide-angle radiation, which is not associated with the underlying hard substructure. There are two main algorithms used by CMS: trimming and soft-drop [103, 104]. They both rely on the same procedure: re-clustering the constituents of a jet, e.g. the PF candidates, while rejecting soft/wide-angle radiation.

For trimming, the k_{T}algorithm is used to re-cluster the constituents into subjets (smaller-radius jets), which have a characteristic radius called $R_{\mathrm{sub}} < R$. Those subjets with a momentum fraction $f < f_{\mathrm{cut}}$ are removed, and the remaining subjets are assembled into the trimmed jet. In CMS, trimming is used at the trigger level, as it will be detailed in the next chapter.

For soft-drop, the CA algorithm is used to re-cluster the constituents but the algorithm imposes a condition on the $2 \rightarrow 1$ subjet clustering step, by

going backwards in the sequence in which the particles were combined. At each node of the CA-clustering tree, the transverse momentum fraction of the softer particle with respect to the merged system:

$$z = \frac{\min(p_{\mathrm{T}}^1, p_{\mathrm{T}}^2)}{p_{\mathrm{T}}^1 + p_{\mathrm{T}}^2} \tag{4.10}$$

is used as a criteria to determine the scale of the soft radiation, and the angular distance ΔR_{12} between the two particles is used as a criteria for identifying wide-angle radiation. Here, p_{T}^i refers to the p_{T} of the i-th subjet of the node. The softer branch of the node is removed unless it satisfies the condition:

$$z > z_{cut} \left(\frac{\Delta R_{12}}{R} \right)^{\beta}, \tag{4.11}$$

where z_{cut} and the angular exponent β are tunable parameters. In CMS, soft-drop is used at the reconstruction level, where these parameters are fixed to $z_{cut} = 0.1$ and $\beta = 0$. The parameter β controls the grooming profile as a function of the subjet angular separation; for $= 0$, the algorithm is independent of subjet separation, and is equivalent to the modified mass-drop tagger (mMDT) detailed in [105]. The remaining constituents are combined to make the groomed or soft-drop jet.

Soft-drop has the benefit of performing jet grooming in a theoretically safer way [104, 106] with a constant behavior across different clustering distance parameters R and p_{T}, which is not true for other grooming algorithms such as the pruning technique [107, 108]. The jet observables computed after applying the soft-drop algorithm exhibit a pileup dependence and therefore it is applied to jets whose momenta has been weighted by the PUPPI algorithm.

Throughout this thesis, the usage of soft-drop is abbreviated as SD.

4.2.2 Jet mass

The simplest way to distinguish a boosted hadronically-decaying resonance from the hadronization products of light quarks and gluons is through the invariant mass of its decay products; in a contamination-free environment this would correspond to the jet mass, defined as the mass of the sum of the four-momenta of the constituents inside the jet. In particular, the jet mass generated by QCD radiation should be typically lower than that generated by the hadronic decay of heavy resonances.

However, in practice, the plain jet mass is very sensitive to contamination from soft and wide angle emissions. This is because the mass of generic quark and gluon jets, i.e. from QCD events, is set by the radiation of gluons from the hard portion of the jet. The probability that a gluon is radiated increases as its energy or angle of emission decreases (soft or collinear radiation, respectively). Hence, the probability of having $m = 0$ vanishes since many soft and/or collinear gluons will be radiated [109]. Above $m/p_{\mathrm{T}} \approx 0.1$, there is an overall decrease in the jet mass spectrum as the dominant contributions come from gluon emissions with high energy or at large angles. The appearance of a large peak at low mass is referred to as a "Sudakov peak" [105].

To separate the hard part of the jet from the soft contributions, the groomed jet mass is used, in our case the SD-jet mass: m_{SD}. By applying the SD algorithm, soft parts of a jet are removed. This results in a dramatically reduced Sudakov peak in the distribution of the jet mass for QCD events

and in a general reduction of the jet mass values. Conversely, the algorithm preserves the mass of jets from heavy boson decays. The effect of grooming the mass distribution of light-quark gluon jets is shown in Figure 4.6. The p_T dependence of the groomed jet mass is also shown in Figure 4.7. Finally, a comparison of the groomed jet mass shape between signal and background jets is shown in Figure 4.8.

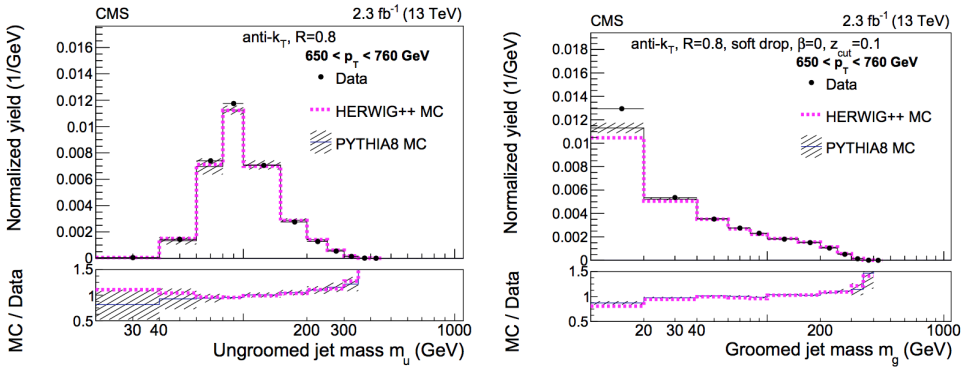

Figure 4.6: Shape of the ungroomed jet mass m_u and the groomed jet mass m_g, equivalent to m_{SD} in the text. The distributions are shown for inclusive multijet events in data and simulation (PYTHIA8 and HERWIG++). Figures reprinted from Reference [109].

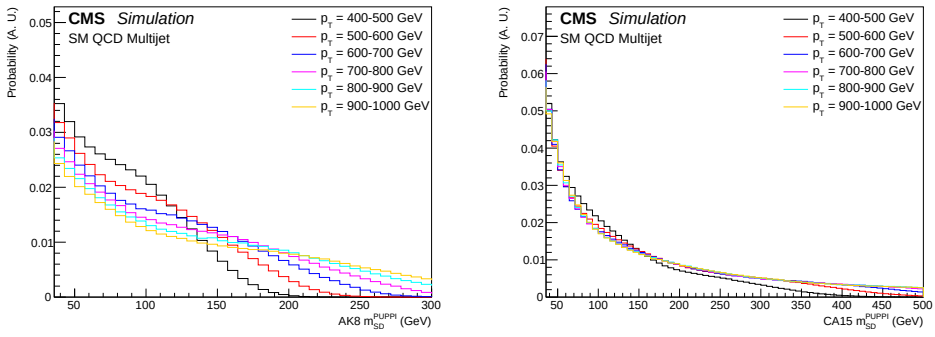

Figure 4.7: Shape of the groomed jet mass m_{SD}, shown for light quark and gluon jets, from simulated QCD multijet events. The distributions are normalized to unity and shown for different $p_T$ranges and different radius parameters, for AK8 (left) and CA15 (right) jets. The grooming procedure tends to lower the jet mass values for QCD and reduces the Sudakov peak, so that the distribution is mostly smoothly falling.

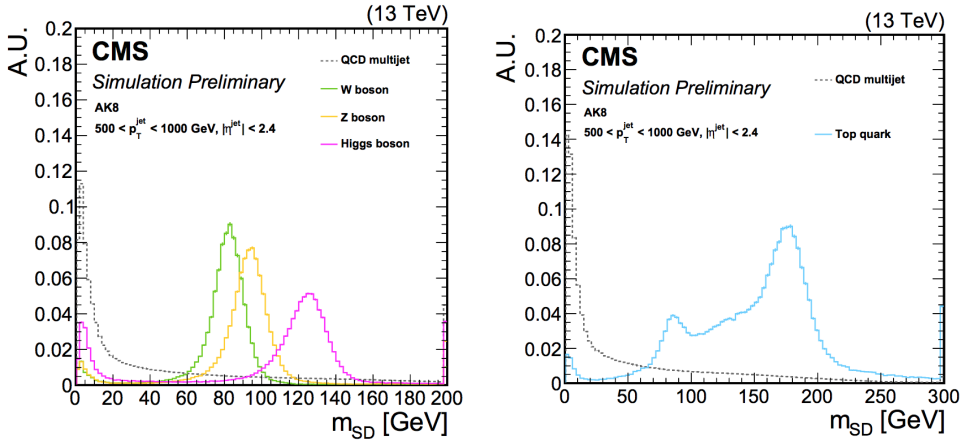

Figure 4.8: Comparison of the shape of the groomed jet mass m_{SD} for signal ($W/Z/h(qq)$) and background (QCD multijet) AK8 jets, in simulation. In the right, the mass distribution of the top quark jets has two peaks since not all its decay products will be merged within a jet radius of $R = 0.8$. Partially-merged $W(qq)$ decays from the top quark correspond to a mass window between 60 and 100 GeV. Figures reprinted from Reference [110].

An extra but important feature of the grooming procedure is that it reduces the sensitivity to non-perturbative contributions and "non-global" logarithms. The latter appear in the perturbative calculations of the jet mass. They arise from radiation that exits the jet re-radiating soft gluons back into the jet. Although it is currently unknown how to fully account for these non-global logarithms in the perturbative integrals, its effects can be reduced through grooming, thereby allowing for more exact predictions of the jet mass in the perturbative regime.

Simulation level corrections for the jet mass scale In CMS, the m_{SD} values are corrected by a p_{T} and η dependent factor to ensure a jet mass distribution that is centered on the nominal resonance mass. These corrections are derived in simulated events where W bosons that decay into a pair quarks are matched

to a large cone size jet. Two type of corrections are derived:

- **Generator-level correction**: That accounts for a small shift in the particle-level vector boson mass when compared to the nominal value of the W boson mass, 80.4 GeV.
- **Reconstruction-level correction**: That accounts for differences between the particle-level jet mass and the reconstructed jet mass, and is applied separately for jets in the barrel and endcaps regions.

The shift in generated soft-drop mass at lower p_{T} is of the order of 2-3% while the difference between reconstructed and generated soft-drop mass is a 5-10% effect. The mass shift introduced at generator level is corrected by a fit to $m_{\mathrm{PDG}}/m_{\mathrm{gen}}$ as a function of jet p_{T}, where $m_{\mathrm{PDG}} = 80.4$ GeV and m_{gen} is the fitted mean of the generator level mass. To correct for the residual shift between generator and reconstruction level, a fit to $(m_{\mathrm{gen}}/m_{\mathrm{reco}})$, where m_{reco} is the reconstructed jet mass. The distribution and corresponding fits for the two weights is shown in Fig. 4.9.

4.2.3 Jet ρ

In addition to the jet mass, it is useful to introduce another observable that characterizes correlations in jet substructure with the jet mass and p_{T}: the jet ρ [105]. This dimensionless variable is defined as:

$$\rho = \log \frac{m^2}{p_{\mathrm{T}}^2}. \tag{4.12}$$

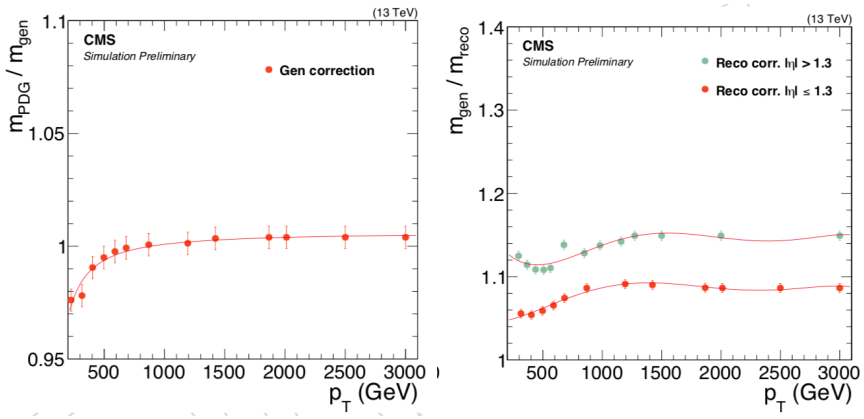

Figure 4.9: Fit to m_{PDG}/m_{gen} as a function of jet p_T (left), where $m_{PDG} = 80.4$ GeV and M_{gen} is the fitted mean of the generator level mass and $(M_{reco} - M_{gen})/M_{reco}$ (right), where M_{reco} is the reconstructed soft-drop mass, as a function of jet p_T in two η bins.

3

The advantage of the ρ observable over the groomed mass is its invariance under boosts along the jet direction, i.e. under a change in the jet p_T. This invariance is present in the perturbative mass regime, where the contributions scale as (m_{SD}/p_T). This behavior is shown in Fig. 4.10, where the distribution of ρ for simulated QCD multijet events varies very slowly as a function of p_T, contrary to the behavior of the jet mass shown in Fig. 4.7.

In general, the region of interest in ρ is determined by an upper or lower boundary. At low masses, non-perturbative effects are large and not well modeled in simulation, this region is avoided by a lower bound on ρ. The upper bound in ρ, that corresponds to higher jet masses, is imposed to avoid

[3]Note that some definitions of ρ factor out the dependence jet distance parameter R, $\rho = \log \frac{m^2}{p_T^2 R^2}$. To enforce consistency with past CMS results, we leave out the radius parameter, and instead account for different R values by varying the selection on ρ for different jet radius sizes.

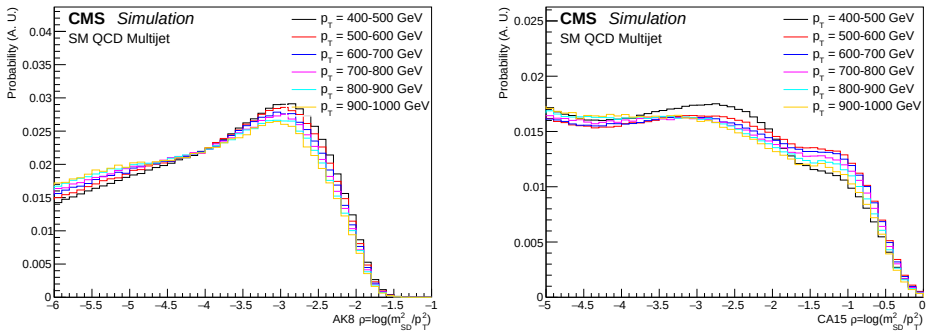

Figure 4.10: Shape of the jet $\rho = \log \frac{m^2}{p_T^2}$ shown for QCD multijet events. The distributions are normalized to unity and shown for different p_T ranges and different radius parameters, for AK8 (left) and CA15 (right) jets.

instabilities because the cone size of the jets becomes insufficient to provide complete cont ainment of the jet radiation.

4.2.4 Two-prong substructure

Along with the jet-mass, other observables are used to distinguish hadronically decaying heavy resonances such as prong-taggers. These aim to identify the multi-pronged structure of a heavy resonance decay. Examples of these observables include $N-$ subjettiness ratios τ_{ij} [111] and energy correlation functions (ECFs) [112]. These are usually constructed using power counting techniques, from a basis of IRC safe observables that probe an $N-$prong substructure within a jet. [4] Power counting [113] can predict which combinations of observables are optimally sensitive to specific parametric features within a jet and can elucidate the underlying physics probed by the observables.

[4] A general IRC safe observable, insensitive to the emission of soft or collinear gluons, can be constructed using all energy deposits and angular information of a hard scattering event.

N-subjettiness

The N-subjettiness (τ_N) is a measure of the compatibility of a jet with an N-axis hypothesis. It is defined as:

$$\tau_N = \sum_{1 \leq i \leq n_J} z_i \min\{\Delta R_{i1}^{\beta}, ..., \Delta R_{iN}^{\beta}\}, \tag{4.13}$$

where z_i refers again to the energy fraction, and ΔR_{iK} refers to the angular separation between the constituent i and the subjet axis K in the jet. A small τ_N indicates a high degree of compatibility with the N-axis hypothesis. In particular, the ratio of "2− subjettiness" to "1− subjettiness" ($\tau_2/\tau_1 = \tau_{21}$) is designed to take small values for a jet with well-resolved 2-prong substructure, and has therefore, excellent capability at separating jets originating from boosted $W/Z/Z'/h(qq)$ bosons from QCD jets.

The $N-$ subjettiness (τ_N) divides a jet into N sectors and correlates the particles in each sector with their corresponding axis. Thus, the definition of N-subjettiness requires an implicit definition of appropriate N-subjettiness axes, which can led to different behaviors of the observable.

Energy correlation functions

Energy correlation functions measure the correlation of the positions of hard particles in a jet [114]. They were first introduced in [112] as observables that would correlate 2 pairwise angles among n particles within a jet. In [114], this definition was generalized to v pairwise angles to introduce more flexibility in the angular scaling. They are denoted by ${}_v e_n^{\beta}$, where the subscript n denotes the number of particles to be correlated and β is an angular exponent that can

be used to adjust the weighting of the pairwise angle.

A generalized definition of the ECFs for a jet j is given by:

$$ {}_{v}e_{n}^{\beta} = \sum_{1 \leq i_1 < i_2 < \ldots < i_n \leq n} z_{i_1} z_{i_2} \ldots z_{i_n} \times \min \left\{ \prod_{m=1}^{v} \Delta R_{st}^{\beta} \,\middle|\, s < t \in i_1, i_2, \ldots, i_n \right\} \quad (4.14) $$

where z_i represents the energy fraction of the constituent i in the jet, and ΔR_{st} is the angular separation between constituents s and t.

Using these generalized correlators, it is possible to apply power counting to identify new jet substructure observables. For boosted vector boson tagging, ratios between the 2-point and 3-point (with its three variants) energy correlation functions can provide a well motivated tagger. For a 2-point (${}_{1}e_{2}$) and 3-point (${}_{2}e_{3}$) correlations, and $\beta = 1$, these observables are defined as:

$$ {}_{1}e_{2} = \sum_{1 \leq i < j \leq n} z_i z_j \Delta R_{ij}, \quad (4.15) $$

$$ {}_{2}e_{3} = \sum_{1 \leq i < j < k \leq n} z_i z_j z_k \min\{\Delta R_{ij}\Delta R_{ik}, \Delta R_{ij}\Delta R_{jk}, \Delta R_{ik}\Delta R_{jk}\}, \quad (4.16) $$

where z_i represents the energy fraction of the constituent i in the jet, and ΔR_{ij} is the angular separation between constituents i and j.

To distinguish a two-prong structure, the observable N_2^1 is defined as:

$$ N_2^1 = \frac{{}_{2}e_{3}}{({}_{1}e_{2})^2}. \quad (4.17) $$

This dimensionless variable is boost invariant, as two angular factors appear in both the numerator and denominator. Its definition is motivated by the behavior of ${}_{1}e_{2}$ and ${}_{2}e_{3}$ for light and quark gluons and signal jets:

- QCD jets, dominated by either soft or collinear radiation, exhibit a single

scaling: ${}_2e_3 \sim ({}_1e_2)^2$.

- Two-prong signal jets have larger 2-point correlations: ${}_2e_3 \ll ({}_1e_2)^2$.

Figure 4.11 shows the shapes of N_2^1 and τ_{21}. N_2^1 has smaller values for signal jets and higher values for background QCD jets. This variable exhibits τ_{21}-like behavior and similar discrimination but does not rely on the definition of axes within the jet. In this thesis, the calculation of N_2^1 is based on the PF jet constituents after application of the SD grooming algorithm to the jet.

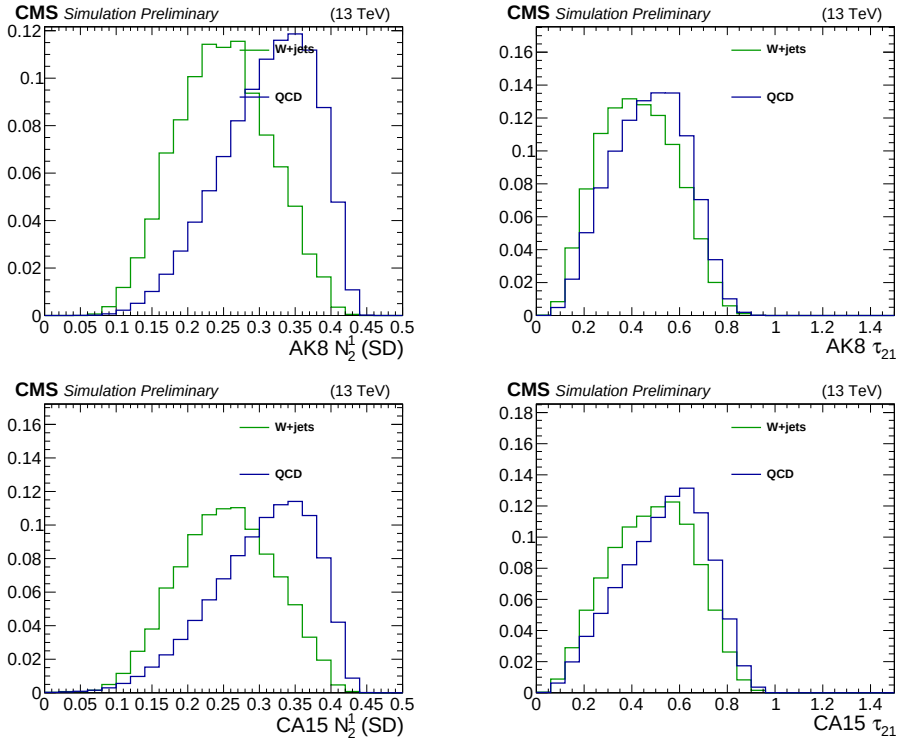

Figure 4.11: Shape of the groomed $N_2^{\beta=1}$ and ungroomed τ_{21} distributions in simulated two-prong W$\rightarrow qq$ boson jets (W+jets) and light quark and gluon jets (QCD). The comparison is shown for AK8 (top) and CA15 (bottom) jets.

Other two-prong taggers

An alternative approach to observables derived from an analytic, first-principles understanding of jet substructure is the development of Machine Learning (ML) classification algorithms. These allow for the construction of discriminators that can exploit low-level (high-dimensional) inputs, such as the properties of the constituents of a jet. In particular, jet taggers can be trained using *supervised learning*. This means that large datasets, that are available from MC simulation and have a known type or origin, can be used to tell the algorithm which jets correspond to signal and which to background.

In CMS, the Deep-AK8 [110] and Particle-Net [115, 116] taggers were recently developed as multi-class classifiers for the identification of hadronically decaying particles. These taggers were trained to distinguish five main categories of heavy resonance jets: $W/Z/H/t/other$. Although the architecture of these networks differs, the inputs that they use are similar. These are the measured properties of up to 100 jet constituent particles such as their p_{T}, the energy deposit, the charge, the angular separation between the particle and the jet axis or the subjet axes, etc. The usage of lower-level variables allows better exploitation of the high granularity of the CMS detector. For two-prong W tagging, this also leads to a significant improvement in performance in simulation when compared to the usage of τ_{21} as shown in Refs. [110, 116]. These developments are somewhat recent and are not used in the search presented in the next chapter.

Merged $b\bar{b}$ tagging

While $b-$tagging in busy hadronic environments plays an important role for final state signatures with $b-$jets, it is a key challenge to identify the merged $b\bar{b}$ signature. This final state is useful to identify boosted decays of the Higgs boson, that has the largest branching fraction to the $h \to b\bar{b}$ decay of about 60%, as well as boosted decays of the Z boson ($Z \to b\bar{b}$).

In the boosted regime there are two approaches to identify jets that likely originate from the merging of the fragmentation products of two b quarks:

- Rely on subjet b-tagging, where the CSVv2 algorithm described in Sec. 4.1.2 is applied to two small-radius AK4 jets. At high-p_{T}, however, when the subjets start to merge, the 'standard' $b-$tagging techniques start to break down due to the overlap in the charged tracks and SVs used when computing the subjet $b-$tag discriminants.

- Develop dedicated algorithms that are based on deep neural networks or multivariate techniques and attempt to fully exploit the strong correlations between the B hadron flight directions and the energy flows of the two decay products.

In CMS, the latter strategy is preferred and a discriminant based on a deep neural network is used. It is referred to as the deep double-b tagger (DDBT) [89, 117]. The algorithm takes as inputs several high-level observables that characterize the distinct properties of B hadrons and their momentum directions in relation to the two subjet candidate axes, as well as low-level track and vertex observables. When compared to the subjet $b-$tagging approach,

the DDBT improves the $h \to b\bar{b}$ tagging efficiency by a factor of about 2 for the same detector conditions and QCD misidentification probability [117].

4.2.5 Mass decorrelation

Jet features used for tagging are often correlated with the resonant feature, i.e. the jet mass. This correlation is present in both jet prong-tagging observables as well as in ML jet classifiers, where the latter can extract features that are correlated to the mass to improve the discrimination power. This means that a fixed-value requirement on these observables will distort the jet mass distribution differently depending on the jet p_{T}. In particular, for inclusive jet production i.e. QCD jets, the shape of the jet mass will no longer be smoothly falling but will resemble the resonance jet mass distribution, as shown in Fig. 4.12.

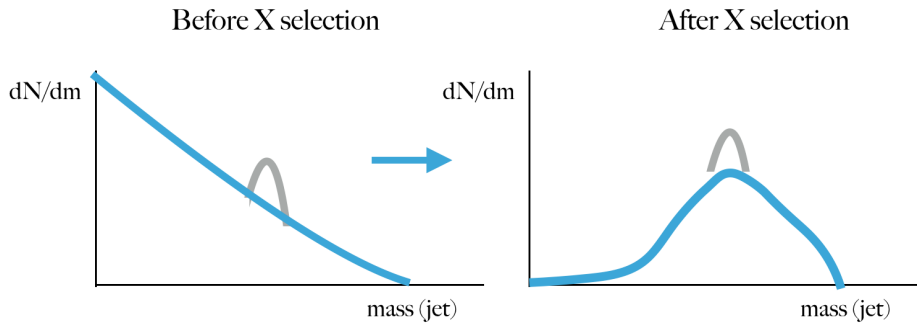

Figure 4.12: Illustration of the distortion or *sculpting* of the shape of the inclusive jet mass distribution after a selection on an observable X that is correlated with the jet mass and p_{T}. The jet mass shape of the background (light blue) becomes similar to that of the resonant signal (gray) after a selection on the tagger. This complicates background predictions on searches or measurements that use the jet mass as the main observable. For example, those that rely on a side-band prediction in the jet mass around the resonance mass or those that rely on a smoothly falling shape constrained by a similar shape in a signal-depleted region.

This "sculpting" behavior is undesirable in analyses where the jet mass is the main observable. These searches or measurements often need to estimate the background contribution from data and require predictable and smooth transitions from signal-depleted regions to signal regions. Furthermore, even if a background prediction was available, fitting a resonant peak on top of a sculpted background would increase the systematic uncertainties on the multijet background and reduce the sensitivity of the analysis.

In the following, we describe one technique to decorrelate jet observables from the jet mass and p_{T} that is called *designed decorrelated tagger* or DDT. This technique was first derived for the τ_{21} observable in Ref. [118] and its generalization to any observable and background efficiency is one of the main results of this thesis. The development of decorrelated jet substructure taggers is an active field and a description of other techniques can be found in Refs. [116, 118–122].

Jet mass correlation We begin by considering the correlations present in prong-taggers such as τ_{21} and N_2^1. These are shown for simulated QCD multijet events in Figures 4.13 and 4.14 with respect to the jet ρ for different p_{T} ranges. The correlations are drawn in a "violin plot" style, where the central black dots represent the mean value of τ_{21} or N_2^1 in each jet ρ bin. We can already observe that, when compared to τ_{21}, the observable N_2^1 is more stable with respect to the jet mass and p_{T}.

Designed decorrelated tagger (DDT) The DDT method relies in *profiling* the linear relationship between τ_{21} and the jet ρ, as seen in Fig. 4.13. This

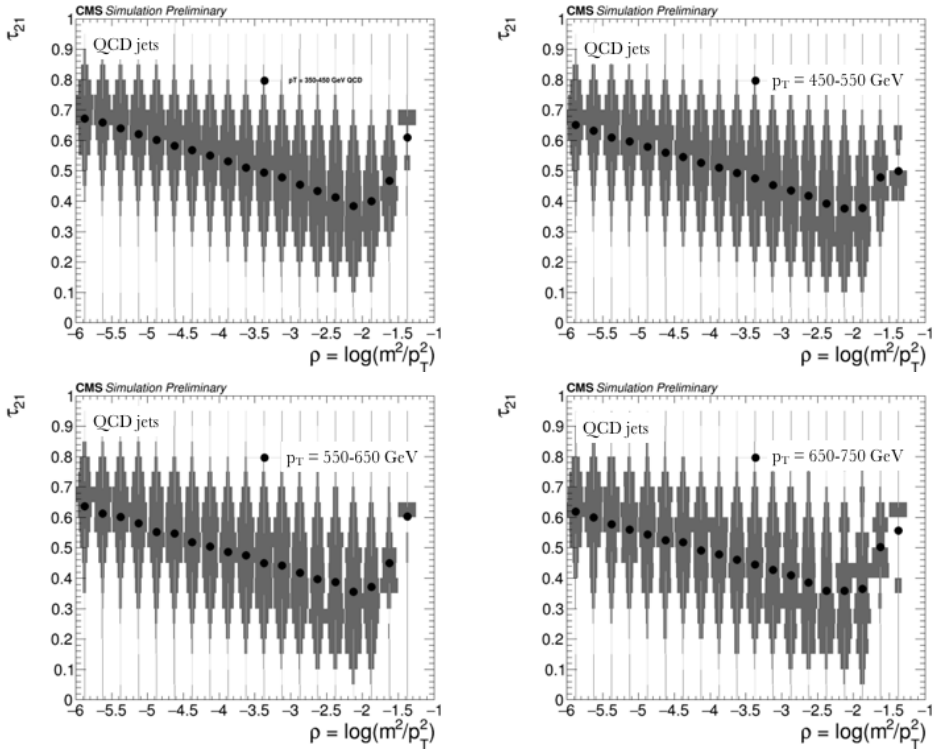

Figure 4.13: Violin plot showing the correlations between the τ_{21} observable and the jet ρ for inclusive jet production (simulated QCD events) and for different p_T ranges from 450 GeV to 750 GeV. The black dot shows the mean value of τ_{21} in each jet ρ bin.

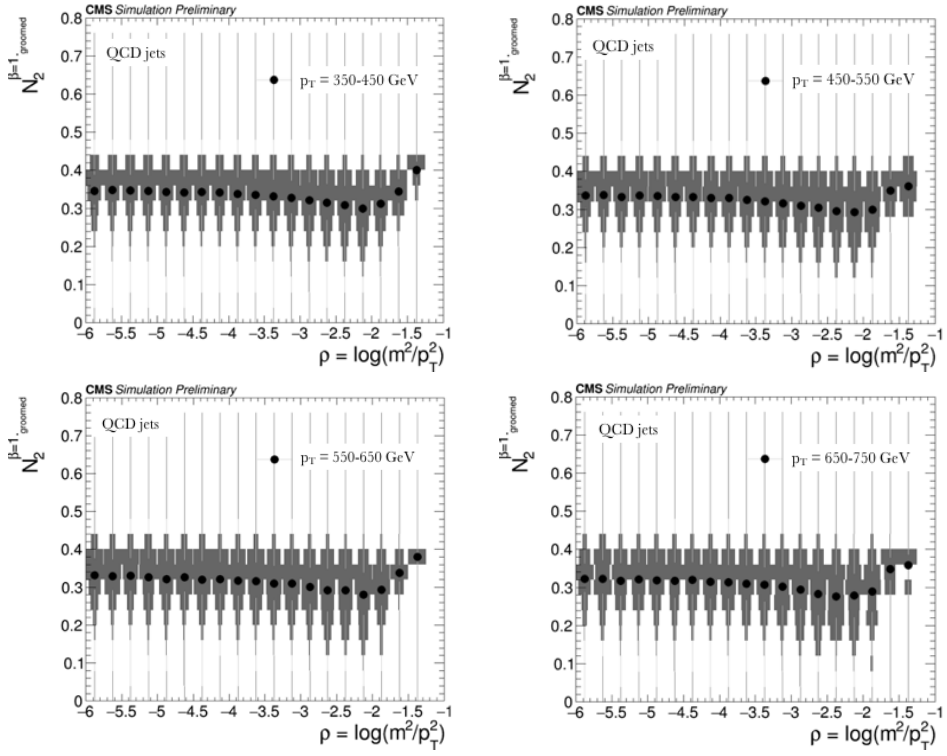

Figure 4.14: Violin plot showing the correlations between the N_2^1 observable and the jet ρ for inclusive jet production (simulated QCD events) and for different p_T ranges from 450 GeV to 750 GeV. The black dot shows the mean value of N_2^1 in each jet ρ bin.

dependence can be exploited to perform a linear transformation, that removes the *mean* bias of τ_{21} with respect to ρ.

In practice, when this method was studied in [118, 123], the jet scaling variable: $\rho^{\text{DDT}} = \log(\frac{m^2}{p_{\text{T}} \times 1\text{GeV}})$ was used. The linear transformation of $\tau_{21} \rightarrow \tau_{21}^{\text{DDT}}$ is given by:

$$\tau_{21}^{\text{DDT}} = \tau_{21} - M \times \rho^{\text{DDT}} \tag{4.18}$$

where $M = 0.063$ is a constant. To obtain M, a linear fit is performed to the simulated τ_{21} profile in the range of $\rho^{\text{DDT}} \in [1.5, 4.0]$. This range of ρ^{DDT} is within the region of interest of simulated QCD events, as explained in Sec. 4.2.3. Figure 4.15 shows how the DDT transformation removes most of the linear correlation of τ_{21} with ρ.

Fixed-efficiency regression The original DDT transform proposes a linear transformation with the jet mass to remove the mean bias of an observable, but such a relationship will not always be linear and is not applicable to every observable.

To generalize this concept, the decorrelation procedure is no longer applied along the profile (50% quantile) of the distribution but instead along a fixed quantile. The quantile should correspond to the inclusive background efficiency, ϵ_{bkg}, at which the selection on a given observable will be applied. For example, a search may want to perform a selection on a jet observable Y that only keeps 5% of the QCD multijet background, i.e. $\epsilon_{\text{bkg}} = 5\%$. By definition, the 5% quantile of Y corresponds to the cut value that divides the multijet events into groups with 5 and 95% of background efficiency. This is

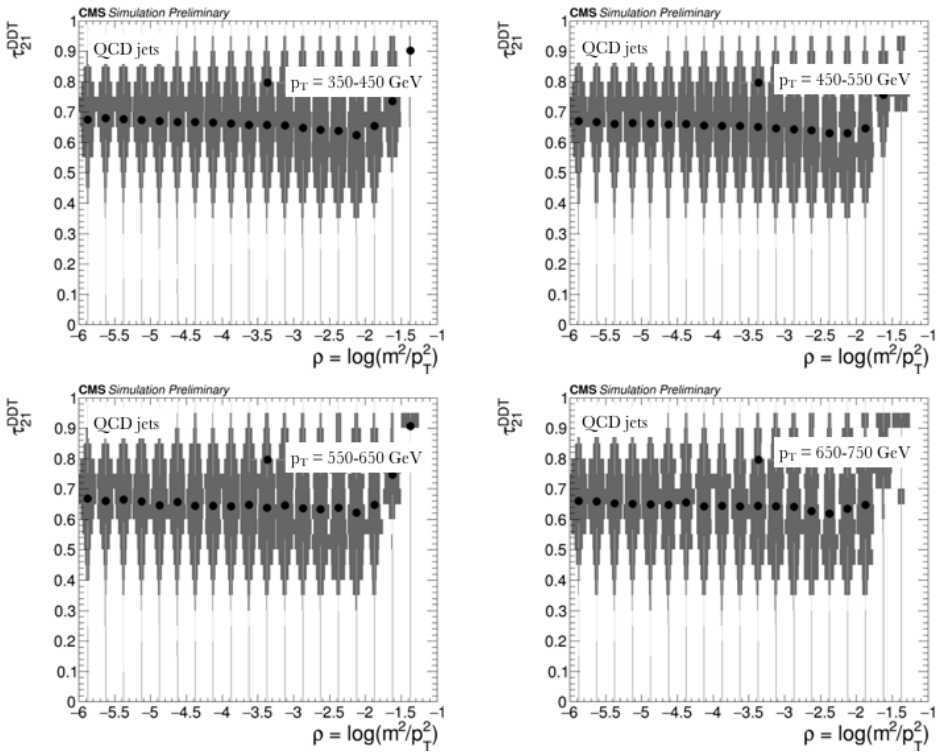

Figure 4.15: Violin plot showing the correlations between the τ_{21}^{DDT} observable and the jet ρ for inclusive jet production and for different p_T ranges from 450 GeV to 750 GeV. The black dots show the profile or mean value of τ_{21}^{DDT} in each jet ρ bin. Since this variable is obtained with respect to a linear fit to ρ^{DDT} instead of ρ, the decorrelation is not perfect but is greatly reduced in the region of interest.

illustrated for the jet N_2^1 observable in Fig. 4.16. This cut value, however, will vary as a function of the jet ρ and p_T.

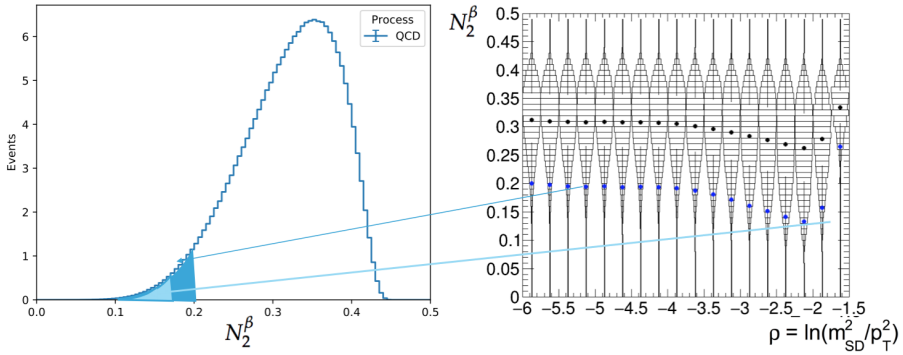

Figure 4.16: Quantile for the jet N_2^1 in simulated multijet events. The left panel shows a distribution of N_2^1 while the right panel shows a violin plot on the correlation of N_2^1 with respect to ρ. The blue points represent the 5% quantile while the black points represent the mean or 50% quantile. A threshold selection on N_2 on the 5% quantile will result on a selection of events in the blue shaded regions. However, given the correlation with the jet ρ, for a fixed quantile the shaded region will vary depending on the jet ρ value.

We can explore the variation of the cut value by building a fine map, in jet ρ and p_T, of the ϵ_{bkg} quantile of N_2^1: $X_{\epsilon_{bkg}}$. This binned map is shown for $\epsilon_{bkg} = 5\%$ in Fig. 4.17.

Thus, for a certain fixed ϵ_{bkg}, we can define the following transformation from $N_2^1 \to N_2^{1,DDT}$.

$$N_2^{1,DDT} = N_2^1 - X_{\epsilon_{bkg}}, \tag{4.19}$$

where $X_{\epsilon_{bkg}}$ is the ϵ_{bkg} percentile of the N_2^1 distribution in simulated QCD multijet events. The selection $N_2^{1,DDT} < 0$, or equivalently $N_2^1 < X_{\epsilon_{bkg}}$, yields a constant ϵ_{bkg} of simulated QCD multijet events.

By definition, this transformation takes into account any variation of the

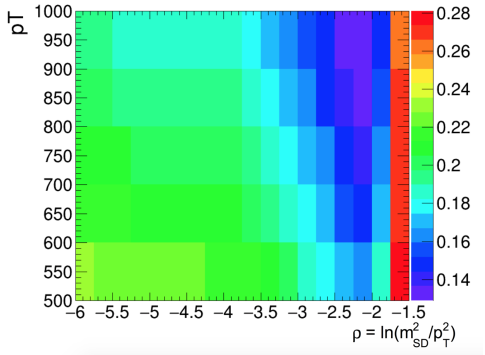

Figure 4.17: Variation of the $X_{5\%}$ of the jet N_2^1 observable, as a function of the jet ρ and p_{T}. This map is built from simulated multijet events and uses a coarse binning of the jet ρ and jet p_{T} variables.

cut with respect to the jet ρ. This is seen in Fig. 4.18, where the violin plot of the $N_2^{1,\mathrm{DDT}}$ variable shows a flat behavior for the quantile corresponding to the transformation. A drawback of this method is that the decorrelation is only perfect for the given working point or ϵ_{bkg} for which the $X_{\epsilon_{\mathrm{bkg}}}$ map is derived.

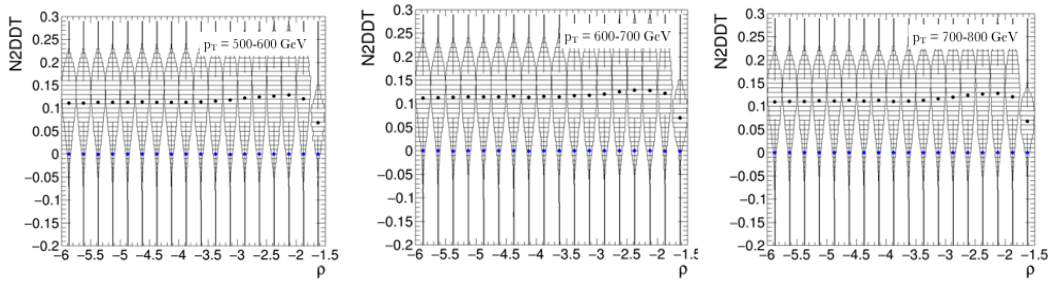

Figure 4.18: Violin plot showing the correlations between the $N_2^{1,\mathrm{DDT}}$ observable and the jet ρ for inclusive jet production and for different p_{T} ranges from 500 GeV to 800 GeV. The black dots shows the profile or mean value of τ_{21}^{DDT} while the blue dots show the 5% quantile, in each jet ρ bin. Since the $N_2^{1,\mathrm{DDT}}$ transformation was derived for a 5% background efficiency, there is no variation on the blue dots across the entire jet ρ range. This leads to a perfect decorrelation in simulation.

The $N_2^{1,\mathrm{DDT}}$ is effectively only a translation of the original jet N_2^1 observable,

and thus a similar discrimination performance is expected. The distributions of both observables are shown in Fig. 4.19 for $\epsilon_{\text{bkg}} = 5\%$.

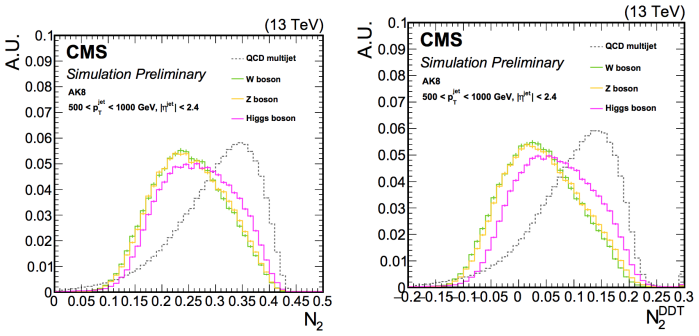

Figure 4.19: Distributions of the jet N_2^1 observable (left) and $N_2^{1,\text{DDT}}$ (right) in two-prong signal and background jets. Figures referenced from [117].

Smooth $N_2^{1,\text{DDT}}$ The binned transformation map shown in Fig. 4.17 is discretized. To ensure that the full differential variation of N_2^1, or any other variable, is exploited, the distribution of $X_{\epsilon_{\text{bkg}}}$ can be smoothed. Three smoothing approaches were studied: (a) a k-Nearest neighbor (kNN) approach, (b) a Gaussian filter, and (c) a "detector-smearing" approach that generates a thousand times more the original number of simulated events and smears them to account for detector effects. While the techniques used in (a) and (b) are straight-forward methods employed in image processing, they are limited by the number of events available in the QCD multijet simulation which are used to create the transformation map. The detector-smearing approach aims to overcome this potential pitfall by generating many more events at particle-level, and smearing the distributions of N_2^1 and jet ρ. In this technique, both jet observables are multiplied by a random number drawn from a Gaussian distribution, such that the smeared jet matches the resolution obtained from

fully simulated events. The advantage of this method over the kNN approach is that it allows better control of the smoothness of the transformation map while maintaining similar performance in terms of the amount of jet mass decorrelation. However, its implementation is time consuming, whereas for example a Gaussian filter is much faster and easier to implement. Thus, the later methods should be preferred in most of the use-cases. Examples of the smooth $N_2^{1,\mathrm{DDT}}$ map are given in Fig. 4.20.

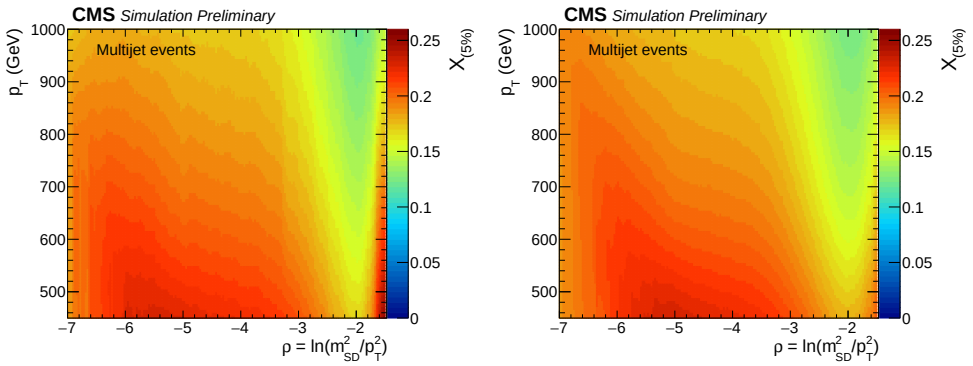

Figure 4.20: Smoothed map of the variation of the $X_{5\%}$ quantile for the jet N_2^1 observable, as a function of the jet ρ and p_T. This map is built from simulated multijet events and uses a fine binning of the jet ρ and jet p_T variables. It is further smoothed with a k-Nearest neighbor approach (left) and a detector-smearing approach (right).

Performance in simulation In Refs. [117, 122], the Jensen-Shannon divergence [124] was used as a figure of merit to quantify the sculpting of the jet mass distribution. This metric is a symmetric version of the Kullback-Leibler divergence [125] and provides a metric for the similarity of the shape between

distributions. [5] The JSD values for a mass decorrelated jet tagger should be lower than for a standard jet tagger. In both independent studies, it was observed that the generalization of the DDT method, e.g. the $N_2^{1,\mathrm{DDT}}$ observable keeps a smoothly falling shape of the jet mass distribution. Furthermore, when compared to other available methods, it is seen to lead to the greatest degree of mass-decorrelation and lowest values of JSD. This is seen in Fig. 4.21.

Summary The fixed-efficiency regression method generalizes the central concept behind DDT, thereby making the method admissible to any fixed background efficiency. When the transformation is applied, a perfect decorrelation should be expected at a given ϵ_{bkg}.

It is worth noting that this method is not only applicable to jet-prong observables. A similar procedure can be applied to the discriminator score which is the output of ML-based classifiers. This has in fact been recently done for the Deep-AK8 tagger with a similar mass decorrelation performance achieved [116]. The simplicity of this decorrelation method has an advantage over other ML decorrelation techniques such as adversarial training [119],

[5] The Jensen-Shannon divergence (JSD) was used to measure the difference between the normalized mass distributions of the background jets passing and failing a jet tagger cut:

$$\mathrm{JSD} \equiv \mathrm{JSD}\left(\frac{N_{\mathrm{bkg}}^{\mathrm{pass}}(m)}{\sum_i N_{\mathrm{bkg},i}^{\mathrm{pass}}(m)} \| \frac{N_{\mathrm{bkg}}^{\mathrm{fail}}(m)}{\sum_i N_{\mathrm{bkg},i}^{\mathrm{fail}}(m)}\right), \tag{4.20}$$

using bins in m between 50 GeV and 300 GeV in increments of 5 GeV. Here, JSD is defined as:

$$\mathrm{JSD}(P\|Q) = \frac{1}{2}\left(\mathrm{KL}(P\|M) + \mathrm{KL}(Q\|M)\right), \text{with } M = \frac{P+Q}{2}. \tag{4.21}$$

$\mathrm{KL}(P\|Q) = -\sum_i P_i \log_n\left(\frac{Q_i}{P_i}\right)$, is the Kullback-Leibler divergence that measures the relative entropy of P with respect to Q, and cand therefore be used to quantify the similarity between discrete distributions P and Q.

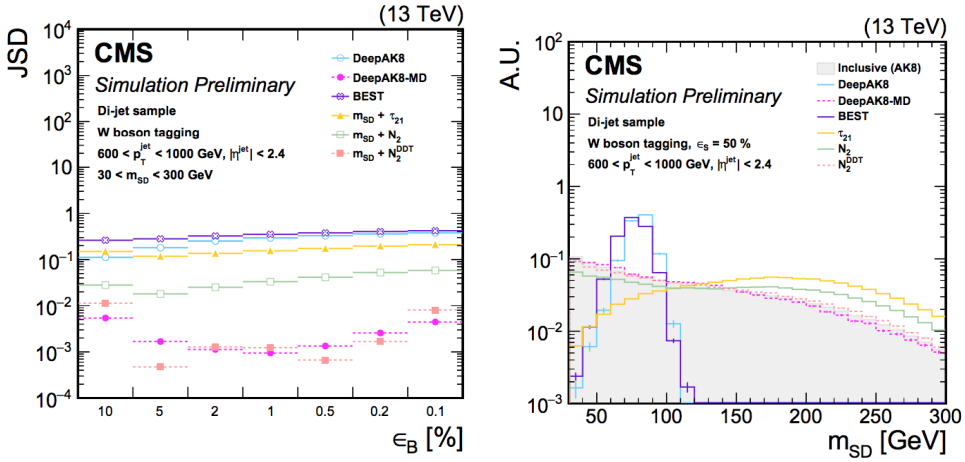

Figure 4.21: Comparison of the mass decorrelation performance of the $N_2^{1,\mathrm{DDT}}$ observable with respect to other W-tagging algorithms in CMS. *Left*: The JSD as a function of successively tighter selections (expressed in terms of ϵ_{bkg}) for the various algorithms. Lower values of JSD indicate larger similarity of the mass shape in QCD multijet events passing and failing the selection on the tagging algorithm. *Right*: The shape of the soft-drop jet mass distribution for background jets with $600 < p_{\mathrm{T}} < 1000$ GeV, inclusively and after a selection by each algorithm. The working point chosen corresponds to $\epsilon_{\mathrm{sig}} = 50\%$. Since $N_2^{1,\mathrm{DDT}}$ it is computed for a fixed background efficiency instead of a fixed signal efficiency, the shape of the jet mass after the selection does not perfectly match the smoothly falling shape of the inclusive selection. but is very close. Figures referenced from [117].

which is not robust or easy to implement.

In the next chapter, the $N_2^{1,\mathrm{DDT}}$ observable is used as the main jet tagger to identify a two-prong signal. The selection $N_2^{1,\mathrm{DDT}} < 0$ yields a constant QCD background efficiency across the ρ and p_T range of the search, and preserves the shape of the jet mass distribution used in the search.

4.2.6 Validation in data

Prior to the usage of the any of the jet substructure tools presented in this Chapter, we must verify the performance in data. The corrections to be obtained are related to the response of the jet mass distribution and the efficiency of a two-prong tagger cut. A data sample enriched in lepton+jets $t\bar{t}$ production is used, that is, where one top quark decays leptonically $t \to W(\ell\nu)b$ and the other decays hadronically $t \to W(qq)b$. The leptonic top quark candidate is selected by identifying a high p_{T} lepton, more concretely a muon, and a $b-$jet and the fully-merged hadronic W jet can be studied in data. The hadronic jet is reconstructed as an AK8 jet with $p_{\mathrm{T}} > 200$ GeV. To identify W jets, a two-prong substructure cut is used, e.g. $N_2^{1,\mathrm{DDT}} < 0$, events in this sample determine the "passing" region. The "failing" region is composed of events with $N_2^{1,\mathrm{DDT}} > 0$ values. Large-radius jets in the passing region show a peak at the W mass in the jet mass distribution, as shown in Figure 4.22.

A simultaneous fit to the two samples can be performed to extract the tagging efficiency of a merged W boson jet, its jet mass scale, and the resolution in simulation and in data. The peak of the W mass in the passing region can be used to determine the scale and resolution of the jet mass. The Jet Mass Scale (JMS) is defined as the ratio of the mean of the distribution in data vs simulation:

$$\mathrm{JMS} = \frac{< m_{\mathrm{Data}} >}{< m_{\mathrm{MC}} >}, \tag{4.22}$$

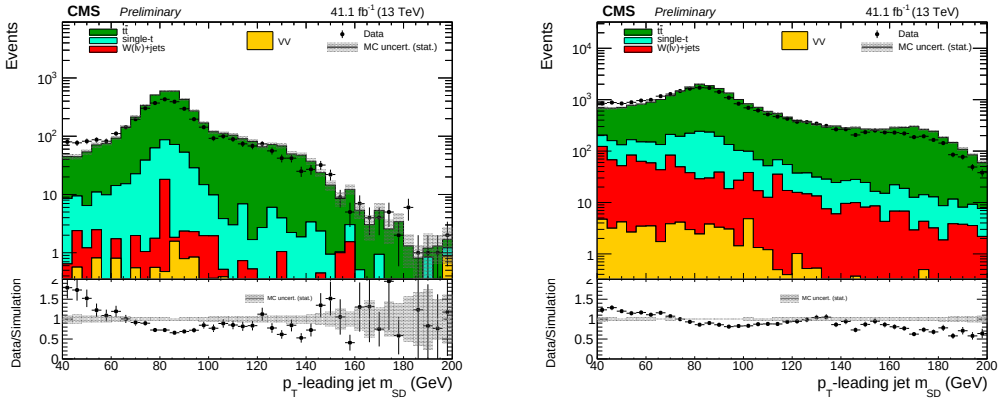

Figure 4.22: Distributions of the soft-drop jet mass that pass (left) and fail (right) the $N_2^{1,\mathrm{DDT}}$ selection. Data, corresponding to 2017 data taking, is shown in black points and simulation is shown in the shaded regions. A large peak is seen around the W mass (80.4 GeV) in the passing region and less so in the failing region, this corresponds to merged hadronic decays of the W boson. In the failing region there is a second and less pronounced bump around the top quark mass (175 GeV) that correspond to fully merged decays of the top quark, where the $b-$jet is close enough to the hadronic W boson jet.

while, the Jet Mass Resolution (JMR) is obtained from the width of the distribution:

$$\mathrm{JMS} = \frac{\sigma(m_{\mathrm{Data}})}{\sigma(m_{\mathrm{MC}})}. \tag{4.23}$$

To account for residual differences in the efficiency of the two-prong tagger between data and simulation, the failing region is also taken into account and a scale factor is defined:

$$\mathrm{SF} = \frac{\epsilon_{\mathrm{Data}}(N_2^{1,\mathrm{DDT}} > 0)}{\epsilon_{\mathrm{MC}}(N_2^{1,\mathrm{DDT}} > 0)}. \tag{4.24}$$

Each of these corrections are obtained from a simultaneous fit to the passing

and failing regions that have the following partial likelihoods:

$$\mathcal{L}_{\text{pass}} = \prod_i [N_W \cdot \epsilon \cdot f^{sig}_{\text{pass}}(m_j) + N_2 \cdot f^{bkg}_{\text{pass}}(m_j) + N^{ST}_{\text{pass}} \cdot f^{ST}_{\text{pass}} + N^{Wjets}_{\text{pass}} \cdot f^{Wjets}_{\text{pass}}]$$

$$\mathcal{L}_{\text{fail}} = \prod_i [N_W \cdot (1-\epsilon) \cdot f^{sig}_{\text{fail}}(m_j) + N_2 \cdot f^{bkg}_{\text{fail}}(m_j) + N^{ST}_{\text{fail}} \cdot f^{ST}_{\text{fail}} + N^{Wjets}_{\text{fail}} \cdot f^{Wjets}_{\text{fail}}],$$

where N_W is the number real W-jets, N_2 is the number of combinatorial background events passing and failing the $N_2^{1,\text{DDT}}$ cut respectively, N_i and f_i with i = ST,W jets are the normalizations and models of the minor backgrounds in this sample. The shapes and normalizations of the minor backgrounds are fixed to what is given by simulations, while the floating parameters of the fit are the rates N_W, N_2, and the mean and sigma of the W-mass distribution defined in $f^{\text{sig}}_{\text{pass}}(m_j)$ and $f^{\text{sig}}_{\text{fail}}(m_j)$. The post-fit shapes are shown in Figs. 4.23 for different data taking periods, for an $N_2^{1,\text{DDT}}$ selection with $\epsilon_{\text{bkg}} = 5\%$.

The scale factors are measured separately for the two data taking periods considered in this thesis and shown in Table 4.1. The excellent performance of the PF algorithm results in a JMR of about 10%. The absolute response and the resolution are well described by the simulation, within 1% for the JMS and about 6% for the JMR, which is about the same size as the statistical uncertainty of this measurement.

Table 4.1: Scale factors for the W-tagging efficiency, the JMS and the JMR.

Run-period	$N_2^{1,\text{DDT}}$ selection	Jet mass scale	Jet mass resolution
2016	0.891±0.066	1.000±0.006	1.110±0.060
2017	0.896±0.088	0.989±0.006	1.082±0.067

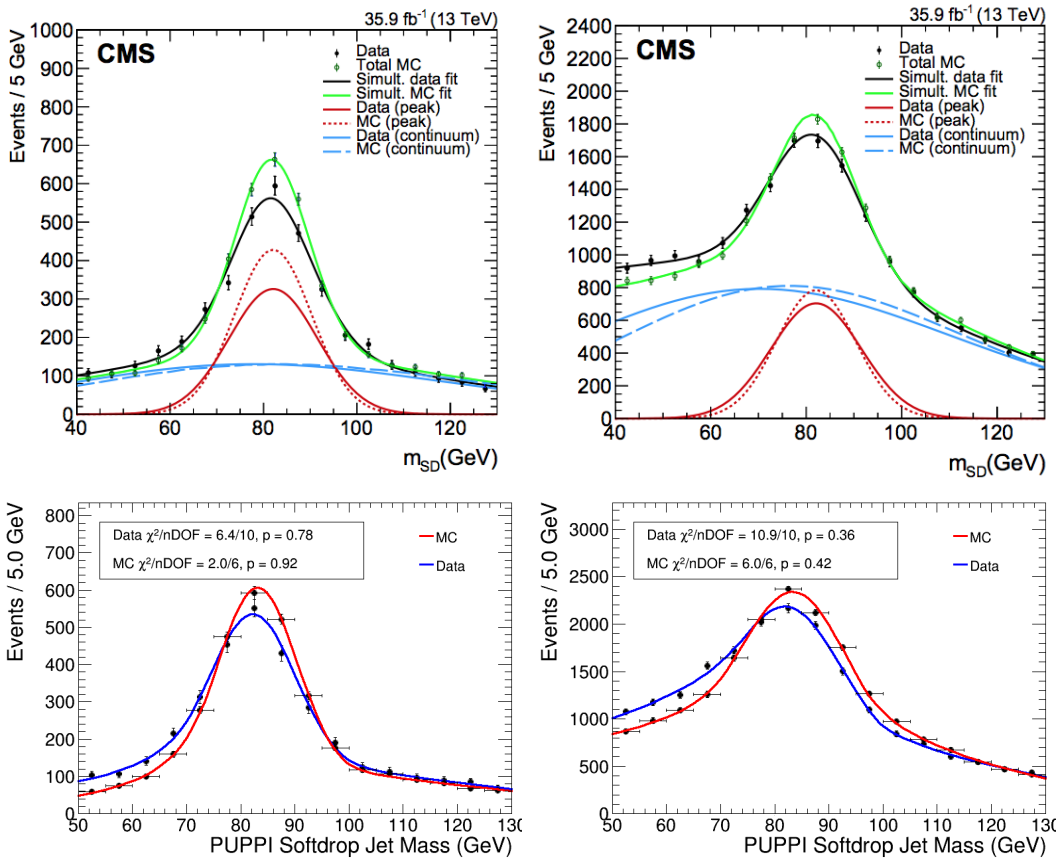

Figure 4.23: Soft-drop jet mass distributions that pass (left) and fail (right) the $N_2^{1,\mathrm{DDT}}$ selection, at 5% background efficiency, in the semileptonic $t\bar{t}$ sample. Data corresponds to 2016 (top) and 2017 (bottom) data taking periods. Note that for 2016, pre-fit and post-fit models are shown, while only post-fit is shown for 2017.

As the semileptonic $t\bar{t}$ sample does not contain a large population of very energetic jets, an additional systematic uncertainty can be computed to account for the extrapolation to very high p_{T} jets. This additional uncertainty is estimated to be 0.5% per 100 GeV, based on a study of fitting the jet mass distributions of p_{T}-binned samples of merged top quark jets with $p_{\mathrm{T}} > 350$ GeV [13]. In total, the jet mass scale uncertainty increases with jet p_{T}, ranging from 1.2% at 450 GeV to 2.1% at 800 GeV.

Chapter 5

Search for a low mass resonance decaying into quarks

In this chapter, we discuss the search for vector resonances decaying to quark-antiquark pairs. The search focuses on resonances with masses below 450 GeV and a natural width small relative to the detector's mass resolution.

As the resonance mass decreases, there is a large increase in the cross section of multijet backgrounds and high energy trigger thresholds are needed to reduce the data recording rate. The energy of the resonance hadronic decay products is therefore not sufficient to meet the trigger requirements. This leads to a loss of sensitivity in dijet searches for resonance masses below 1 TeV. However, a sufficiently light resonance can be triggered when produced in association with initial state radiation. The ISR constraint provides enough energy in the event to satisfy the trigger, either by the ISR jet or by the resonance itself. Furthermore, the minimum p_{T} of the resonance in this regime is sufficiently high that the hadronization products of the daughter quarks merge and are reconstructed as a single, large-radius jet. This topology is illustrated in Fig. 5.1.

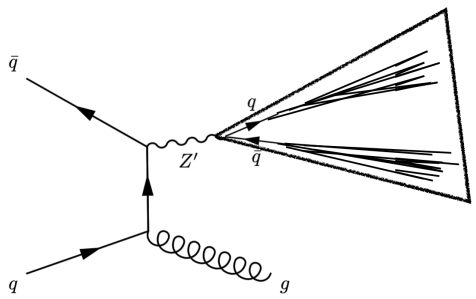

Figure 5.1: An s-channel Z′ mediator resonance produced in association with an ISR jet.

We interpret the results of this search within the framework of a leptophobic vector resonance (Z′) model. This model includes a new U(1) gauge sector and a new vector boson Z′. As anticipated in Sec. 2.3.2, these type of low-mass Z′ models are especially appealing as a potential mediator between the standard model and the dark sector. Furthermore, they can constrain other extension theories of the SM, such as models with new gauge symmetries, amongst others [126–135] Such models can avoid flavor constraints if the couplings to quarks are the same for each generation. Here, we consider a simple extension to the SM with a single extra Z′ boson which couples exclusively and equally to all quarks by adding the Lagrangian term:

$$\mathcal{L} \sim g'_{\mathrm{q}} \bar{\mathrm{q}} \gamma^{\mu} \mathrm{q} Z'_{\mu}, \tag{5.1}$$

where g'_{q} is the coupling to quarks, and q and $\bar{\mathrm{q}}$ are the quark fields.

Previous results from dijet searches can be summarized in the g'_{q} coupling-mass plane. The latest results from searches at the LHC experiments have excluded resonance masses between 1.0 and 7.6 TeV [3, 4]. For masses between 0.45 and 1 TeV, limits on resonances have been set by trigger-level analyses

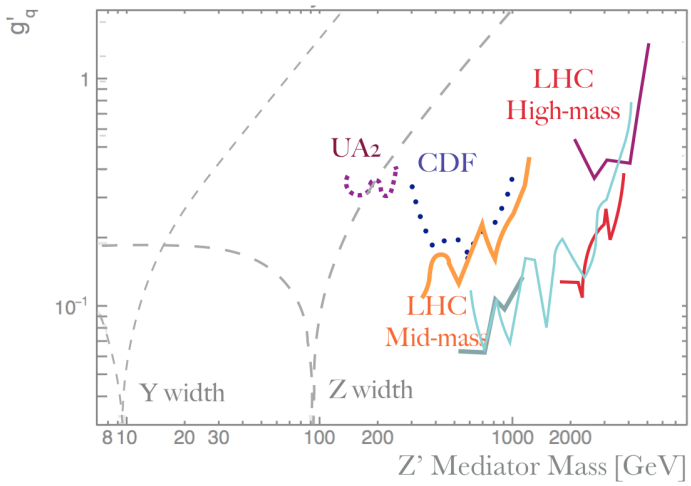

Figure 5.2: Limits on the universal coupling g'_q between a leptophobic ZIJôóš boson and quarks for various dijet analyses from UA2, CDF/D0 and CMS experiments that aim both lower and higher Z′ masses. Also shown are indirect constraints on g'_q from the Y and Z boson widths, which are valid for all values of Γ_Z/M_Z. This figure is reproduced and simplified from Ref. [136].

that record only partial event information and perform searches in the dijet mass spectrum with lower trigger thresholds [4–7]. A rough summary of the previous results, obtained by the UA2, CDF/D0, and CMS experiments is given in Fig. 5.2. The figure roughly illustrates most of the excluded mass and coupling range but emphasizes that for masses below 140 GeV no exclusions have been placed since the UA2 experiment. A detailed analysis of the most recent CMS results can be found in Ref. [136] and a full picture, including the results of this search, will be shown at the end of this Chapter.

5.1 Event simulation and selection

In this section, we describe how events with boosted Z′ candidates are simulated and selected.

The main criteria in this search is the selection of a *high-p_{T} large radius jet*. The most energetic jet in the event is assumed to correspond to the $Z' \to qq$ system, and is reconstructed as a single AK8 or CA15 jet. The search using 2016 data only used AK8 jets and probed resonances with masses up to 300 GeV, while the 2017 search extended the mass region up to 450 GeV by using larger-radius CA15 jets. In the 2017 search, the AK8 jets provide better sensitivity for signal mass hypotheses below 175 GeV, while the CA15 jets provide better sensitivity at mass hypotheses above 175 GeV. The large radius jet is required to have the *two-prong substructure* expected from the Z' signal.

From these simple statements we can identify the main processes that would mimic our signal, i.e. the background processes of this search:

- **Inclusive jet production or QCD multijet**: The cross section of this process increases as the jet mass decreases which makes it the dominant background in the low-mass regime. The hadronization products of light quarks and gluons produced in pp collisions can easily be clustered together in a single large radius jet. Grooming and jet substructure observables can help lower the QCD contribution and distinguish a two-prong jet, as explained in Sec. 4.2.

- **W and Z + jets**: When the W or Z bosons are produced in association with an ISR jet, the event signature will exactly mimic the Z' signal. Large radius jets coming from the hadronization products of the W and Z boson will also have a two-prong substructure consistent with the signal. The main handle to identify these backgrounds is the jet mass, for which these processes will exhibit a peak at the W and Z pole masses,

respectively.

- **Top quark production**: The semi-leptonic of fully hadronic decay of a top quark in association with a jet (Single t) or the decay a pair of top quarks ($t\bar{t}$) can mimic a large-radius jet signature. The two-prong substructure selection can help lower the contribution from these events.

5.1.1 Simulation

To estimate the behavior and contribution of the three main background processes and the Z' signal, MC generators are used to produce simulated datasets. The generation of simulated events includes the matrix-element, parton shower and detector response elements described in Sections 2.2 and 3.3.

The Z'+jet(s) signal events are generated at leading order (LO) with the MADGRAPH5_aMC@NLO 2.4.3 generator [35], for various mass hypotheses in the range 50–450 GeV. The events are generated with one or two jets in the matrix element calculations and a parton-level filter requires the scalar sum of transverse energies of all the jets in the event (HT) to satisfy the condition $HT > 400$ GeV. These signal events generally satisfy the event topology with the presence of large ISR.

The MADGRAPH5 generator is also used to simulate background processes, including multijet, Z+jets, and W+jets events, at LO accuracy. The POWHEG 2.0 [36–38] generator at next-to-leading order (NLO) precision is used to model the $t\bar{t}$ and single top quark processes. The generators used for signal and background processes are interfaced with PYTHIA 8.230 [39] to simulate parton showering and hadronization. The PYTHIA parameters for the underlying

event description are set with the CP5 tune as described in Ref. [42]. The parton distribution function set NNPDF3.1 [33] is used to produce all simulated samples.

The generation of W+jets and Z+jets processes at LO accuracy is purely due to technical constraints, owing to the large number of simulated events needed to accurately describe W and Z processes. Their cross sections include higher-order QCD and electroweak (EW) differential corrections, as a function of the boson p_{T}, to improve the modeling of high-p_{T} W and Z bosons events [137–141]. The NLO QCD and EW corrections to the cross sections for the Z′ boson signal do not yet exist. The NLO QCD corrections to the Z boson cross section are assumed to be valid for the Z′ boson, within the p_{T} range of this analysis, and are applied to the signal events. However, since the EW couplings of the Z′ could differ from those of the Z boson, the NLO EW corrections are not applied to the signal events.

5.1.2 Online selection

Data events are first selected with the L1 trigger system by requiring a single jet topology or large hadronic deposits. At the HLT level, a combination of several online signatures is required for the trigger selection. All of these signatures require that the total hadronic transverse energy in the event (HT) or the AK8 jet p_{T} exceed a certain threshold, this threshold is around 1050 GeV and 500 GeV, respectively. In addition, soft radiation remnants are removed with the jet trimming technique [103] before the mass selection. This allows the jet p_{T} thresholds to be reduced to around 400 GeV while triggering at a

similar rate, thereby improving the signal acceptance.

Since the online environment and reconstruction are limited as compared to the offline reconstruction, we do not expect the trigger efficiency to be a step function at each of the trigger thresholds. Instead, we define a trigger efficiency that depends both on the jet mass and the jet p_{T} as:

$$\epsilon_{\text{trigger}}(m, p_{\mathrm{T}}) = \frac{N_{\text{pass trigger}}}{N}. \tag{5.2}$$

The efficiency is measured in data events that are triggered by single muon triggers and contain one or more high-p_{T} jets. We use single muon triggers because they have lower thresholds and their efficiencies are close to 1 in this part of the phase-space. We then require events to have at least one jet with $p_{\mathrm{T}} > 180$ GeV and $|\eta| < 2.5$. Figure 5.3 shows the efficiency measured as a function of jet mass and p_{T} for the 2016 data taking year.

In general, the trigger selection efficiency is greater than 95% for events with at least one jet with $|\eta| < 2.5$, jet mass greater than 50 GeV and $p_{\mathrm{T}} >$ 500 GeV. This measurement, however, varies within data taking periods. For example, the trimmed jet mass triggers were not available early in the 2017 data collection, corresponding to the first 4.8 fb^1 of data recorded. This condition motivated the use of a higher p_{T} threshold for the analysis of 2017 data. Table 5.1 shows the different p_{T} lower bounds that determine events for which the trigger selection is greater than 95% efficient.

Figure 5.4 shows the efficiency as a function of jet mass for both AK8 and CA15 jets in 2017 for the given p_{T} thresholds. Jets that are reconstructed offline as the larger CA15 jet also use online triggers that reconstruct the jet with a

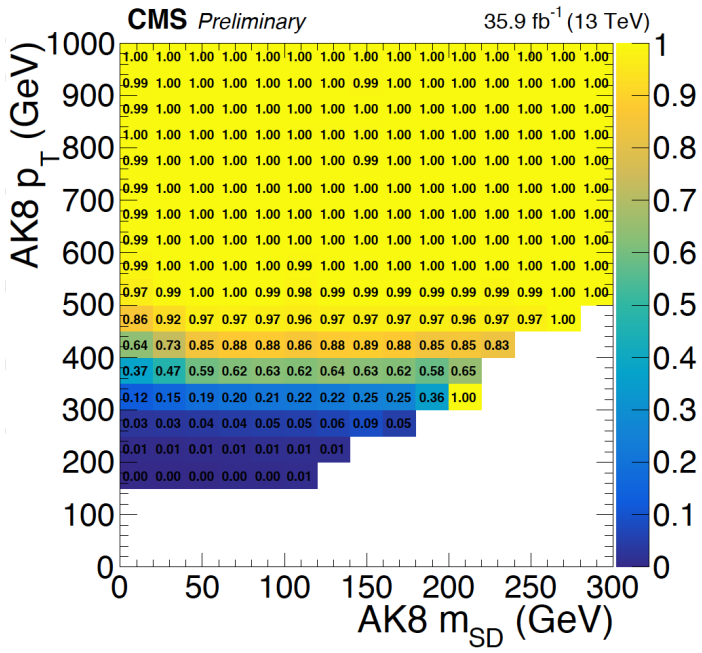

Figure 5.3: Measured trigger efficiency of the combination of single jet and high HT triggers as a function of the AK8 jet p_{T} and jet soft drop mass for the 2016 data taking period. The trigger is above 90% efficient above $m_{\mathrm{SD}} > 50$ GeV and $p_{\mathrm{T}} > 500$ GeV.

Table 5.1: Lower bound thresholds on the jet p_{T} determined by the measured trigger efficiency in each data taking period, and for the different large jets used in this search.

Run-period	AK8	CA15
2016	500 GeV	-
2017	525 GeV	550 GeV

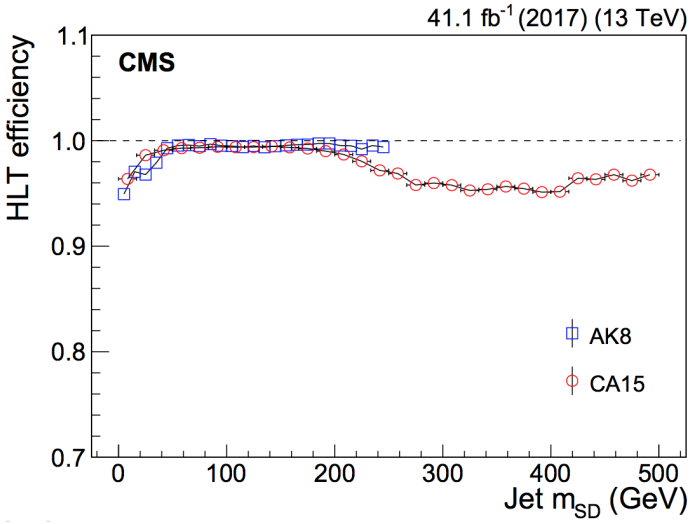

Figure 5.4: High-level trigger efficiency as a function of the soft-drop jet mass (m_{SD}) for AK8 jets with $p_T > 525$ GeV (blue squares) and CA15 jets with $p_T > 575$ GeV (red circles). The trigger selection is $> 95\%$ efficient for 2017 data for both cone sizes and is applied to AK8 jets with masses between 50 and 275 GeV and CA15 jets with masses between 150 and 450 GeV. For jet masses above 200 GeV, the trigger efficiency for the larger CA15 jet decreases slightly. This is due to events for which a reconstructed jet passing the CA15 jet selection does not satisfy the AK8 jet selection at the trigger level.

smaller radius AK8 jet. This leads to a slight decrease in the trigger efficiency for high jet masses. This decrease is due to events for which the jet passes the CA15 jet selection but fails the trigger-level AK8 jet p_T and trimmed mass requirements.

5.1.3 Offline selection

To reduce backgrounds from SM electroweak processes, events are vetoed if they contain isolated electrons, isolated muons, or hadronically decaying τ leptons with $p_T > 10$, 10, or 18 GeV and $|\eta| < 2.5$, 2.4, or 2.3, respectively. For electrons and muons, the isolation criteria require that the pileup-corrected

sum of the p_{T} of charged hadrons and neutral particles surrounding the lepton divided by the lepton p_{T} be less than approximately 15 or 25%, respectively, as described in Sections. 3.6.3 and 3.6.2. Tau leptons, reconstructed by combining information from charged hadrons and π^0 candidates, are required to satisfy the loose working point of a multivariate-based identification discriminant that combines information on the isolation and lifetime of the tau lepton [142].

The remaining offline selection aims to identify the signal jets using jet substructure. As explained in Sec. 4.2.1, jets are identified using the soft-drop algorithm, the p_{T} invariant variable ρ and the decorrelated version of N_2^1: $N_2^{1,\mathrm{DDT}}$. The soft-drop jet mass, that preserves the masses of merged $W/Z/Z'$ jets and reduces the mass of QCD jets, is the main observable of this search. To avoid departure from the p_{T} invariance and the non-perturbative region of QCD, only events with jets in the range $-5.5 < \rho < -2.0$ $(-4.7 < \rho < -1.0)$ are considered for the AK8 (CA15) jets (see Sec. 4.2.3). This results in a p_{T} dependence of the m_{SD} range under study.

Finally, jets are required to have $N_2^{1,\mathrm{DDT}} < 0$. This selection rejects 95% of the multijet background independently of the jet mass and p_{T}. Events failing this requirement, with $N_2^{1,\mathrm{DDT}} > 0$, are used in the background estimate from data described in the next section.

5.2 Background estimate

The background is dominated by QCD multijet events with smaller contributions from $W(q'q)$+jets, $Z(qq)$+jets, and top quark processes. Backgrounds from other EW processes are found to be negligible.

Top quark contribution The contributions from top pair and single top quark production are obtained from simulation. Scale factors correct for:

- the overall top quark background *normalization*, that accounts for the overall agreement for data and MC in a $t\bar{t}$ dominated region; and,
- the $N_2^{1,\mathrm{DDT}}$ *mistag efficiency* for jets originating from top quark decays, which corrects the efficiency as modeled in simulation for a top quark to pass the substructure selection.

These corrections are computed from a dedicated $t\bar{t}$-enriched control region in data, in which an isolated high-p_T muon is required. The control sample consists of events with an energetic muon, a leading AK8 or CA15 jet with $p_\mathrm{T} > 400$ GeV, and an additional $b-$tagged AK4 jet that is separated from the leading large radius jet by $\Delta R > 0.8$. Using the same candidate jet requirements that define the signal selection, $N_2^{1,\mathrm{DDT}}$ pass and fail regions are constructed in both data and simulation. Both the absolute normalization and the $N_2^{1,\mathrm{DDT}}$ efficiency of the $t\bar{t}$ contribution are allowed to vary without prior constraint from the simulation expectation, but are constrained to vary identically in the $t\bar{t}$ control region and the signal region in the simultaneous fit. This allows to constrain in situ the background expectation and $N_2^{1,\mathrm{DDT}}$ mistag

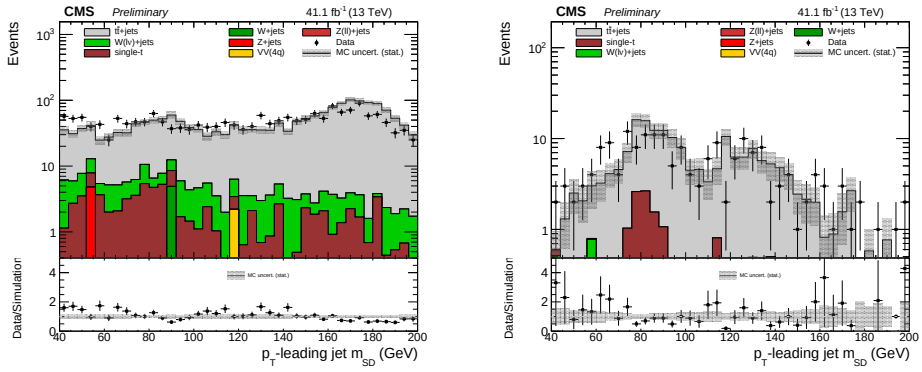

Figure 5.5: Jet mass distributions in the single muon control sample that fail (left) and pass (right) the $N_2^{1,\mathrm{DDT}}$ selection. Due to the relatively low statistics in the control sample, the inclusive event counts are used to constrain the $t\bar{t}$ contribution in the signal region.

efficiency of this process. The data and MC distributions of the jet mass in this control sample are shown in Fig. 5.5.

W/Z + jets contribution The W+jets and Z+jets backgrounds are modeled using simulation. Their cross sections are corrected for NLO QCD and EW effects, following Refs. [137, 139–141].

Multijet contribution The main background in the passing region, QCD multijet production, has a nontrivial jet mass shape that is difficult to model parametrically and depends on the jet p_T. Therefore, we constrain it using a signal-depleted region in data. Because of the decorrelation of $N_2^{1,\mathrm{DDT}}$ from ρ and p_T, the QCD jet mass distributions for events passing and failing the $N_2^{1,\mathrm{DDT}}$ selection exhibit the same smoothly falling shape. This is shown for simulated events in Fig. 5.6.

Given this similarity in shape, we can use the jet mass distribution of

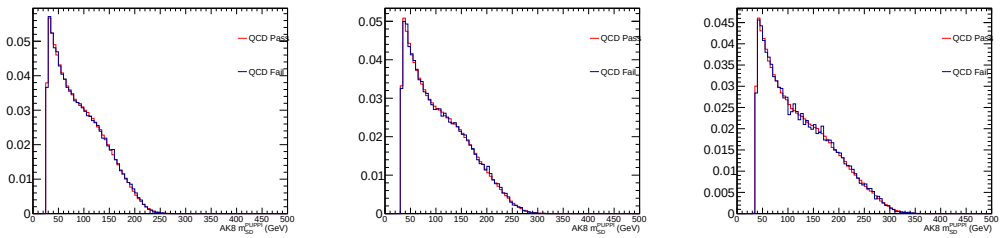

Figure 5.6: Jet mass distributions in the signal region that fail (blue) and pass (red) the $N_2^{1,\mathrm{DDT}}$ selection. The distributions are normalized to unity and are shown for different p_T ranges: [500-600] GeV (left), [600-700] GeV (center), [700-800] GeV (right). Because of the decorrelation procedure, that varies the N_2^1 cut so that the background efficiency is fixed for *any* value of the jet mass and p_T, the shape of the jet mass is the same in the passing and failing region in *simulation*.

events failing the selection to constrain the jet mass distribution of QCD events passing the selection as:

$$n_\mathrm{pass}^\mathrm{QCD} = R_\mathrm{p/f}\, n_\mathrm{fail}^\mathrm{QCD}\,, \tag{5.3}$$

where $n_\mathrm{pass}^\mathrm{QCD}$ and $n_\mathrm{fail}^\mathrm{QCD}$ are the number of passing and failing events in a given m_SD, p_T bin, and $R_\mathrm{p/f}$ is the "pass-to-fail ratio".

The fraction of events, p, passing the $N_2^{1,\mathrm{DDT}}$ selection in simulated QCD multijet events is, by construction, 5% irrespective of ρ and p_T. Therefore, the correction $R_\mathrm{p/f}$ is flat at $\mathrm{p} = 5\%$ and $\mathrm{f} = 95\%$ in the QCD background simulation. To account for residual differences between data and simulation, $R_\mathrm{p/f}$ is allowed to deviate from a constant. This procedure is illustrated schematically in Fig. 5.7.

This deviation is modeled by parametrizing $R_\mathrm{p/f}$ as a function of ρ and p_T

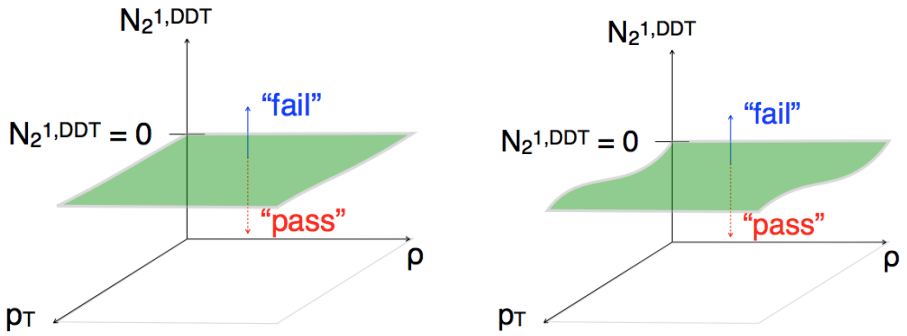

Figure 5.7: A schematic of the background estimation method. The pass-to-fail ratio, $R_{\mathrm{p/f}}$, is defined from the events passing and failing the $N_2^{1,\mathrm{DDT}}$ selection. The variable $N_2^{1,\mathrm{DDT}}$ is constructed so that, for simulated multijet events, $R_{\mathrm{p/f}}$ is constant at $\mathrm{p} = 5\%$ and $\mathrm{f} = 95\%$ (left). To account for residual differences between data and simulation, $R_{\mathrm{p/f}}$ is extracted by performing a two-dimensional fit to data in (ρ, p_{T}) space (right).

and expanding it in a Bernstein polynomial basis of the form:

$$R_{\mathrm{p/f}}(\rho, p_{\mathrm{T}}) = \mathrm{p/f} \sum_{k=0}^{n_\rho} \sum_{\ell=0}^{n_{p_{\mathrm{T}}}} a_{k\ell} b_{\ell,n_{p_{\mathrm{T}}}}(p_{\mathrm{T}}) b_{k,n_\rho}(\rho), \tag{5.4}$$

where $a_{k\ell}$ are the polynomial coefficients, and

$$b_{\nu,n}(x) = \binom{n}{\nu} x^\nu (1-x)^{n-\nu} \tag{5.5}$$

is a polynomial of degree n in the Bernstein basis. The Bernstein basis is chosen over a standard polynomial because, with the variable x bounded between 0 and 1, it is more stable numerically and the function is nonnegative.

With the exception of a_{00}, which is fixed to unity by choice, the coefficients $a_{k\ell}$ and p are unconstrained and determined together with the signal yield from a simultaneous fit to the data events passing and failing the $N_2^{1,\mathrm{DDT}}$ selection. The minimum number of coefficients needed to model the $R_{\mathrm{p/f}}$ shape is determined using a Fisher F-test on data [143]. The test is performed

by iteratively comparing two parametrizations of the $R_{\mathrm{p/f}}$, one with higher polynomial order than the other, and computing the expected change in the log likelihood, i.e. using the goodness-of-fit as the F-statistic. [1] [2]

To determine whether the polynomial order is sufficient, we compare the F-statistic observed in data to that computed from a set of simulated samples generated from the default fit model and fit with the higher order polynomial using the background-only fit. If one provides a significantly better fit (p-value $<5\%$) [3] , we choose that as the new default.

[1] A test statistic is a quantity calculated from our sample of data. Its value can be used to estimate how probable is the result that we observe with respect to some null hypothesis. The goodness-of-fit (GOF) test is used to test the null hypothesis that the data follows a specified distribution, e.g. the distribution given by the background estimate.

The GOF test statistic is given by [22, 144]:

$$-2\log\lambda = 2\sum_{j}^{n_{\mathrm{bins}}} = \left(f_j - d_j + d_j\log(d_j/f_j)\right), \tag{5.6}$$

where d_j is the data in the j-th bin and f_j is the prediction in the j-th bin. It asymptotically follows a chi-squared distribution with degrees of freedom and is an appropriate metric for the GOF of maximum likelihood fits using a Poisson likelihood, such as the one in this chapter.

[2] An F-test is any statistical test in which the test statistic is distributed according to the F-distribution under the null hypothesis. A random variate X of the F-distribution with parameters (d_1, d_2) arises as the ratio of two appropriately scaled chi-squared variates: $X = \frac{U_1/d_1}{U_2/d_2}$, where U_1 and U_2 are independent and follow chi-squared distributions with d_1 and d_2 degrees of freedom respectively.

In our case, we have two fit models, 1 and 2, where model$_1$ has p_1 parameters and is "nested" within model$_2$ with p_2 parameters $(p_2 > p_1)$. The chi-squared variates are determined from the expected changed in the goodness of fit test statistic. Thus, to determine whether model$_2$ gives a significantly better fit to the data, we compute the F-statistic as:

$$F = \frac{-2\log(\lambda_1/\lambda_2)/(p_2 - p_1)}{-2\log\lambda_2/n_{\mathrm{bins}} - p_2)}, \tag{5.7}$$

where n_{bins} is the number of bins, $-2\log\lambda_i$ is the goodness-of-fit test statistic for a model i.

[3] Under the null hypothesis that *model$_2$ does not provide a significantly better fit than model$_1$*, F will have an $F-$distribution with $(p_2 - p_1,\ p_2 - n_{\mathrm{bins}})$ degrees of freedom. The p-value is the area in the tail of this $F-$distribution, i.e. a statement about the evidence against a null hypothesis. We may reject the null hypothesis if F from data is greater than the critical value of the $F-$distribution with probability $\alpha = 0.05$, i.e. p-value $<5\%$.

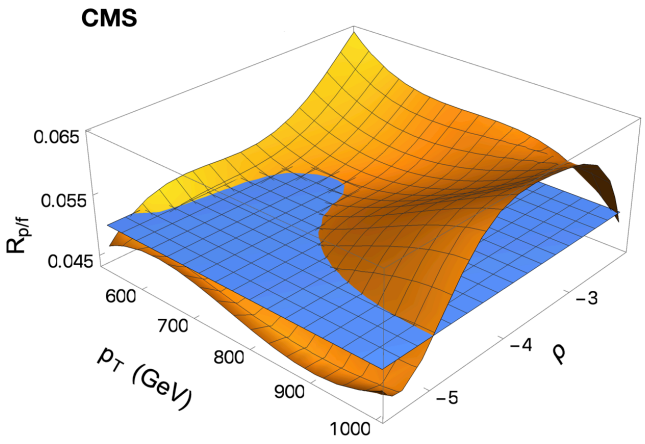

Figure 5.8: Pass-to-fail ratio, $R_{\mathrm{p/f}}(\rho(m_{\mathrm{SD}}, p_{\mathrm{T}}))$, defined from the ratio of events passing and failing the $N_2^{1,\mathrm{DDT}}$ selection. The $R_{\mathrm{p/f}}$ shown in orange is derived for AK8 jets using 41.1 fb^{-1}of data collected in 2017 and corresponds to a polynomial in the Bernstein basis of third order in p_{T} and fifth order in ρ. In simulation, shown in blue, the $R_{\mathrm{p/f}}$ is constant at p $= 5\%$ and f $= 95\%$.

As the magnitude of the data-to-simulation discrepancies can vary among the data samples and their corresponding simulation samples, the F-test is performed independently for each data taking year and jet cone size. The optimal parametrization found for each data sample is the following:

2016 (AK8 jets)	$n_\rho = 4, n_{p_{\mathrm{T}}} = 3$
2017 (AK8 jets)	$n_\rho = 3, n_{p_{\mathrm{T}}} = 5$
2017 (CA15 jets)	$n_\rho = 2, n_{p_{\mathrm{T}}} = 5$

The result of the fit is a slow variation of $R_{\mathrm{p/f}}$ over the m_{SD}–p_{T} plane, with p bounded between 4.5–6.5%. The parametric shape of $R_{\mathrm{p/f}}$ derived from data for the AK8 jet analysis in 2017 is shown in Fig. 5.8.

The fact that $R_{\mathrm{p/f}}$ varies slowly across the m_{SD}–p_{T} domain is essential, since it allows one to estimate the background under a narrow signal resonance based on the events across the whole jet mass range.

5.2.1 Likelihood fit

A binned maximum likelihood fit to the shape of the observed m_{SD} distribution is performed using the sum of the Z′ signal,W,Z,$t\bar{t}$, and multijet contributions. We search for a signal from a Z′ resonance in the mass range from 50 to 450 GeV and take the signal shapes directly from simulation. The fit is performed simultaneously in the passing and failing regions of five (four) p_{T} categories for AK8 (CA15) jets, as well as in the passing and failing components of the $t\bar{t}$-enriched control region. The boundaries of the p_{T} categories are chosen so that approximately the same number of events are used to constrain $R_{\mathrm{p/f}}$ in each p_{T} bin.

The likelihood [4] for the signal extraction can be written as:

$$\begin{aligned}\mathcal{L}(\mathrm{data}|\mu,\boldsymbol{\theta}) = &\prod_{i,j}\mathrm{Poisson}\left(N^{\mathrm{data}}_{\mathrm{fail},i,j}|N^{\mathrm{QCD}}_{\mathrm{fail},i,j}+N^{t\bar{t}}_{\mathrm{fail},i,j}+N^{V}_{\mathrm{fail},i,j}+\mu N^{Z'(qq)}_{\mathrm{fail},i,j}\right)\\ &\times\prod_{i,j}\mathrm{Poisson}\left(N^{\mathrm{data}}_{\mathrm{pass},i,j}|N^{\mathrm{QCD}}_{\mathrm{pass},i,j}+N^{t\bar{t}}_{\mathrm{pass},i,j}+N^{V}_{\mathrm{pass},i,j}+\mu N^{Z'(qq)}_{\mathrm{pass},i,j}\right)\\ &\times\mathcal{L}_{1\mu\ \mathrm{CR}}(\mathrm{data}|\mu,\boldsymbol{\theta})\times\mathrm{Constraint}(\boldsymbol{\theta}|\bar{\boldsymbol{\theta}},\delta\boldsymbol{\theta})\ ,\end{aligned}\tag{5.8}$$

where:

[4] A function $f(x)$ can describe the probability density for the observable x for a single event. The probability density function refers to the value of f as a function of x, a data point. The likelihood function refers to the value of f as a function of the parameters $\mu,\boldsymbol{\theta}$ given a fixed value of x. A maximum likelihood estimate of the parameters $\mu,\boldsymbol{\theta}$ are the values for which $\mathcal{L}(\mathrm{data}|\mu,\boldsymbol{\theta})$ has a global maximum.

- $N^{\text{data}}_{\text{fail},i,j}$ ($N^{\text{data}}_{\text{pass},i,j}$) is the observed number of events in the i^{th} m_{SD} bin and j^{th} p_{T} bin of the failing (passing) region,
- $N^{\text{process}}_{\text{fail},i,j}$ ($N^{\text{process}}_{\text{pass},i,j}$) is the expected number of events for each background or signal process in the i^{th} m_{SD} bin and j^{th} p_{T} bin of the failing (passing) region,
- μ is the Z′ signal strength modifier, [5]
- $\text{Poisson}(x|y) = y^x e^{-y}/x!$,
- $\mathcal{L}_{1\mu\ \text{CR}}(\text{data}|\mu, \boldsymbol{\theta})$ is the likelihood for the single muon control region, which takes on a very similar Poisson counting experiment form, and
- $\text{Constraint}(\boldsymbol{\theta}|\bar{\boldsymbol{\theta}}, \delta\boldsymbol{\theta})$ is the constraint term for the nuisance parameters.

Nuisance parameters, denoted by $\boldsymbol{\theta}$, are parameters whose values are not taken as known a priori but rather must be fitted from the data. The nuisance parameters in the last term model the effects of systematic uncertainties. These are constrained in the fit and detailed in the next section. However, note that some of the nuisance parameters do not have an additional constraint in the fit, namely: the multijet QCD yields in the failing category $N^{\text{QCD}}_{\text{fail},i,j}$, the Bernstein coefficients of the QCD transfer factor $a_{k,\ell}$, and the $t\bar{t}$-related scale factors.

[5] In order to test a signal model, we assume a signal plus background hypothesis (H_1). This test hypothesis includes the new physical process (e.g. the Z′ signal). The null hypothesis (H_0) describes the known physical process, i.e. the SM background only hypothesis.
The signal strength modifier is defined as the ratio:

$$\mu = \frac{\sigma_{\text{obs}}}{\sigma_{\text{SM}}}$$

and acts as a multiplicative factor on the signal cross-section appearing in the signal-plus-background hypothesis, i.e that assuming SM and the Z′ signal. The null hypothesis corresponds to the case where $\mu = 0$, and the alternate signal-plus-background hypothesis corresponds to $\mu = 1$.

5.3 Systematic Uncertainties

The dominant uncertainty in this analysis is the uncertainty in the fit for $R_{\mathrm{p/f}}$, as described in Eq. 5.4 (1–3%), arising from the parameters $a_{k\ell}$, and the statistical uncertainty on the data in the $N_2^{1,\mathrm{DDT}} < 0$ region.

The systematic uncertainties in the shapes and normalization of the W and Z boson backgrounds and the signal are correlated since they are affected by similar systematic effects. The uncertainties in the jet mass scale and resolution, and the $N_2^{1,\mathrm{DDT}}$ selection efficiency, are estimated using an independent sample of merged W boson jets in semileptonic $t\bar{t}$ events in data. The derivation of these scale factors and uncertainties was described in Sec. 4.2.6. This efficiency of the $N_2^{1,\mathrm{DDT}} < 0$ requirement is used to correct overall yields for resonant backgrounds obtained from simulation in the signal region. The uncertainty on this scale factor is in the order of 6-10%. The jet mass resolution has an uncertainty of 10% and jet mass scale uncertainty is below 1%. The variation of the jet mass scale with jet p_{T} is also studied using large cone size jets, where all the decay products of the top quark are contained in a single jet. By performing simultaneous fits to data and simulation of this peak binned in p_{T}, a small (1%) variation in jet mass scale is observed and applied in the fit as an additional $p_{\mathrm{T}}-$dependent nuisance parameter. These scale factors determine the initial shape and normalization of the jet mass distribution for the W, Z boson, and signal but they are further constrained in the fit to data because of the presence of the W and Z resonances in the jet mass distribution.

To account for potential deviations due to missing higher-order corrections, uncertainties are applied to the W and Z boson yields. These uncertainties

increase with the jet p_{T} and are correlated per p_{T} bin. An additional systematic uncertainty is included to account for potential differences between the W and Z boson higher-order corrections (NLO EW W/Z decorrelation). The uncertainties associated with the modeling of the Z′ boson p_{T} spectrum when considering extra jets in the generation and similar NLO QCD corrections to the Z boson are propagated to the overall normalization of the Z′ signal. Finally, uncertainties associated with the jet energy resolution [101], trigger efficiency (2%), variations in the amount of pileup (1–2%) and the integrated luminosity determination (2-2.3%) [145] are also applied to the W, Z, and Z′ boson signal yields.

A quantitative summary of the systematic effects considered for signal and W/Z boson background processes is given in Table 5.2.

5.3.1 Fit validation

In order to validate the robustness of the fit and its associated systematic uncertainties, we perform a goodness-of-fit test and signal injection studies on background-only fits that estimate the possible bias on the background estimate due to the presence of a signal. We generate pseudo-experiments, with and without the injection of simulated signal, and then fit with the signal plus background model, for different values of the Z′ boson mass. No significant bias in the fitted signal strength is observed; specifically, the means of the differences between the fitted and injected signal strengths divided by the fitted uncertainty are found to be less than 15%.

As a further test of the $R_{\mathrm{p/f}}$ fit robustness, we split the subset of events

Table 5.2: Summary of the uncertainties for signal (Z′), W/Z boson and multijet background processes. The reported ranges denote a variation of the uncertainty across p_T bins, from 500/525 to 1500 GeV (AK8 jets) and from 575 to 1500 GeV (CA15 jets). The symbol $^{\triangle}$ denotes uncorrelated uncertainties for each p_T bin. For the uncertainties related to the jet mass scale and resolution, the reported percentage reflects a one standard deviation effect on the nominal jet mass shape. Three dots (—) indicates that the uncertainty does not apply.

Uncertainty source	Sample	Systematic Uncertainty		
		AK8 2016	AK8 2017	CA15 2017
Statistical				
QCD pass-fail ratio (data correction)	Multijet	1–3%	1–4%	1–4%
$t\bar{t}$ normalization and misidentification	$t\bar{t}$	2–10%	2–8%	2–8%
Systematic				
Simulation sample size	W/Z/Z′/$t\bar{t}$	—	1–12%	1–12%
$N_2^{1,\mathrm{DDT}}$ selection efficiency	W/Z/Z′	9%	10%	7%
Jet mass scale	W/Z/Z′	0.5%	1%	1%
Jet mass resolution	W/Z/Z′	10%	10%	7%
Jet mass scale (% / (p_T [GeV]/100)) $^{\triangle}$	W/Z/Z′	0.5–2%	0.5–2%	0.5–2%
Jet energy scale and resolution	W/Z/Z′	10%	1–7%	1–7%
Other experimental uncertainties	W/Z/Z′	0.5–2%	0.5–2%	0.5–2%
Theoretical				
NLO EW corrections$^{\triangle}$	W/Z	15–35%	15–35%	15–35%
NLO QCD corrections	W/Z/Z′	10%	10%	10%
NLO EW W/Z decorrelation$^{\triangle}$	W/Z	5–15%	5–15%	5–15%

failing the $N_2^{1,\mathrm{DDT}}$ selection into two smaller subsets mimicking the passing and failing selection in the data fit. The mimicked passing-like events also reject 95% of the QCD background events in the failing region. We repeat our background estimation procedure on this selection and use the coefficients $a_{k\ell}$ from this fit to generate pseudo-experiments. We then fit the data with the signal plus background model and find the biases in the fitted signal strength to be negligible.

5.4 Results

The number of observed events is consistent with the predicted background from SM processes. Figures 5.9, 5.10 and 5.11 show the m_{SD} distributions for data in the passing region with the fitted SM background for the different jets and data taking periods. The number of observed events is consistent with the predicted background from SM processes. For AK8 jets, the W and Z boson contributions are clearly visible as a merged peak in the data, while for CA15 jets, due to the ρ selection and increased QCD background, the W/Z contributions are only visible in the lower p_{T} categories.

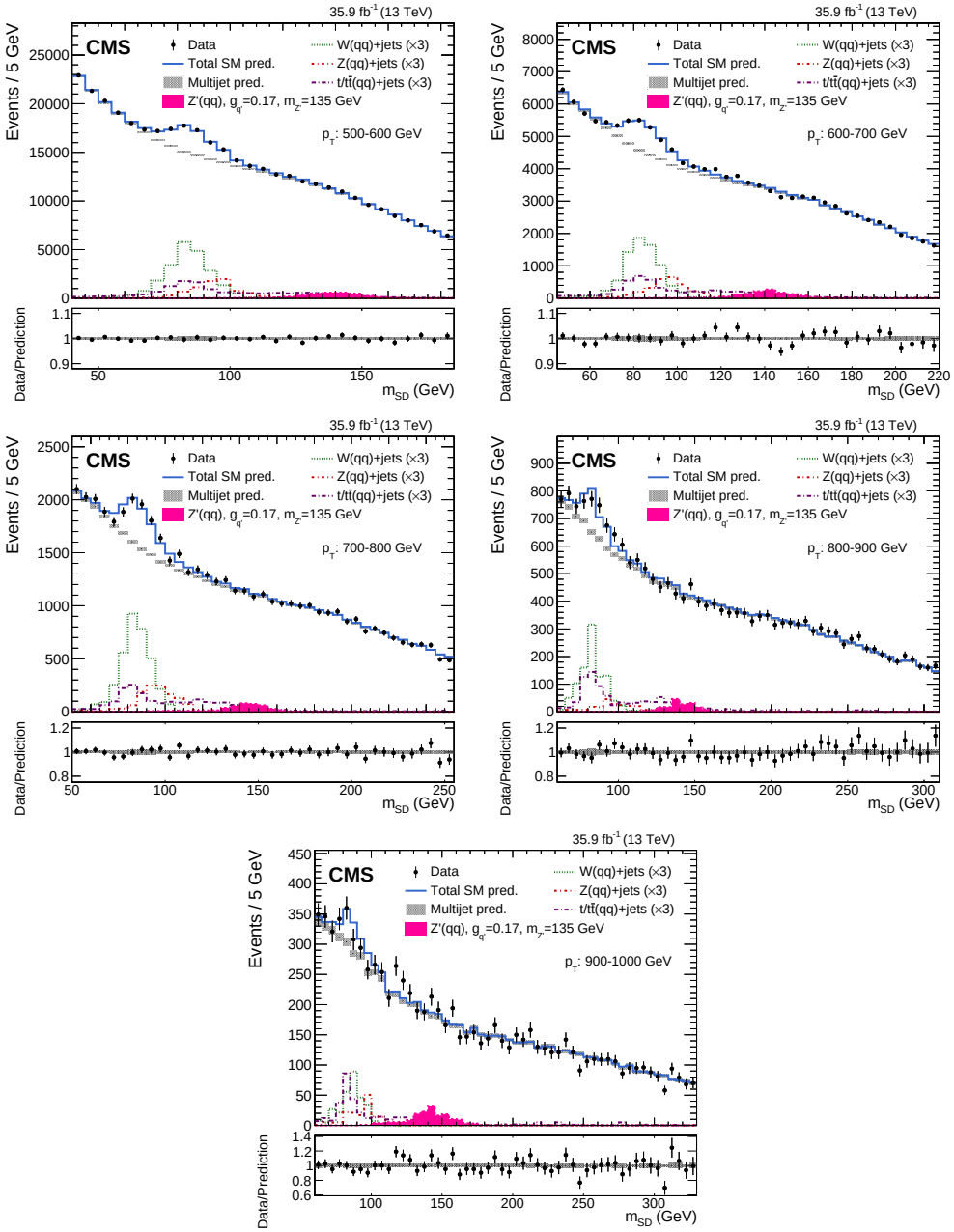

Figure 5.9: Jet m_{SD} distribution for each p_{T} category of the fit in the 2016 AK8 jet data sample. Data are shown by the black points. The multijet background prediction, including uncertainties, is shown by the shaded bands. Contributions from the W and Z bosons, and top quark background processes are shown as well. A hypothetical Z' boson signal with a mass of 135 GeV is also indicated. In the bottom panel, the ratio of the data to its statistical uncertainty, after subtracting the non-resonant backgrounds, is shown. The scale on the x-axis differs for each p_{T} range due to the kinematic selection on ρ.

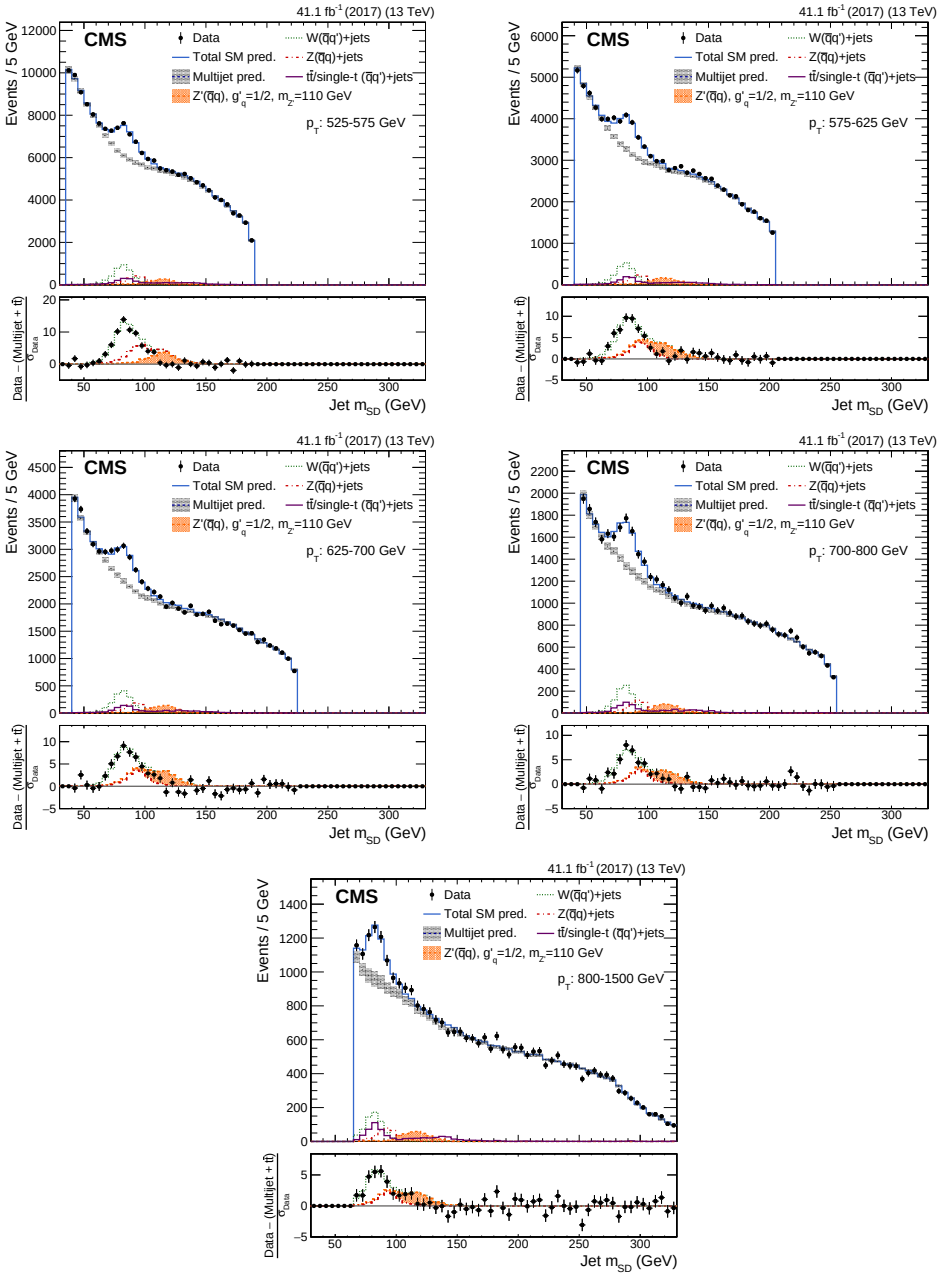

Figure 5.10: Jet m_{SD} distribution for each p_{T} category of the fit in the 2017 AK8 jet data sample. Data are shown by the black points. The multijet background prediction, including uncertainties, is shown by the shaded bands. Contributions from the W and Z bosons, and top quark background processes are shown as well. A hypothetical Z' boson signal with a mass of 110 GeV is also indicated. In the bottom panel, the ratio of the data to its statistical uncertainty, after subtracting the non-resonant backgrounds, is shown.

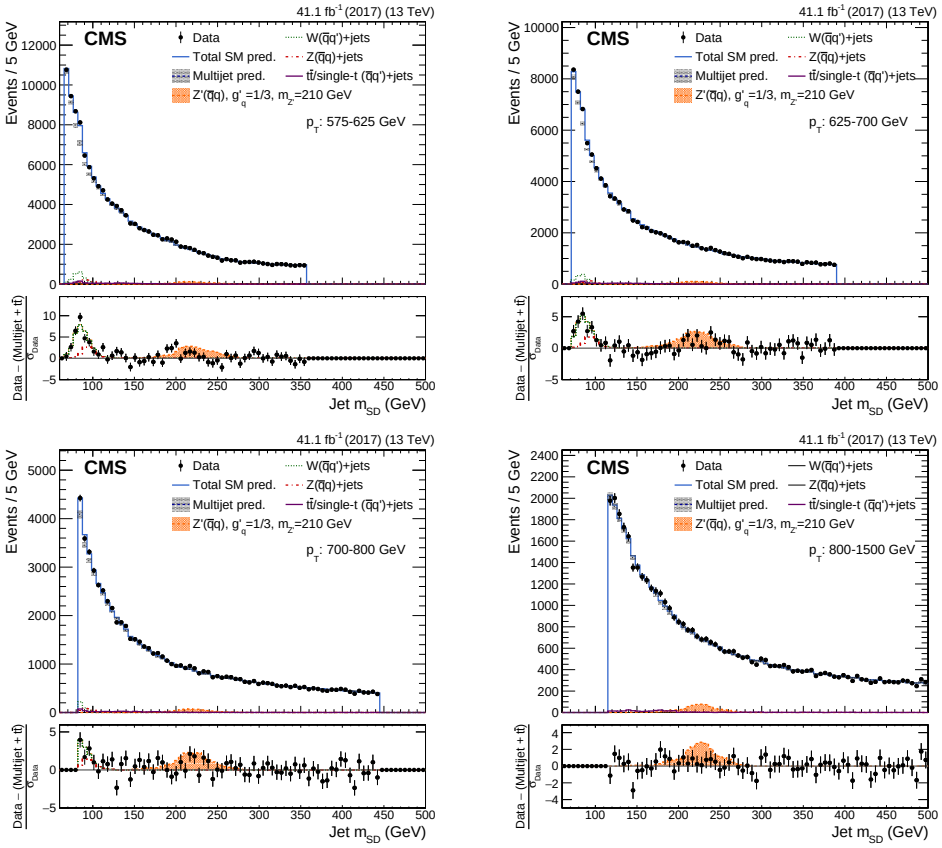

Figure 5.11: Jet m_{SD} distribution for each p_{T} category of the fit in the 2017 CA15 jet data sample, from 575 to 1500 GeV. Data are shown by the black points. The multijet background prediction, including uncertainties, is shown by the shaded bands. Smaller contributions from the W and Z bosons, and top quark background processes are shown as well. A hypothetical Z' boson signal with a mass of 210 GeV is also indicated. In the bottom panel, the ratio of the data to its statistical uncertainty, after subtracting the non-resonant backgrounds, is shown.

5.4.1 Limits on Z′ boson coupling to quarks

In the absence of an excess compatible with our signal hypothesis, we use the results of the fit to set 95% confidence level (CL) upper limits of the Z′ boson coupling to quarks g'_{q}.

Upper limits are computed using the modified frequentist approach for CL, taking the profile likelihood ratio [6] as the test statistic [146, 147] in the asymptotic approximation [148]. [7]

[6] The profile likelihood ratio test statistic q_μ is defined for an analysis-specific likelihood $\mathcal{L}$:

$$q_\mu = -2 \ln \frac{\mathcal{L}(\mathrm{d}|\mu, \hat{\hat{\boldsymbol{\theta}}})}{\mathcal{L}(\mathrm{d}|\hat{\mu}, \hat{\boldsymbol{\theta}})}.$$

In the numerator, $\hat{\boldsymbol{\theta}}$ are the values that maximize $\mathcal{L}$ for the assumed μ. The denominator is maximized in an unconstrained way, i.e. the full set of parameters $\hat{\mu}, \hat{\boldsymbol{\theta}}$ are the true maximum likelihood estimators. The likelihood ratio in Equation 6 is used to measure the compatibility of the data with an non-negative signal strength smaller than the hypothesis μ.

[7] The exclusion of a signal hypothesis is an important statement in the search for new physics. Once a signal is excluded, the exclusion regime is no longer considered to be important to be searched for. Therefore, in order to reduce the likelihood of excluding signal hypotheses that a search is not a-priori sensitive to, the $\mathrm{CL_s}$ statistic is used in the LHC experiments.

The $\mathrm{CL_s}$ statistic is defined as the ratio of the confidence levels of the signal-plus-background hypothesis, CL_{s+b}, to the background only hypothesis $\mathrm{CL}_b(\mu)$:

$$\mathrm{CL_s}(\mu) = \frac{\mathrm{CL}_{s+b}(\mu)}{\mathrm{CL}_b(\mu)}.$$

The confidence levels CL_{s+b} and CL_b are the probabilities to obtain a value of the test statistic q_μ, as low or lower than the observed value q_μ^{obs}, under the corresponding hypothesis:

$$\mathrm{CL}_{s+b}(\mu) = P_1(q_\mu < q_\mu^{\mathrm{obs}}|H_1) = \int_{q_\mu^{\mathrm{obs}}}^{\infty} \mathrm{d}q_\mu \, f(q_\mu|\mu) = p_\mu,$$

$$\mathrm{CL}_b(\mu) = P_0(q_\mu < q_\mu^{\mathrm{obs}}|H_1) = \int_{q_\mu^{\mathrm{obs}}}^{\infty} \mathrm{d}q_\mu \, f(q_\mu|\mu = 0) = p_0$$

Here:

- p_μ and p_0: are the p-values for each hypothesis, and,
- $f(q_\mu|\mu)$ is the probability distribution function.

Limits on g'_q as a function of the Z′ boson mass are shown in Fig. 5.12. The results are shown for the independent data samples of each running period 2016 and 2017. Coupling values above the solid curves are excluded at the 95% CL.

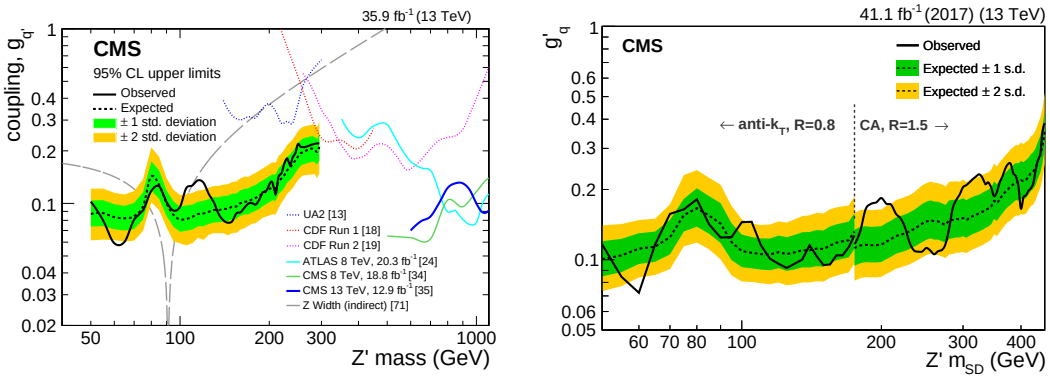

Figure 5.12: Upper limits at 95% C.L. on the coupling g'_q as a function of the resonance mass for a leptophobic Z′ boson that couples only to quarks. The limits based on the 2016 analysis are shown in the left and those based on the 2017 analysis are shown in the right. The observed limits (solid), expected limits (dashed), and their variation at the 1 and 2 standard deviation levels (shaded bands) are shown. In the right plot, the vertical line at 175 GeV corresponds to the transition between the AK8 and CA15 jet selections.

We summarize the results of this search in the mass vs. coupling plane in Fig. 5.13. For masses between 50 and 220 GeV, the most restrictive limits for this search are obtained from the statistical combination of the upper limits set by the 2016 and 2017 data sets using AK8 jets. For the mass range between 175 and 220 GeV, this combination is as sensitive as that obtained from the limits set by the 2016 AK8 jet and 2017 CA15 jet searches. The limits correspond

A given signal hypothesis with $\mu = 1$ is considered excluded at 95% CL when $CL_s \leq 0.05$. This prescription for exclusion, $CL_s \leq \alpha$, is generally a stronger requirement than the standard prescription, $p_\mu \leq \alpha$. This metric is also used to compute upper limits. These measure the amount of excluded region in the theory's parameter space resulting from the negative results of a search for a new signal.

to a total integrated luminosity of 77.0fb^{-1}. For higher masses, between 220 and 450 GeV, the most stringent limits come from the analysis of 2017 data using CA15 jets, corresponding to an integrated luminosity of 41.1fb^{-1}. The sensitivity is driven by the multijet background uncertainty on the parametric fit of $R_{\mathrm{p/f}}$, which is modeled with different polynomial orders for the 2016 and 2017 data sets. In 2017 a loss of sensitivity of 20%, relative to the results set by the 2016 search [12], is observed, due to the higher p_{T} threshold determined by the trigger turn-on for the 2017 data set.

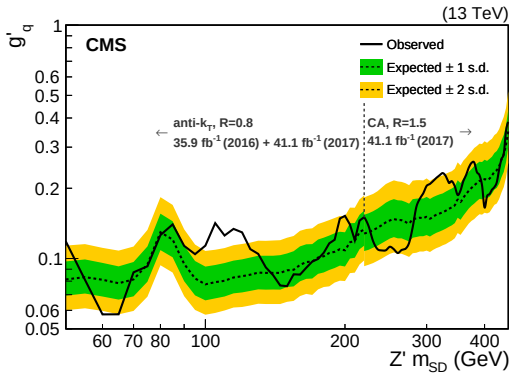

Figure 5.13: Upper limits at 95% CL on the coupling g'_{q} as a function of the resonance mass for a leptophobic Z′ boson that couples only to quarks. The limits are based on the statistical combination of 2016 and 2017 results. The observed limits (solid), expected limits (dashed), and their variation at the 1 and 2 standard deviation levels (shaded bands) are shown. For masses between 50 and 220 GeV the limits correspond to a Z′ boson reconstructed in AK8 jets using 77.0fb^{-1}of statistically combined data from 2016 and 2017. For masses above 220 up to 450 GeV, the results correspond to a Z′ resonance reconstructed in CA15 jets using 41.1fb^{-1}of data collected in 2017.

A local excess in the observed limit over the expected limit, corresponding to 2.9 standard deviations, was observed at a Z′ mass hypothesis near 115 GeV in the 2016 analysis with 35.9fb^{-1}of integrated luminosity. This excess is not confirmed by the 2017 analysis, where the local observed p-value for

a Z′ boson mass of 115 GeV is 0.5 and the data agrees with the prediction. The combined observed limit with the full 2016 and 2017 dataset at a Z′ mass hypothesis of 115 GeV in Fig. 5.13, corresponds to 2.2 standard deviations from the background-only expectation.

5.4.2 Sensitivity with respect to other results

Figure 5.14 shows the limits on g'_{q} for the current search and other dijet searches performed at the UA2, CDF/D0, ATLAS and CMS experiments. In the mass range between 50 and 300 GeV this analysis places the most sensitive limits to date. Above 300 GeV the most sensitive limits are set by the searches for dijet resonances in the non-boosted regime produced in association with a jet [8] or with a photon [149]. The sensitivity of the CA15 jet analysis is lower than that reached with the non-boosted dijet searches due to the lack of a dedicated CA15 jet trigger-level selection.

5.4.3 Translation to dark matter constraints

The results of this analysis can be used to constrain simplified models of DM. Figure 5.15 shows the excluded values at 95% CL of mediator mass (m_{Med}) as a function of the dark matter particle mass (m_{DM}) for vector mediators, in simplified models that assume a leptophobic mediator that couples only to quarks and DM particles [1, 2]. Limits are shown for a choice of universal quark coupling $g'_{\mathrm{q}} = 0.25$ and a DM coupling $g_{\mathrm{DM}} = 1.0$. The difference in limits between axial-vector and vector mediator couplings is small and thus only constraints for the latter coupling scenario are shown.

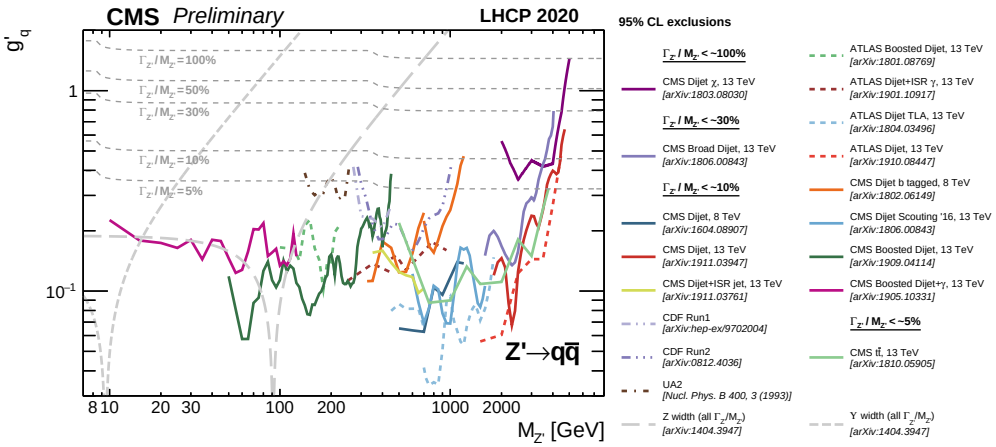

Figure 5.14: Upper limits on the universal coupling g'_q between a leptophobic Z′ boson and quarks for various dijet analyses from UA2, CDF/D0, ATLAS and CMS. The limits are shown in solid lines, with the excluded area above the lines. The hashed areas show the direction of the excluded area from the observed limits. The grey dashed lines show the g'_q values at fixed values of Γ'_Z/M'_Z. Most of the analyses, with the exception of Dijet χ and Broad Dijet, assume that the intrinsic width is negligible compared to the experimental resolution, and hence are valid for $\Gamma'_Z/M'_Z \leq 10\%$. The results of the search presented in this Chapter are shown in green (CMS Boosted Dijet). Also shown are indirect constraints on g'_q from the Y and Z boson widths, which are valid for all values of Γ'_Z/M_Z. This figure is reproduced from Ref. [136].

The excluded range of mediator mass (light blue) is between 50 and 450 GeV. In Fig. 5.15 the exclusion range above 300 GeV is superimposed with results from searches in the non-boosted regime produced in association with a jet (light red) [8]. The upper bound decreases to 240 GeV when $m_{\mathrm{Med}} > 2m_{\mathrm{DM}}$, because the branching fraction (BR) to qq decreases as the BR to DM becomes kinematically favorable. If $m_{\mathrm{Med}} < 2m_{\mathrm{DM}}$, the mediator cannot decay to DM particles and the dijet cross section from the mediator model becomes identical to that in the leptophobic Z′ model, meaning that the limits on the mediator mass in Fig. 5.15 are identical to the limits on the Z′ mass with a coupling $g_{\mathrm{q}'} = g_{\mathrm{q}} = 0.25$. For axial-vector mediators, the excluded values of mediator mass are expected to be identical to the excluded values in Fig. 5.15 when $m_{\mathrm{DM}} > m_{\mathrm{Med}}/2$ or $m_{DM} = 0$, with differences only expected in the transition region $m_{\mathrm{Med}} \simeq 2m_{\mathrm{DM}}$. Additional limits (light violet, yellow) in Fig. 5.15 come from traditional dijet searches [6].

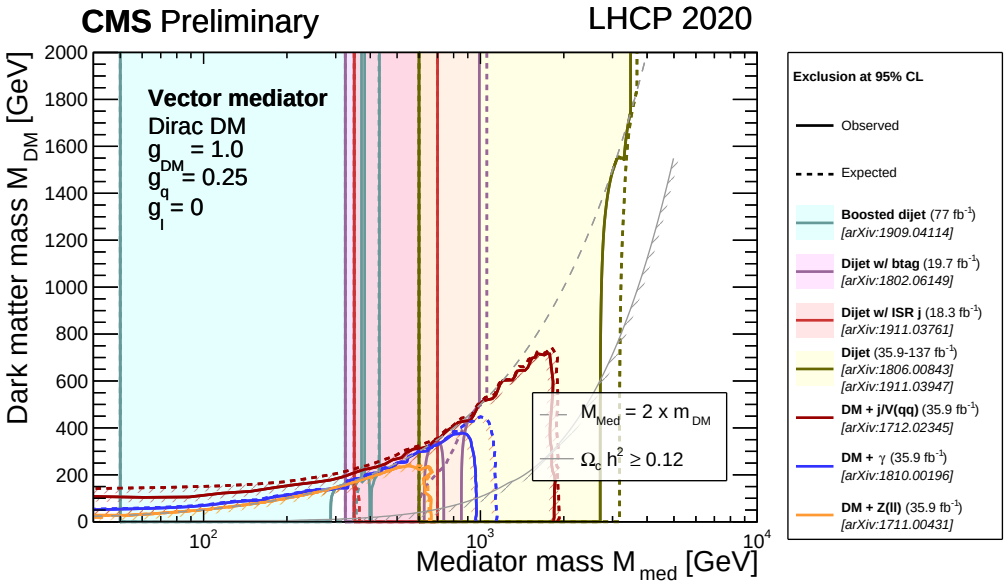

Figure 5.15: The 95% CL observed (solid) and expected (dashed) excluded regions in the plane of dark matter particle mass (m_{DM}) vs. mediator mass (m_{Med}), for vector mediators. The results are shown for dijet searches and different $p_{\mathrm{T}}^{\mathrm{miss}}$ based DM searches from CMS. A branching fraction of 100% is assumed for a leptophobic vector mediator decaying to dijets. The exclusion is computed for a quark coupling choice $g_q = 0.25$ and for a dark matter coupling $g_{\mathrm{DM}} = 1$. The results of the search presented in this chapter are shown in light blue. The excluded regions from the dijet resolved analysis (Dijet + ISR) using early 2016 data [8] are also shown. Results are compared to constraints from the cosmological relic density of DM (light gray) determined from astrophysical measurements [150, 151] and MADDM version 2.0.6 [152, 153] as described in Ref. [154]. It should also be noted that the absolute exclusion of the different searches as well as their relative importance, will strongly depend on the chosen coupling and model scenario. Therefore, the exclusion regions, relic density contours, and unitarity curve shown in this plot are not applicable to other choices of coupling values or model. This figure is reproduced from Ref. [136].

Chapter 6

Triggering on boosted Higgs bosons

The SM Higgs boson (H) is responsible for the electroweak symmetry breaking and the mechanism from which elementary particles acquire mass. Following the discovery of the 125.1 GeV Higgs-like particle at the LHC [44, 45], its production has now been observed in all of its production modes and several decay modes. One of the most important goals now is to precisely understand all of its properties and its couplings to other particles. Any deviation from the SM predictions would reveal the existence of new physics in the Higgs sector.

Recently, there has been considerable interest in the measurement of Higgs bosons produced with high-p_{T}, where traditional channels have limited sensitivity due to small production rates or branching fractions. This is because new physics contributions could modify the tail of the kinematic spectrum of the Higgs particle and give direct hints of new couplings to the Higgs sector [155–158].

Due to their large branching fraction[1], the hadronic decays of the Higgs boson can provide a starting point for high-p_{T} H measurements. However, the search for these decays in a hadron collider faces the same challenge as the search presented in Chapter 5. It is impossible to trigger events in data because of the overwhelming multijet SM background at low jet mass. Therefore, one is led to consider the production of high-p_{T} Higgs bosons in association with a jet. As with the Z'+jet signature, the extra radiation ensures that the event meets the restrictive trigger criteria. In this highly energetic regime, hadronic decays of the Higgs boson can be reconstructed as single-large radius jets, with the jet mass consistent with that of the observed Higgs boson [159].

The search for high-p_{T} $H(\bar{b}b)$ events by the CMS Collaboration has already demonstrated the experimental sensitivity of this hadronic channel [160, 161]. The latest result, using the full Run-2 integrated luminosity, observes an excess of events above the background assuming no Higgs boson production with a significance of 2.5 standard deviations, while the expectation is 0.7. It is worth noting that the analysis strategy is similar to that presented in Sec. 5.1:

- it relies on the identification of the Higgs boson candidates reconstructed as large-radius jets,
- the dominant multijet background contribution is estimated from data by inverting the jet tagging requirement, which is designed to have reduced correlation with jet mass and p_{T}, and,
- the Higgs boson signal can be observed as a resonant peak on top of a

[1]The branching fraction is the ratio of particles which decay by an individual decay mode with respect to the total number of particles which decay. For example, $H \to \bar{b}b$ decays have a branching ratio of $\mathcal{B} = 58.1\%$.

smoothly falling jet mass distribution.

The main difference, with respect to the search for $Z'(qq)$ decays, is that, to further distinguish the $H(\bar{b}b)$ candidates, the two-pronged jet is required to have displaced tracks and decay vertices consistent with a merged $H(\bar{b}b)$ decay, as briefly described in Sec. 4.2.4.

In view of these observations, we can take advantage of this analysis strategy, which enables triggering on hadronically decaying Higgs bosons, to expand the boosted Higgs program. In the following chapter, we explore the possibility of using cross section measurements of boosted Higgs bosons to jets to provide a constraint on a fundamental property of the Higgs: its total width.

6.1 The SM Higgs width

The H boson is produced at a hadron collider via the interactions between the quarks or the gluons in the collisions. The dominant modes are through gluon fusion (ggH), weak vector boson fusion (VBF), associated production with a weak vector boson (VH), and the associated production with two top or bottom quarks ($\bar{t}tH$ and $\bar{b}bH$ respectively). The leading order diagrams of these four important production mechanisms are given in Fig 6.1.

The cross section of each production mode will depend on the partonic contribution of quarks and gluons, the center-of-mass energy and the Higgs boson mass. Fig. 6.2 shows the expected production cross sections and the expected decay mode branching ratios as a function of Higgs boson mass for

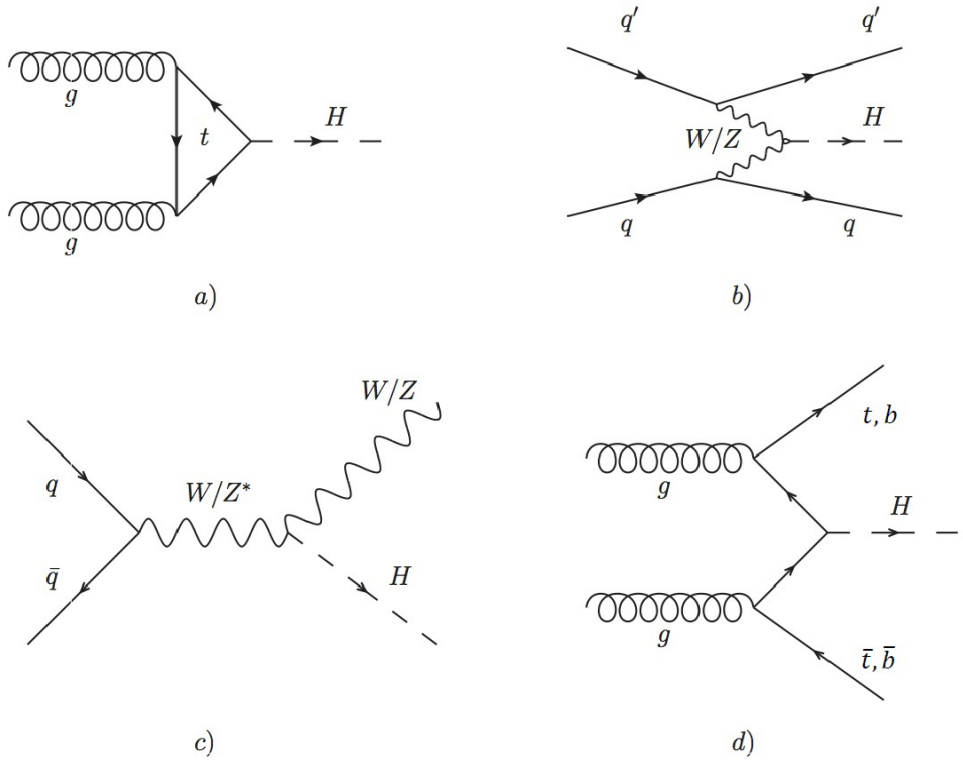

Figure 6.1: Main leading order diagrams contributing to the Higgs boson production at the LHC, shown for (a) ggH, (b) VBF, (c) WH or ZH, and (d) $\bar{t}tH$ and $\bar{b}bH$.

$\sqrt{s} = 13$ TeV. The most abundant production channel at the LHC for a SM 125 GeV Higgs boson is via gluon fusion, while the most probable decay is through $H \to \bar{b}b$ decays.

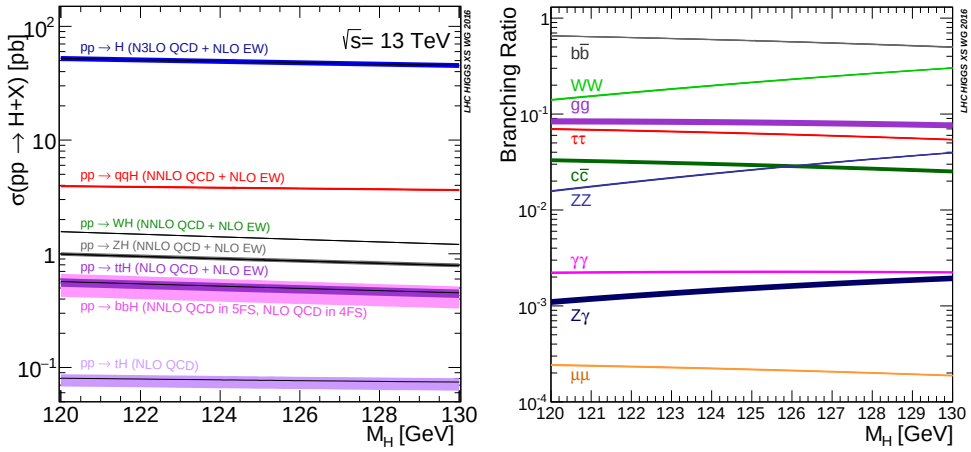

Figure 6.2: Minimal SM Higgs production and decay modes at the LHC [162]. In the left panel, the production cross sections at $\sqrt{s} = 13$ TeV are shown for $m_H = 120$–130 GeV. The right panel shows the decay Branching Fractions for $m_H = 120$–130 GeV.

The total width [2] of the SM Higgs boson for a mass of 125.1 GeV is predicted to be $\Gamma_{SM} = 4.2$ MeV. Measuring this property is of intrinsic interest because a sizable deviation would directly indicate new physics. However, performing this measurement is an experimental challenge because the H width is very narrow compared to its mass.

[2] The total width Γ of a particle is directly related to its decay lifetime: a large decay width leads to a short lifetime. The total width can be found by summing the particle's partial widths. The partial width of a given decay channel is nothing but the product of Γ and the corresponding branching ratios.

6.1.1 Constraints in hadron colliders

At the LHC, the Higgs boson width has been measured using two different approaches.

The first is through the direct measurement of the Higgs mass line-shape using the resonant Higgs decays to diphoton [163] and four lepton final states ($H \to ZZ \to \ell\ell\ell\ell$) [164]. While this approach is directly sensitive to the Higgs boson width, it is heavily limited by systematic uncertainties from lepton and photon detector resolution. The current precision on Γ_H using this approach is 1.1 GeV, equivalent to $270 \times \Gamma_{SM}$ [164].

The second approach involves the comparison of the on- and off-shell [3] Higgs boson cross sections. In particular, it uses the interference of gluon fusion production of the Higgs boson with gluon fusion production of di-bosons. An interesting feature of this interference is that, while the amount of Higgs-boson signal observed on the 125 GeV on-shell peak depends on the Higgs boson width, the far off-shell signal is independent of it. This results on a modification of the high mass distribution of the diboson mass spectrum, which yields a constraint on the width[4] . The extraction of the Higgs width

[3] Particles are said to be "on the mass shell", or simply "on-shell" if their behavior satisfies Einstein's relationship between energy and momentum: $E^2 = p^2 + m^2$. Virtual particles are those that do not satisfy this relationship and thus do not have the right mass, they are said to be "off-shell". A virtual Higgs boson can be produced "off-shell", at much larger mass than 125 GeV.

[4] This is plainly seen from the Breit-Wigner expression for a resonance with mass M [22]:

$$\sigma(s) \sim \frac{\Gamma^2 M^2}{(s - M^2)^2 + M^2\Gamma^2} \tag{6.1}$$

which (for a narrow resonance $\Gamma / M \ll 1$) implies that the on-shell cross section is independent of the width.

in this manner requires a knowledge of the interference pattern, which intrinsically implies a SM-like behavior of the product of the couplings g_{ggH} and g_{VVH} across a large mass range. Further details of this approach and its model-dependent limitations are discussed extensively in the literature [165–177].

This approach has been applied using several diboson final states; most recently, the four lepton final state [178] [5] . The current best measurement for the total width using this approach is $\Gamma_H = 3.2^{+2.8}_{-2.2}$ MeV, while the expected constraint based on simulation is $\Gamma_H = 4.1^{+5.0}_{-4.0}$ MeV [178]. Projections for measuring the Higgs width in the four lepton channel alone at the High Luminosity LHC (HL-LHC), with assumptions similar to the ones mentioned above, suggest that, with 100 times more data, Γ_H can be constrained with a precision of $\Gamma_H = 4.2^{+1.5}_{-2.1}$ MeV (ATLAS) and $\Gamma_H = 4.1^{+1.0}_{-1.1}$ MeV (CMS) [179].

6.1.2 Constraints in lepton colliders

Model-independent measurements of the Higgs boson total width are possible through the use of lepton-colliders. With a muon collider, the width can be probed through the direct production of $\mu^+\mu^- \rightarrow h$ by a precise scan of the center of mass energy about the Higgs boson total mass [180]. With an electron–positron collider, the Higgs boson width can be measured through the Higgs boson recoil approach whereby one measures the inclusive Higgs

[5] The analysis measures the cross section ratio, which depends linearly on Γ_H:

$$\frac{\sigma^{\text{off-shell}}_{gg\rightarrow H\rightarrow ZZ}}{\sigma^{\text{on-shell}}_{gg\rightarrow H\rightarrow ZZ}} \sim \frac{m_H \Gamma_H}{(2m_X)^2}. \tag{6.2}$$

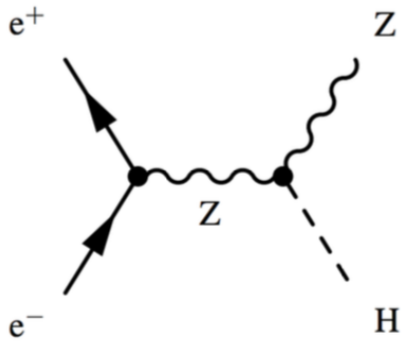

Figure 6.3: Feynman diagram for the Higgsstrahlung process given by $e^+e^- \to Zh$ that dominates the Higgs boson production at $\sqrt{s}$=250 GeV in an electron linear collider.

boson cross section [181, 182]. In this method, a Higgs boson is produced through the ZH production mode (Fig. 6.3). The recoiling Z boson is identified and through conservation of energy of the collision a missing mass can be computed. The inclusive Z+Higgs boson cross section, $\sigma_{e^+e^-\to ZH}$, can then be deduced from the missing mass distribution.

From a measurement of the inclusive cross section, the Higgs boson width can be determined as follows. Noting that the cross sections for the exclusive final-state decays $H \to XX$ can be expressed as:

$$\sigma_{ZH(XX)} = \sigma_{e^+e^-\to ZH} \times \mathrm{BR}_{\mathrm{H}\to\mathrm{XX}} \quad \propto \quad g^2_{hZZ}\frac{g^2_{HXX}}{\Gamma_H}, \tag{6.3}$$

and, that the total ZH cross section is proportional to the square of the coupling

between the Higgs and Z bosons, g^2_{HZZ}: $\sigma_{e^+e^-\to ZH} \propto g^2_{HZZ}$, we have:

$$\Gamma_H \propto g^2_{HZZ} \frac{g^2_{HZZ}}{\sigma_{e^+e^-\to ZH(ZZ)}} \tag{6.4}$$

$$\propto \frac{(\sigma_{e^+e^-\to ZH})^2}{\sigma_{e^+e^-\to ZH(ZZ)}} \tag{6.5}$$

$$\propto \frac{\sigma_{e^+e^-\to ZH}}{\mathrm{BR(H\to ZZ)}} \propto \frac{g^2_{HZZ}}{\mathrm{BR(H\to ZZ)}}. \tag{6.6}$$

That is, we can extract Γ_H by directly measuring the total ZH cross section and correcting it by the branching ratio of $H \to ZZ$. [6]

In summary, to measure the width one would need e.g. the following ratio:

$$\Gamma_H \propto \frac{[\sigma_{e^+e^-\to ZH}]^2}{\sigma_{e^+e^-\to ZH(ZZ)}} \tag{6.8}$$

6.1.3 Analogy of the lepton collider measurement at the LHC

In this section, we explore the possibility of constraining Γ_H at the LHC with an analogy of the recoil measurement as used at a lepton collider. However, we make two changes. First, in place of a recoiling Z boson, we study a Higgs + jet(s) topology, as shown in Fig. 6.4. The ZH production mode is sub-dominant at the LHC, in contrast to the lepton collider, and prevents a direct application of the above approach that is sensitive to Γ_H at the LHC. Second, we assume that the recoiling jet(s) give sufficiently high energy to the Higgs boson such

[6] Note that we could have chosen another final state, e.g. $H \to XX$. In that case we would have:

$$\Gamma_H \propto \frac{g^2_{HXX}}{\mathrm{BR(H\to XX)}} \tag{6.7}$$

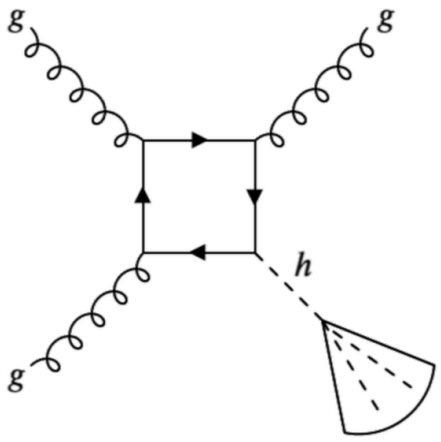

Figure 6.4: Sample Feynman diagram contributing to the gluon fusion Higgs production process at the LHC where the Higgs is recoiling against a jet.

that its decay products all fall into a single cone. We then reconstruct the decay products as a single jet and extract the Higgs boson signal from this jet by cutting as minimally as possible on the decay components.

With these modifications, we *assume* that we are able to perform a measurement of the inclusive Higgs boson cross section in the boosted regime: $\sigma_{gg\to H}$. This assumption implies that we can constrain SM and possible BSM Higgs decay modes. The latter is not always true, making this measurement model dependent. Additionally, the purity of the Higgs boson signal extracted from the reconstructed jet cone may affect the cross section measurement. The definition of a pure H signal includes invisible and partially visible decays that may not be captured in the jet cone. We leave the discussion of the extent of this dependence and presence of a possible bias till the end of the chapter.

Under this assumption, however, we can write the cross section as proportional to the gluon coupling: $\sigma_{gg\to H} \propto \tilde{g}^2_{gg}(p_T)$ [7]. Then, to arrive to a

[7]We have written the gluon coupling as $\tilde{g}_{gg}(p_T)$ to make it clear that this is really an

similar expression than Eq. 6.8, we need to consider three other cross section measurements:

- the gluon fusion production of the boosted Higgs boson to b-quarks: $\sigma_{gg\to H(b\bar{b})}$,
- the W boson associated production of the Higgs boson to b-quarks: $\sigma_{W+H(b\bar{b})}$, and
- the W boson associated production of the Higgs boson to W bosons: $\sigma_{W+H(WW)}$.

We have chosen these exclusive cross section measurements since the LHC experiments have already explored these final states and set initial bounds on their production.

We can constrain the Higgs boson width by computing a scale factor $\mu_\Gamma = \mu_H/\mu_{SM}$ defined as:

$$\mu_\Gamma = \mu^2_{gg\to H}\frac{\mu^2_{WH\to b\bar{b}}}{\mu^2_{ggH\to b\bar{b}}\,\mu_{W+H\to WW}}, \text{with uncertainty} \tag{6.9}$$

$$\delta\mu^2_\Gamma = 4\delta\mu^2_{gg\to H} + \delta\mu^2_{W+H\to WW} + 4\delta\mu^2_{W+H\to b\bar{b}} + 4\delta\mu^2_{ggH\to b\bar{b}}, \tag{6.10}$$

Where $\delta\mu$ signifies the uncertainty on the respective scale factor. [8]

To estimate the precision on the μ_Γ measurement, we take into account the current and expected precision of these cross section measurements. If we consider the latest measurements of the LHC data, with 36 fb^{-1}of integrated

effective coupling which is dependent on the p_T.

[8]These expressions can be derived as follows.

The Higgs to b-quark coupling can be written for two production modes, which have been proven sensitive to $\bar{b}b$ decays: gluon fusion ($gg \to H$) and W boson associated production

luminosity, for $W + H \to \bar{b}b$ [183, 184], $W + H \to WW$ [185, 186], and $gg \to H \to \bar{b}b$ [160, 187], we find a 1σ uncertainty of 20%, 18% and 80%, respectively. Projections on these uncertainties to a full 3000 fb^{-1} dataset, or equivalently 3 ab^{-1}, which is expected with the HL-LHC, can be obtained from [188]. We assume these uncertainties to be 9%, 5% and 25%, respectively. The missing piece is a measurement of the inclusive cross section $\delta\mu_{ggH}$. In the following two sections, we discuss the strategy to constrain this uncertainty and how this can help us bound the total width of the Higgs boson.

$(W + H)$. Following the narrow width assumption, the cross sections can be expressed as:

$$\sigma_{gg\to H(b\bar{b})} \propto \frac{\tilde{g}^2_{gg}(p_T) g^2_{b\bar{b}}}{\Gamma_H}, \text{and} \tag{6.11}$$

$$\sigma_{W+H(b\bar{b})} \propto \frac{g^2_{WW} g^2_{b\bar{b}}}{\Gamma_H}, \tag{6.12}$$

and their ratio yields:

$$\frac{\sigma_{W+H(b\bar{b})}}{\sigma_{gg\to H(b\bar{b})}} \propto \frac{g^2_{WW}}{\tilde{g}^2_{gg}(p_T)} \tag{6.13}$$

Multiplying this by the inclusive cross section, we get:

$$\sigma_{gg\to H} \times \frac{\sigma_{W+H(b\bar{b})}}{\sigma_{gg\to H(b\bar{b})}} \propto g^2_{WW} \tag{6.14}$$

Moreover, for the W boson decay by W associated production, $W + H \to WW$, the cross section is proportional to the W boson coupling over the width:

$$\sigma_{W+H(WW)} \propto \frac{g^4_{WW}}{\Gamma_H} \tag{6.15}$$

Thus, we can take Γ_H and square the ratio from 6.14 to write the total Higgs boson width as:

$$\Gamma_H \propto \frac{1}{\sigma_{W+H\to WW}} \times \left(\sigma_{gg\to H} \times \frac{\sigma_{W+H\to b\bar{b}}}{\sigma_{gg\to h\to b\bar{b}}}\right)^2 \tag{6.16}$$

6.2 Strategy for boosted Higgs boson measurements

Since our goal is to measure all the SM Higgs boson decay modes, we ultimately aim to select Higgs bosons using only event and jet properties common for all its decays. In practice, we can first utilize a basic selection that mimics the selection used in the search for $Z'(qq)$ resonances, presented in Chapter 5. The main requirement in this analysis is for the Higgs to be produced at high p_{T} such that its decay products are collimated and reconstructed into one single large radius jet. Beyond this selection, we have three main handles to enhance the Higgs boson signal: the jet p_{T}, the substructure of the jet, and the jet mass.

All of these handles were already discussed in Chapter 4 but have different effects on the purity of the H signal. A high-p_{T} jet selection has a high level of purity and does not add any model dependence. This is because the p_{T} spectrum of the Higgs boson is harder than the background, when a fixed mass window is considered. In this study, we assume that advances in the capabilities of the trigger system will allow a $p_{\mathrm{T}} > 400$ GeV threshold across most of the LHC and HL-LHC running. A substructure jet-prong selection, however, can enhance only certain decay modes, e.g. two-prong modes such as $H \to \bar{b}b$ or $H \to \bar{c}c$, while leaving out non-two-pronged decays such as $H \to WW \to qq'qq'$. This leads us to consider other substructure observables that enhance the radiation pattern of a color singlet like the Higgs particle. Finally, the groomed jet mass still provides a very effective way to discriminate the Higgs signal but its reconstruction can leave out partially-visible decays, such as those with neutrinos in the final state like $H \to \tau\tau$.

To address the last two observations, we can start by optimizing the jet mass reconstruction. To include invisible particles present in Higgs boson final states, we can utilize the missing transverse energy $p_{\mathrm{T}}^{\mathrm{miss}}$ or MET. Adding the $p_{\mathrm{T}}^{\mathrm{miss}}$ to the jet will recover the lost energy and produce a better estimate of the true Higgs boson properties. Thus, the first improvement to the analysis strategy is to require our Higgs jet to be the leading jet in $p_{\mathrm{T}}^{\mathrm{jet}+\mathrm{p}_{\mathrm{T}}^{\mathrm{miss}}}$ in the event. To perform the vector addition of the jet and MET, we assume the missing energy vector is aligned with the jet axis.

A complication to this procedure is that the MET resolution in events with high energy jets is quite poor; Higgs decays without neutrinos in the final state can still produce over 100 GeV of $p_{\mathrm{T}}^{\mathrm{miss}}$. Since this artificial $p_{\mathrm{T}}^{\mathrm{miss}}$ would worsen the mass resolution when it is added to the jet, we first perform a dedicated regression for the true MET and utilize the regressed MET as our default MET calculation. The regression is designed to remain model independent and to eliminate artificial MET as efficiently as possible. Figure 6.5 shows a comparison between the groomed mass for the Higgs jet and the jet mass as computed using the jet - regressed MET combination.

The dominant multijet background can be further suppressed by using information that captures the internal structure of the jet. While most SM Higgs boson decays result in a jet consistent with either two (ex. $H \to q\bar{q}$), three (ex. $H \to qql\nu$) or four (ex. $H \to qqqq$) prongs, the multijet background consists primarily of jets consistent with one prong. Thus, one can use a jet-prong variable, such as τ_{21} to mitigate the multijet background. Figure 6.6 shows the discriminating power of this variable and its mass-decorrelated

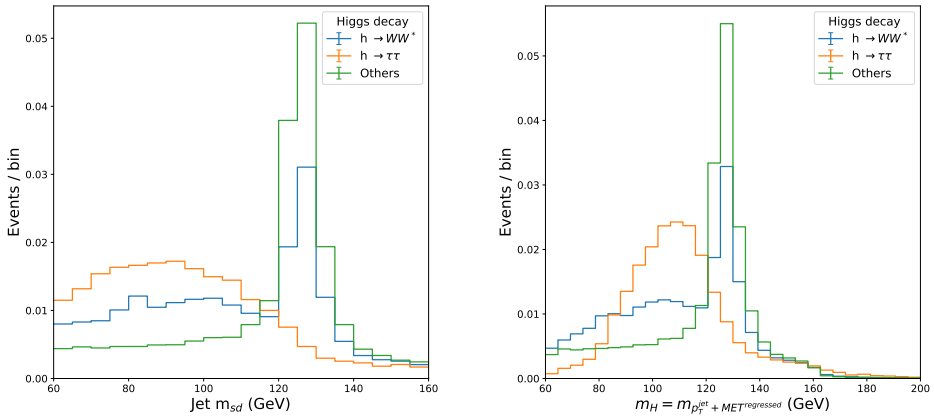

Figure 6.5: Comparison of the performance of the groomed mass (m_{SD}) and the reconstructed Higgs mass $m(\text{jet} + \text{MET})$. Those Higgs decays without genuine MET (ex. $H \to gg$) are unaffected by the new mass reconstruction, those with minimal genuine MET (ex. $H \to b\bar{b}$) are improved slightly, and those with large genuine MET (ex. $H \to \tau\tau$) are greatly improved.

version, τ_{21}^{DDT}, for the various decay modes of the SM Higgs boson. It is clear that the jet τ_{21} distinguishes $H \to \bar{b}b, \bar{c}c$ and even $H \to \tau\tau$ decays, but provides poor discrimination for four or three-pronged decays decays. Therefore, the use of a single jet-prong variable would clearly result in a biased measurement.

Moreover, the definition of an "N-prong" jet is ambiguous and can leave out $H \to gg$ decays. To distinguish these type of background-like decays of the Higgs boson from colored states of quarks and gluon jets, one can rely on the color singlet nature of the Higgs jet. In particular, we study the performance of collinear drop jet substructure observables, recently introduced in [189]. The purpose of these variables is to retain components of the soft radiation in a jet while removing collinear radiation. Such observables can be exploited for a study of the color radiation pattern of the particle initiating a jet. Thus, they

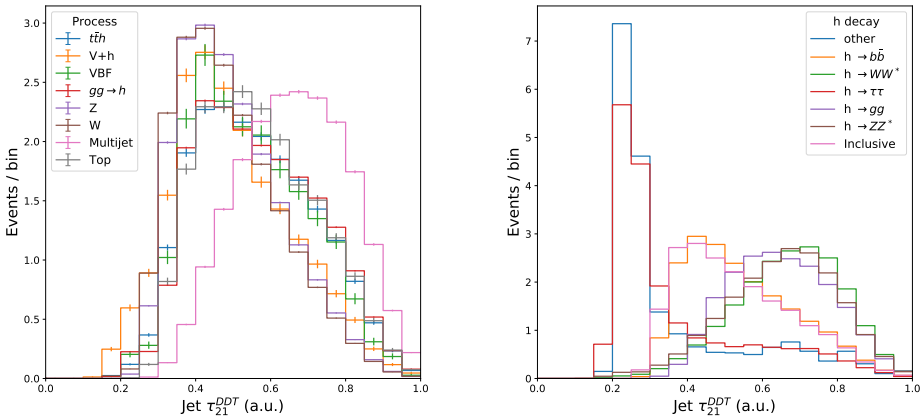

Figure 6.6: τ_{21}^{DDT} observable shown for the inclusive Higgs signal vs other backgrounds (left) and for different SM Higgs decays (right). For decays other than $H \to b\bar{b}$ or $h \to c\bar{c}$, the discriminant is not able to separate very well the signal from the main multijet background.

could provide a handle to isolate the color singlet Higgs jet, without added assumptions of its decay.

In Fig. 6.7, we test the performance of one of these variables, formed by the ratio of the ungroomed mass of the jet vs the groomed mass:

$$m_{\mathrm{jet}} / m_{\mathrm{groomed\ jet}}, \tag{6.17}$$

using the soft-drop algorithm. For background QCD jets the mass-ratio distribution is slightly harder than for signal Higgs jets since more soft radiation is removed. This is only one example of color jet identification and no further studies have been performed, but provide a hint on a possible new direction for boosted particle tagging.

Finally, we also explore the use of machine learning to exploit particle level information of the jet. We construct a deep neural network discriminant that

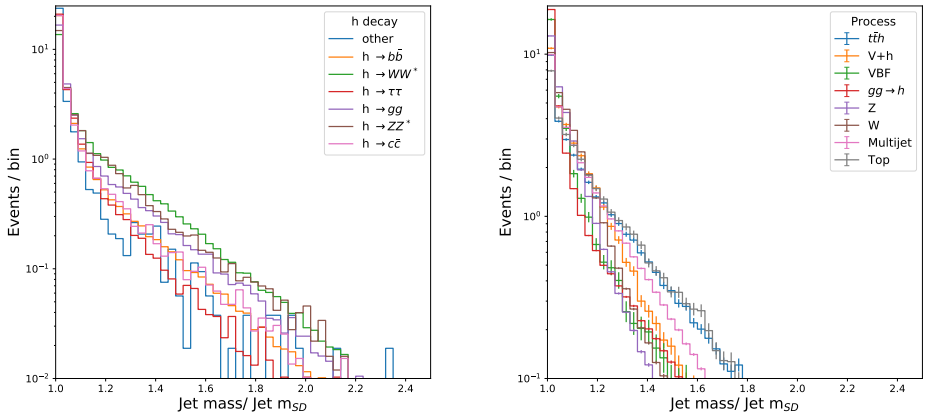

Figure 6.7: Ratio of the ungroomed mass of the jet with respect to the groomed mass, using the soft-drop algorithm. The two jet masses differ more for QCD background decays leading to a discrimination favoring color-singlet jets.

employs jet particles, similar to those presented in Sec. 4.2.4. The details of its architecture and performance can be found in Ref. [14]. For brevity, we refer to this discriminant as the GRU classifier [190, 191]. We compare the background rejection power of this algorithm to the τ_{21}^{DDT} and jet mass-ratio observables for different signal efficiencies in Fig. 6.8. The GRU classifier significantly improves the Higgs boson signal efficiency for a fixed background efficiency well beyond the critical point where $\epsilon_S = \sqrt{\epsilon_B}$.

Finally, we note that for each of these jet observables we apply a DDT transformation, as detailed in Sec. 4.2.5, to prevent any distortion of the mass spectrum after a jet substructure selection.

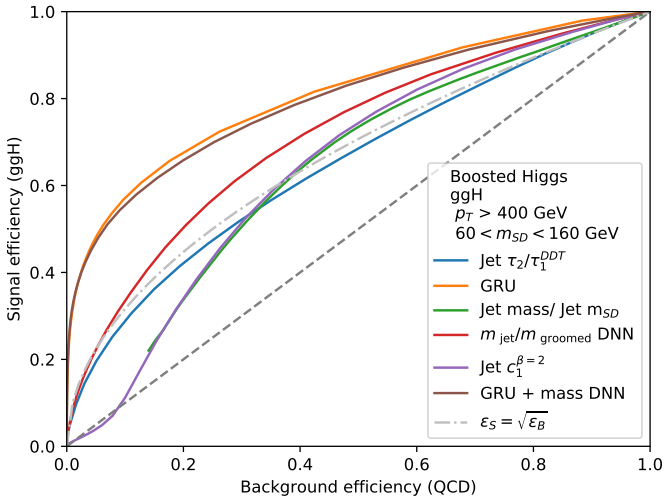

Figure 6.8: Comparison of the performance for the different algorithms that identify an inclusive SM Higgs boson signal against the dominant QCD multijet background. The GRU and jet mass-ratio neural networks developed in this Chapter show improved performance with respect to the baseline τ_{21}^{DDT} selection.

6.3 Results

After a selection of events that follows the strategy detailed in the last Section, i.e. high-p_{T} jet, a jet mass reconstruction that includes the event's $p_{\text{T}}^{\text{miss}}$, and a jet substructure selection, we can extract the Higgs boson signal from a fit to the jet mass distribution. Figure 6.9 shows the mass distribution after a selection on the decorrelated versions of the jet τ_{21}, jet-mass ratios and GRU observables. All of the jet tagger selections are such that keep only 1% of the multijet background. For our fit, we scale the MC events to the point that they reach a total integrated luminosity of 3 ab^{-1}, expected at the HL-LHC. Since the aim of this study is to only provide a rough estimate of the sensitivity of the boosted Higgs boson signal, we only use simulation to model small resonant

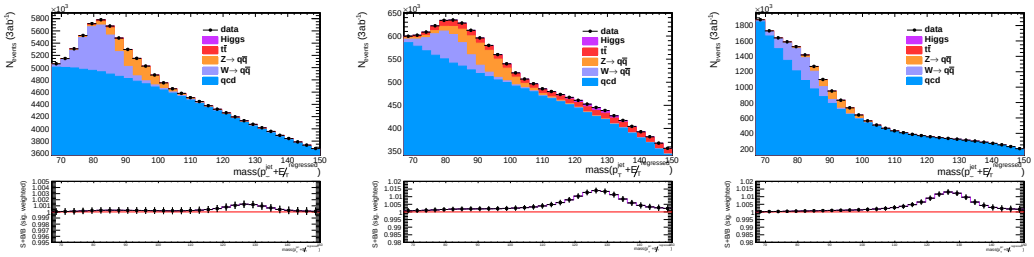

Figure 6.9: Combined mass distribution of the inclusive Higgs signal extraction for different tagger selections. The reconstructed Higgs jet mass distribution is shown after a τ_{21}^{DDT}, GRU$^{\text{DDT}}$ and decorrelated version of jet-mass ratio observable. The cuts on all discriminants correspond to 1% of multijet background efficiency. The Higgs signal (violet) is small and hidden by the background processes, but its significance is illustrated in the bottom panel.

backgrounds and estimate the contribution of the main multijet background with a simple polynomial fit. Finally, we use a SM Higgs boson simulation for our signal extraction. Detailed descriptions of the uncertainties used in this fit can be found in [14].

We express the results of the extracted signal from the fit in terms of the 1σ uncertainty on $\delta\mu_{gg\to H}$. Figure 6.10 shows the results of the inclusive limit using various discriminators and working points. While the figure shows various results under different assumptions and uncertainties on the fit, we can focus on the upper bounds that indicate the sensitivity based on an estimate of the multijet background known to the percent level. The application of a combined GRU$^{\text{DDT}}$ and mass ratios discriminator gives a full factor of 2 improvement from the inclusive result.

Our benchmark result, with the combined GRU and mass ratios discriminant, gives us a 1σ bound on the inclusive cross section of $\delta\sigma = 0.14 \times \sigma\text{SM}$. Using Eq. 6.9, this translates to a bound on the total width with an uncertainty of $\delta\Gamma_H < 1.4$ MeV. This model-dependent bound assumes, roughly, a standard

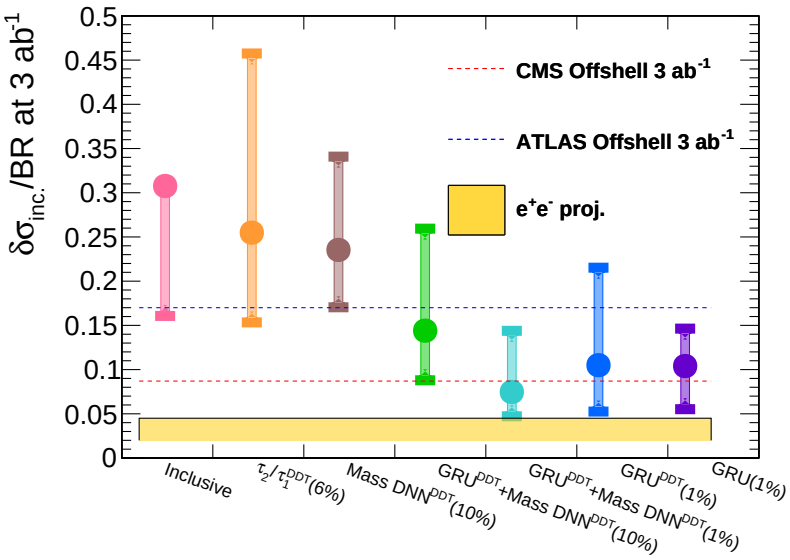

Figure 6.10: Estimated 1σ sensitivity for selections on various working points and discriminators extrapolated to 3 ab^{-1}. For each point, the upper bound of the band indicates the sensitivity using a fourth-order Bernstein polynomial to fit the background. The lower bound of the band shows the result when no systematic uncertainties are included in the calculation and the point indicates the performance when a template fit is utilized for the signal extraction. The inclusive bin has a tadpole-like point indicating that the Bernstein polynomial fit did not converge. Our benchmark result, $\text{GRU}^{\text{DDT}} + \text{mass} - \text{DNN}^{\text{DDT}}$, results in a 1σ bound on the inclusive SM Higgs boson signal of $\delta\sigma = 0.14 \times \sigma_{\text{SM}}$. Additionally, we quote the projected extrapolations of the width for both the ATLAS and CMS off-shell interference analyses. This is done by taking the benchmark numbers used in Eq. 6.9 and subtracting the width uncertainty with the projected uncertainties on the various Higgs boson projection modes; if the projected uncertainties are reduced then the lines from the off-shell width measurements will increase. Lastly, we show a band for the quoted uncertainties from electron colliders. The various projected results range from 2% to 4.5% [181, 182, 192–194].

model admixture of decays. Given the variation over decay modes, we find that this bound can be applied to final states with varying admixture of heavy vector-bosons, τ leptons, and quarks.

Considering the worst possible sensitivity over all decay modes, we find that with the di-gluon final state we obtain an uncertainty on the total cross section of $0.41 \times \sigma_{SM}$. This corresponds to a width constraint of $\delta\Gamma_H < 3.5$ MeV. These results are comparable to the ATLAS and CMS projected off-shell measurements of the Higgs boson total width, which are found to be 1.6 MeV and 1 MeV respectively.

6.3.1 Model assumptions and bias

In this study we assume we can obtain bounds on the inclusive cross section by extracting a SM Higgs boson signal from a large radius jet. This statement implies three additional assumptions that can translate to model dependence or bias in our measurement:

1. All objects from the Higgs boson decay, both with missing energy signatures and visible signatures, occur within a single cone in the event.

   This assumption may not always be true if Higgs decay products, such as neutrinos or possible DM particles, escape detection. In our analysis, we have attempted to recover these signals by including the missing energy in the jet reconstruction. While we have only explicitly shown this modified reconstruction helps recover $H \to \tau\tau$ and $H \to WW^*$ sensitivity, we believe this approach is broadly applicable to all semi-visible decays.

In the instance where the Higgs boson decays completely invisibly this analysis is not applicable anymore. The equivalent search for an invisibly decaying Higgs boson would occur with a single jet recoiling against missing energy. Projected results on the bounds of the Higgs boson to invisible cross section, at the HL-LHC, range from roughly 1% to 4% [179, 195].

2. We treat all SM Higgs boson decays as a proxy for all signatures, including possible BSM signatures.

   While BSM signatures are not included in our signal model, we can take background-like decays of the Higgs as a proxy for all visible background-like BSM signatures. In this study, we have made a first attempt at isolating $H \to gg$ decays by exploiting the color-singlet nature of the Higgs boson through the mass-ratios selection. If we can isolate background-like signals by using a universal property of the Higgs boson then we can argue that other non-SM signatures that look like background could be isolated with this approach. However, it remains to be shown that SM Higgs boson decays covering a broad range of signatures can really serve as a proxy for all visible signatures.

3. A SM Higgs jet substructure selection can improve the purity of our measurement.

   While this statement is true for a given Higgs boson decay, a selection on a jet discriminant can introduce a bias in our measurement if the properties exploited in the discriminant only appear in one particular decay mode. We encourage then the development of Higgs boson jet

identification methods with properties that are universal to all Higgs boson decays.

In summary, we have performed a study of the sensitivity of the LHC to all the decay modes of the Standard Model Higgs. We have assumed that the inclusive Higgs boson cross section can be constrained by catching all the Higgs boson decays in a high-p_{T} jet cone. The analysis strategy to identify these boosted Higgs decays to jets is similar to that presented in Chapter 5. While our constraint is model dependent and may exclude new physics contributions to the Higgs jet, it explores for the first time the reconstruction and identification of all the SM Higgs boson decays, in particular the unexplored $H \to gg/WW/\tau\tau$ hadronic decay modes. This study motivates a rich program of exploring the Higgs boson at high p_{T} in *all* the SM final states.

Chapter 7

Summary

In this dissertation, we have demonstrated that hadronically decaying low mass resonances can be triggered on by using initial state radiation jets in the final event signature. In particular, we have shown that this triggering approach can be used to search for a narrow vector resonance (Z′) decaying into a quark-antiquark pair and reconstructed as a single jet, that it can allow for the clear observation of a merged W and Z boson peak in the jet mass distribution, the first of this type at the LHC, and, that the study of a boosted resonance and an extra jet can allow for the study of highly energetic Higgs bosons and their decays into jets. In this thesis, jet substructure techniques have been employed to identify a jet containing a signal candidate over a smoothly falling jet mass distribution in data. In particular, we have developed a novel technique that prevents the distortion of the jet mass distribution after a jet substructure selection. No significant excess above the standard model prediction has been observed in the search for low-mass resonances and we have set upper limits at 95% confidence level on the Z′ boson coupling to quarks, g'_{q}, as a function of the Z' boson mass. We have excluded coupling

values of $g'_q > 0.4$ over the signal mass range from 50 to 450 GeV, with the most stringent constraints set for masses below 250 GeV where we have excluded coupling values of $g'_q > 0.2$. For masses between 50 and 300 GeV these are the most sensitive limits to date. Future searches for low mass hadronically decaying resonances will greatly benefit from continued work in the identification of large radius jets, such as machine-learning jet tagging techniques, and a precise understanding of the W and Z boson contributions and the jet mass scale and resolution. The ideas presented in this thesis to identify boosted Higgs boson signatures are far from comprehensive but constitute a step beyond the already explored properties of Higgs boson jets that may allows to build a boosted Higgs exploration program at the LHC.

References

[1] Daniel Abercrombie et al. "Dark Matter Benchmark Models for Early LHC Run-2 Searches: Report of the ATLAS/CMS Dark Matter Forum". 2015.

[2] Giorgio Busoni et al. "Recommendations on presenting LHC searches for missing transverse energy signals using simplified s-channel models of dark matter". 2016.

[3] Morad Aaboud et al. "Search for new phenomena in dijet events using 37fb^{-1} of pp collision data collected at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *Phys. Rev. D* 96 (2017), p. 052004. DOI: 10.1103/PhysRevD.96.052004.

[4] Albert M Sirunyan et al. "Search for narrow and broad dijet resonances in proton-proton collisions at $\sqrt{s} = 13$ TeV and constraints on dark matter mediators and other new particles". In: *JHEP* 08 (2018), p. 130. DOI: 10.1007/JHEP08(2018)130. arXiv: 1806.00843 [hep-ex].

[5] Vardan Khachatryan et al. "Search for narrow resonances in dijet final states at $\sqrt{s} = 8$ TeV with the novel CMS technique of data scouting". In: *Phys. Rev. Lett.* 117 (2016), p. 031802. DOI: 10.1103/PhysRevLett.117.031802. arXiv: 1604.08907 [hep-ex].

[6] Albert M Sirunyan et al. "Search for dijet resonances in proton-proton collisions at $\sqrt{s} = 13$ TeV and constraints on dark matter and other models". In: *Phys. Lett. B* 769 (2017), p. 520. DOI: 10.1016/j.physletb.2017.02.012. arXiv: 1611.03568 [hep-ex].

[7] M. Aaboud et al. "Search for low-mass dijet resonances using trigger-level jets with the ATLAS detector in pp collisions at $\sqrt{s} = 13$ TeV". In: *Phys. Rev. Lett.* 121 (2018), p. 081801. DOI: 10.1103/PhysRevLett.121.081801. arXiv: 1804.03496 [hep-ex].

[8] Albert M Sirunyan et al. "Search for dijet resonances using events with three jets in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *Phys. Lett. B* 805 (2020), p. 135448. DOI: 10.1016/j.physletb.2020.135448. arXiv: 1911.03761 [hep-ex].

[9] Peter W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons". In: *Phys. Rev. Lett.* 13 (1964). Ed. by J.C. Taylor, pp. 508–509. DOI: 10.1103/PhysRevLett.13.508.

[10] Peter W. Higgs. "Broken symmetries, massless particles and gauge fields". In: *Phys. Lett.* 12 (1964), pp. 132–133. DOI: 10.1016/0031-9163(64)91136-9.

[11] F. Englert and R. Brout. "Broken Symmetry and the Mass of Gauge Vector Mesons". In: *Phys. Rev. Lett.* 13 (1964). Ed. by J.C. Taylor, pp. 321–323. DOI: 10.1103/PhysRevLett.13.321.

[12] Albert M Sirunyan et al. "Search for low mass vector resonances decaying into quark-antiquark pairs in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *JHEP* 01 (2018), p. 097. DOI: 10.1007/JHEP01(2018)097. arXiv: 1710.00159 [hep-ex].

[13] Albert M Sirunyan et al. "Search for low mass vector resonances decaying into quark-antiquark pairs in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *Phys. Rev. D* 100.11 (2019), p. 112007. DOI: 10.1103/PhysRevD.100.112007. arXiv: 1909.04114 [hep-ex].

[14] Philip Coleman Harris, Dylan Sheldon Rankin, and Cristina Mantilla Suarez. "An approach to constraining the Higgs width at the LHC and HL-LHC". In: (2019). arXiv: 1910.02082 [hep-ph].

[15] S.L. Glashow. "Partial Symmetries of Weak Interactions". In: *Nucl. Phys.* 22 (1961), pp. 579–588. DOI: 10.1016/0029-5582(61)90469-2.

[16] Steven Weinberg. "A Model of Leptons". In: *Phys. Rev. Lett.* 19 (1967), pp. 1264–1266. DOI: 10.1103/PhysRevLett.19.1264.

[17] Abdus Salam. "Weak and Electromagnetic Interactions". In: *Conf. Proc. C* 680519 (1968), pp. 367–377. DOI: 10.1142/9789812795915_0034.

[18] David J. Gross and Frank Wilczek. "Ultraviolet Behavior of Nonabelian Gauge Theories". In: *Phys. Rev.* 30 (1973). Ed. by J.C. Taylor, pp. 1343–1346. DOI: 10.1103/PhysRevLett.30.1343.

[19] H. David Politzer. "Reliable Perturbative Results for Strong Interactions". In: *Phys. Rev. Lett.* 30 (1973). Ed. by J.C. Taylor, pp. 1346–1349. DOI: 10.1103/PhysRevLett.30.1346.

[20] W. Pauli. “The Connection Between Spin and Statistics”. In: *Phys. Rev.* 58 (1940), pp. 716–722. DOI: 10.1103/PhysRev.58.716.

[21] Eugene P. Wigner. “On Unitary Representations of the Inhomogeneous Lorentz Group”. In: *Annals Math.* 40 (1939). Ed. by Y.S. Kim and W.W. Zachary, pp. 149–204. DOI: 10.2307/1968551.

[22] M. et. al. Tanabashi. “Review of Particle Physics”. In: *Phys. Rev. D* 98 (3 2018 and 2019 update), p. 030001. DOI: 10.1103/PhysRevD.98.030001. URL: https://link.aps.org/doi/10.1103/PhysRevD.98.030001.

[23] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson. “Experimental Test of Parity Conservation in Beta Decay”. In: *Phys. Rev.* 105 (4 1957), pp. 1413–1415. DOI: 10.1103/PhysRev.105.1413. URL: https://link.aps.org/doi/10.1103/PhysRev.105.1413.

[24] Peter W. Higgs. “Spontaneous Symmetry Breakdown without Massless Bosons”. In: *Phys. Rev.* 145 (4 1966), pp. 1156–1163. DOI: 10.1103/PhysRev.145.1156. URL: https://link.aps.org/doi/10.1103/PhysRev.145.1156.

[25] T.W.B. Kibble. “Symmetry breaking in nonAbelian gauge theories”. In: *Phys. Rev.* 155 (1967). Ed. by J.C. Taylor, pp. 1554–1561. DOI: 10.1103/PhysRev.155.1554.

[26] Philip W. Anderson. “Plasmons, Gauge Invariance, and Mass”. In: *Phys. Rev.* 130 (1963). Ed. by J.C. Taylor, pp. 439–442. DOI: 10.1103/PhysRev.130.439.

[27] G.S. Guralnik, C.R. Hagen, and T.W.B. Kibble. “Global Conservation Laws and Massless Particles”. In: *Phys. Rev. Lett.* 13 (1964). Ed. by J.C. Taylor, pp. 585–587. DOI: 10.1103/PhysRevLett.13.585.

[28] Wikimedia Commons. *Standard Model of Elementary Particles*. 2020 (accessed May 19, 2020). URL: https://commons.wikimedia.org/wiki/File:Standard_Model_of_Elementary_Particles_modified_version.svg.

[29] John C. Collins, Davison E. Soper, and George F. Sterman. “Factorization of Hard Processes in QCD”. In: vol. 5. 1989, pp. 1–91. DOI: 10.1142/9789814503266_0001. arXiv: hep-ph/0409313.

[30] Guido Altarelli and G. Parisi. “Asymptotic Freedom in Parton Language”. In: *Nucl. Phys. B* 126 (1977), pp. 298–318. DOI: 10.1016/0550-3213(77)90384-4.

[31] V.N. Gribov and L.N. Lipatov. "Deep inelastic e p scattering in perturbation theory". In: *Sov. J. Nucl. Phys.* 15 (1972), pp. 438–450.

[32] Yuri L. Dokshitzer. "Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics." In: *Sov. Phys. JETP* 46 (1977), pp. 641–653.

[33] Richard D. Ball et al. "Parton distributions from high-precision collider data". In: *Eur. Phys. J. C* 77.10 (2017), p. 663. DOI: 10.1140/epjc/s10052-017-5199-5. arXiv: 1706.00428 [hep-ph].

[34] J. Alwall, S. Höche, F. Krauss, N. Lavesson, L. Lönnblad, F. Maltoni, M. L. Mangano, M. Moretti, C. G. Papadopoulos, F. Piccinini, S. Schumann, M. Treccani, J. Winter, and M. Worek. "Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions". In: *Eur. Phys. J. C* 53 (2008), p. 473. DOI: 10.1140/epjc/s10052-007-0490-5. arXiv: 0706.2569 [hep-ph].

[35] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S Shao, T. Stelzer, P. Torrielli, and M. Zaro. "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations". In: *JHEP* 07 (2014), p. 079. DOI: 10.1007/JHEP07(2014)079. arXiv: 1405.0301 [hep-ph].

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

[37] Stefano Frixione, Paolo Nason, and Carlo Oleari. "Matching NLO QCD computations with Parton Shower simulations: the POWHEG method". In: *JHEP* 11 (2007), p. 070. DOI: 10.1088/1126-6708/2007/11/070. arXiv: 0709.2092 [hep-ph].

[38] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX". In: *JHEP* 06 (2010), p. 043. DOI: 10.1007/JHEP06(2010)043. arXiv: 1002.2581 [hep-ph].

[39] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, and Ilten. "An introduction to PYTHIA 8.2". In: *Comput. Phys. Commun.* 191 (2015), p. 159. DOI: 10.1016/j.cpc.2015.01.024. arXiv: 1410.3012 [hep-ph].

[40] Johannes Bellm et al. “Herwig 7.0/Herwig++ 3.0 release note”. In: *Eur. Phys. J. C* 76.4 (2016), p. 196. DOI: 10.1140/epjc/s10052-016-4018-8. arXiv: 1512.01178 [hep-ph].

[41] Bo Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand. “Parton Fragmentation and String Dynamics”. In: *Phys. Rept.* 97 (1983), pp. 31–145. DOI: 10.1016/0370-1573(83)90080-7.

[42] Albert M Sirunyan et al. “Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements”. 2019.

[43] CMS Collaboration. *Summary of the cross section measurements of Standard Model processes*. 2020. URL: https://twiki.cern.ch/twiki/pub/CMSPublic/PhysicsResultsCombined/SigmaNew_v0.pdf.

[44] Georges Aad et al. “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”. In: *Phys. Lett. B* 716 (2012), pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex].

[45] Serguei Chatrchyan et al. “Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC”. In: *Phys. Lett. B* 716 (2012), pp. 30–61. DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235 [hep-ex].

[46] Tamara M. Davis. “Cosmological constraints on dark energy”. In: *Gen. Rel. Grav.* 46 (2014), p. 1731. DOI: 10.1007/s10714-014-1731-1. arXiv: 1404.7266 [astro-ph.CO].

[47] N. Aghanim et al. “Planck 2018 results. VI. Cosmological parameters”. In: (2018). arXiv: 1807.06209 [astro-ph.CO].

[48] A.D. Sakharov. “Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe”. In: *Sov. Phys. Usp.* 34.5 (1991), pp. 392–393. DOI: 10.1070/PU1991v034n05ABEH002497.

[49] Y. Fukuda et al. “Evidence for oscillation of atmospheric neutrinos”. In: *Phys. Rev. Lett.* 81 (1998), pp. 1562–1567. DOI: 10.1103/PhysRevLett.81.1562. arXiv: hep-ex/9807003.

[50] Bogdan A. Dobrescu and Felix Yu. “Coupling-mass mapping of dijet peak searches”. In: *Phys. Rev. D* 88 (2013), p. 035021. DOI: 10.1103/PhysRevD.88.035021. arXiv: 1306.2629 [hep-ph].

[51] F. Zwicky. “On the Masses of Nebulae and of Clusters of Nebulae”. In: *Astrophysical Journal* 86 (1937), p. 217. DOI: 10.1086/143864.

[52] T.S. van Albada, John N. Bahcall, K. Begeman, and R. Sancisi. "The Distribution of Dark Matter in the Spiral Galaxy NGC-3198". In: *Astrophys. J.* 295 (1985), pp. 305–313. DOI: 10.1086/163375.

[53] A. A. Penzias and R. W. Wilson. "A Measurement of Excess Antenna Temperature at 4080 Mc/s." In: *Astrophysical Journal* 142 (1965), pp. 419–421. DOI: 10.1086/148307.

[54] Wayne Hu. "Cosmic Microwave Background Anisotropies". In: *Ann. Rev. Astron. Astrophys.* 40 (2002), pp. 171–216. DOI: 10.1146/annurev.astro.40.060401.093926. arXiv: astro-ph/0110414.

[55] John R. Ellis, J.S. Hagelin, Dimitri V. Nanopoulos, Keith A. Olive, and M. Srednicki. "Supersymmetric Relics from the Big Bang". In: *Nucl. Phys. B* 238 (1984). Ed. by M.A. Srednicki, pp. 453–476. DOI: 10.1016/0550-3213(84)90461-9.

[56] G. Bertone. *Particle Dark Matter: Observations, Models and Searches*. Cambridge University Press, 2010. ISBN: 9780521763684. URL: https://books.google.com/books?id=szrTSWO7D-8C.

[57] Abdelhak Djouadi, Oleg Lebedev, Yann Mambrini, and Jeremie Quevillon. "Implications of LHC searches for Higgs–portal dark matter". In: *Phys. Lett. B* 709 (2012), pp. 65–69. DOI: 10.1016/j.physletb.2012.01.062. arXiv: 1112.3299 [hep-ph].

[58] Georges Aad et al. "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". In: *JHEP* 08 (2016), p. 045. DOI: 10.1007/JHEP08(2016)045. arXiv: 1606.02266 [hep-ex].

[59] Antonio Boveia and Caterina Doglioni. "Dark Matter Searches at Colliders". In: *Ann. Rev. Nucl. Part. Sci.* 68 (2018), pp. 429–459. DOI: 10.1146/annurev-nucl-101917-021008. arXiv: 1810.12238 [hep-ex].

[60] Lyndon Evans and Philip Bryant. "LHC Machine". In: *Journal of Instrumentation* 3.08 (2008), S08001–S08001. DOI: 10.1088/1748-0221/3/08/s08001.

[61] CMS Collaboration. *Public Results of CMS Luminosity Information*. 2019. URL: https://cmslumi.web.cern.ch/cmslumi/publicplots/int_lumi_cumulative_pp_2_run2.png.

[62] Albert M Sirunyan et al. "Pileup mitigation at CMS in 13 TeV data". In: (2020). arXiv: 2003.00503 [hep-ex].

[63] S. Chatrchyan et al. “The CMS Experiment at the CERN LHC”. In: *JINST* 3 (2008), S08004. DOI: 10.1088/1748-0221/3/08/S08004.

[64] CMS Collaboration. “Commissioning and performance of the CMS pixel tracker with cosmic ray muons”. In: *Journal of Instrumentation* 5.03 (2010), T03007âĂŞT03007. ISSN: 1748-0221. DOI: 10.1088/1748-0221/5/03/t03007. URL: http://dx.doi.org/10.1088/1748-0221/5/03/T03007.

[65] M Swartz, D Fehling, G Giurgiu, P Maksimovic, and V Chiochia. “A new technique for the reconstruction, validation, and simulation of hits in the CMS Pixel Detector”. In: *PoS* Vertex 2007.CMS-NOTE-2007-033 (2007), 035. 37 p. DOI: 10.22323/1.057.0035. URL: https://cds.cern.ch/record/1073691.

[66] “CMS Technical Design Report for the Pixel Detector Upgrade”. In: (2012). Ed. by David Aaron Matzner Dominguez et al. DOI: 10.2172/1151650.

[67] Serguei Chatrchyan et al. “Description and performance of track and primary-vertex reconstruction with the CMS tracker”. In: *JINST* 9.10 (2014), P10009. DOI: 10.1088/1748-0221/9/10/P10009. arXiv: 1405.6569 [physics.ins-det].

[68] P. Adzic et al. “Energy resolution of the barrel of the CMS electromagnetic calorimeter”. In: *JINST* 2 (2007), P04004. DOI: 10.1088/1748-0221/2/04/P04004.

[69] Serguei Chatrchyan et al. “Energy Calibration and Resolution of the CMS Electromagnetic Calorimeter in pp Collisions at $\sqrt{s} = 7$ TeV”. In: *JINST* 8 (2013), p. 9009. DOI: 10.1088/1748-0221/8/09/P09009. arXiv: 1306.2016 [hep-ex].

[70] Vardan Khachatryan et al. “Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at $\sqrt{\mathrm{s}} = 8$ TeV”. In: *JINST* 10.06 (2015), P06005. DOI: 10.1088/1748-0221/10/06/P06005. arXiv: 1502.02701 [physics.ins-det].

[71] S. Abdullin et al. “The CMS barrel calorimeter response to particle beams from 2-GeV/c to 350-GeV/c”. In: *Eur. Phys. J. C* 60 (2009), pp. 359–373. DOI: 10.1140/epjc/s10052-009-0959-5.

[72] A.M. Sirunyan et al. "Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *JINST* 13.06 (2018), P06015. DOI: 10.1088/1748-0221/13/06/P06015. arXiv: 1804.04528 [physics.ins-det].

[73] G. Weiglein et al. "Physics interplay of the LHC and the ILC". In: *Phys. Rept.* 426 (2006), pp. 47–358. DOI: 10.1016/j.physrep.2005.12.003. arXiv: hep-ph/0410364.

[74] S. Agostinelli et al. "GEANT4: A Simulation toolkit". In: *Nucl. Instrum. Meth.* A506 (2003), pp. 250–303. DOI: 10.1016/S0168-9002(03)01368-8.

[75] John Allison et al. "Geant4 developments and applications". In: *IEEE Trans. Nucl. Sci.* 53 (2006), p. 270. DOI: 10.1109/TNS.2006.869826.

[76] G. Lindstrom. "Radiation damage in silicon detectors". In: *Nucl. Instrum. Meth. A* 512 (2003). Ed. by P. Holl, G. Lutz, F. Schopper, L. Struder, S. Masciocchi, C. Fiorini, A. Longoni, and M. Sampietro, pp. 30–43. DOI: 10.1016/S0168-9002(03)01874-6.

[77] Morad Aaboud et al. "Modelling radiation damage to pixel sensors in the ATLAS detector". In: *JINST* 14.06 (2019), P06012. DOI: 10.1088/1748-0221/14/06/P06012. arXiv: 1905.03739 [physics.ins-det].

[78] M. Swartz et al. "Observation, modeling, and temperature dependence of doubly peaked electric fields in irradiated silicon pixel sensors". In: *Nucl. Instrum. Meth. A* 565 (2006). Ed. by J. Grosse-Knetter, H. Krueger, and N. Wermes, pp. 212–220. DOI: 10.1016/j.nima.2006.05.002. arXiv: physics/0510040.

[79] Vincenzo et. al. Chiochia. "Simulation of Heavily Irradiated Silicon Pixel Sensors and Comparison with Test Beam Measurements". In: *IEEE Trans. Nucl. Sci.* 52.physics/0411143 (2004), 1067–1075. 6 p. DOI: 10.1109/TNS.2005.852748. URL: https://cds.cern.ch/record/804214.

[80] *CMS Offline Software*. https://github.com/cms-sw/cmssw.

[81] H. Bichsel. "A method to improve tracking and particle identification in TPCs and silicon detectors". In: *Nucl. Instrum. Meth. A* 562 (2006), pp. 154–197. DOI: 10.1016/j.nima.2006.03.009.

[82] G. Kramberger, V. Cindro, I. MandiÄĞ, M. MikuÅ¿, and M. Zavrtanik. "Field engineering by continuous hole injection in silicon detectors irradiated with neutrons". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 497.2 (2003), pp. 440 –449. ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(02)01927-7. URL: http://www.sciencedirect.com/science/article/pii/S0168900202019277.

[83] S. Cucciarelli and D. Kotlinski. *Pixel Hit Reconstruction*. Tech. rep. 2004.

[84] Gavril Giurgiu, D. Fehling, P. Maksimovic, M. Swartz, and V. Chiochia. "Pixel Hit Reconstruction with the CMS Detector". In: (2008). arXiv: 0808.3804 [physics.ins-det].

[85] Benedikt Roland Vormwald. *The CMS Phase-1 pixel detector – experience and lessons learned from two years of operation*. Tech. rep. CMS-CR-2019-019. 07. Geneva: CERN, 2019. DOI: 10.1088/1748-0221/14/07/C07008. URL: https://cds.cern.ch/record/2687017.

[86] P. Billoir and S. Qian. "Simultaneous pattern recognition and track fitting by the Kalman filtering method". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 294.1 (1990), pp. 219 –228. ISSN: 0168-9002. DOI: https://doi.org/10.1016/0168-9002(90)91835-Y. URL: http://www.sciencedirect.com/science/article/pii/016890029091835Y.

[87] R. Fruehwirth. "Application of Kalman filtering to track and vertex fitting". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 262.2 (1987), pp. 444 –450. ISSN: 0168-9002. DOI: https://doi.org/10.1016/0168-9002(87)90887-4. URL: http://www.sciencedirect.com/science/article/pii/0168900287908874.

[88] R Fruehwirth, Wolfgang Waltenberger, and Pascal Vanlaer. *Adaptive Vertex Fitting*. Tech. rep. CMS-NOTE-2007-008. Geneva: CERN, 2007. URL: https://cds.cern.ch/record/1027031.

[89] A.M. Sirunyan et al. "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV". In: *JINST* 13.05 (2018), P05011. DOI: 10.1088/1748-0221/13/05/P05011. arXiv: 1712.07158 [physics.ins-det].

[90] Gianluca Cerminara. *Alignment and Calibration of the CMS Detector*. Tech. rep. CMS-CR-2011-268. Geneva: CERN, 2011. URL: http://cds.cern.ch/record/1399496.

[91] Federico De Guio. *CMS Alignement and Calibration workflows: lesson learned and future plans*. Tech. rep. CMS-CR-2014-335. Geneva: CERN, 2014. DOI: 10.1016/j.nuclphysbps.2015.09.143. URL: https://cds.cern.ch/record/2121268.

[92] Jeffrey Roskes. "A boson learned from its context, and a boson learned from its end". PhD thesis. Baltimore, US: Johns Hopkins University, 2019.

[93] A.M. Sirunyan et al. "Particle-flow reconstruction and global event description with the CMS detector". In: *JINST* 12.10 (2017), P10003. DOI: 10.1088/1748-0221/12/10/P10003. arXiv: 1706.04965 [physics.ins-det].

[94] W Adam, R FrÃijhwirth, A Strandlie, and T Todorov. "Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at the LHC". In: *Journal of Physics G: Nuclear and Particle Physics* 31.9 (2005), N9. URL: http://stacks.iop.org/0954-3899/31/i=9/a=N01.

[95] S. Baffioni, C. Charlot, F. Ferri, D. Futyan, P. Meridiani, I. Puljak, C. Rovelli, R. Salerno, and Y. Sirois. "Electron reconstruction in CMS". In: *Eur. Phys. J. C* 49 (2007), pp. 1099–1116. DOI: 10.1140/epjc/s10052-006-0175-5.

[96] Albert M Sirunyan et al. "Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13$ TeV using the CMS detector". In: *JINST* 14.07 (2019), P07004. DOI: 10.1088/1748-0221/14/07/P07004. arXiv: 1903.06078 [hep-ex].

[97] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "The anti-k_t jet clustering algorithm". In: *JHEP* 04 (2008), p. 063. DOI: 10.1088/1126-6708/2008/04/063. arXiv: 0802.1189 [hep-ph].

[98] Stephen D. Ellis and Davison E. Soper. "Successive combination jet algorithm for hadron collisions". In: *Phys. Rev.* D48 (1993), pp. 3160–3166. DOI: 10.1103/PhysRevD.48.3160. arXiv: hep-ph/9305266 [hep-ph].

[99] Yuri L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber. "Better jet clustering algorithms". In: *JHEP* 08 (1997), p. 001. DOI: 10.1088/1126-6708/1997/08/001. arXiv: hep-ph/9707323 [hep-ph].

[100] "FastJet User Manual". In: *Eur. Phys. J.* C72 (2012), p. 1896. DOI: 10.1140/epjc/s10052-012-1896-2. arXiv: 1111.6097 [hep-ph].

[101] Vardan Khachatryan et al. "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV". In: *JINST* 12.02 (2017), P02014. DOI: 10.1088/1748-0221/12/02/P02014. arXiv: 1607.03663 [hep-ex].

[102] Daniele Bertolini, Philip Harris, Matthew Low, and Nhan Tran. "Pileup Per Particle Identification". In: *JHEP* 10 (2014), p. 059. DOI: 10.1007/JHEP10(2014)059. arXiv: 1407.6013 [hep-ph].

[103] D. Krohn, J. Thaler, and L. Wang. "Jet trimming". In: *JHEP* 02 (2010), p. 084. DOI: 10.1007/JHEP02(2010)084. arXiv: 0912.1342 [hep-ph].

[104] Andrew J. Larkoski, Simone Marzani, Gregory Soyez, and Jesse Thaler. "Soft Drop". In: *JHEP* 05 (2014), p. 146. DOI: 10.1007/JHEP05(2014)146. arXiv: 1402.2657 [hep-ph].

[105] Mrinal Dasgupta, Alessandro Fregoso, Simone Marzani, and Gavin P. Salam. "Towards an understanding of jet substructure". In: *JHEP* 09 (2013), p. 029. DOI: 10.1007/JHEP09(2013)029. arXiv: 1307.0007 [hep-ph].

[106] Andrew J. Larkoski and Jesse Thaler. "Unsafe but Calculable: Ratios of Angularities in Perturbative QCD". In: *JHEP* 09 (2013), p. 137. DOI: 10.1007/JHEP09(2013)137. arXiv: 1307.1699 [hep-ph].

[107] Stephen D. Ellis, Christopher K. Vermilion, and Jonathan R. Walsh. "Techniques for improved heavy particle searches with jet substructure". In: *Phys. Rev.* D80 (2009), p. 051501. DOI: 10.1103/PhysRevD.80.051501. arXiv: 0903.5081 [hep-ph].

[108] Stephen D. Ellis, Christopher K. Vermilion, and Jonathan R. Walsh. "Recombination Algorithms and Jet Substructure: Pruning as a Tool for Heavy Particle Searches". In: *Phys. Rev.* D81 (2010), p. 094023. DOI: 10.1103/PhysRevD.81.094023. arXiv: 0912.0033 [hep-ph].

[109] Albert M. Sirunyan et al. "Measurements of the differential jet cross section as a function of the jet mass in dijet events from proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *JHEP* 11 (2018), p. 113. DOI: 10.1007/JHEP11(2018)113. arXiv: 1807.05974 [hep-ex].

[110] Albert M Sirunyan et al. "Identification of heavy, energetic, hadronically decaying particles using machine-learning techniques". In: (2020). DOI: 10.1088/1748-0221/15/06/p06005. arXiv: 2004.08262 [hep-ex].

[111] Jesse Thaler and Ken Van Tilburg. "Identifying Boosted Objects with N-subjettiness". In: *JHEP* 03 (2011), p. 015. DOI: 10.1007/JHEP03(2011)015. arXiv: 1011.2268 [hep-ph].

[112] Andrew J. Larkoski, Gavin P. Salam, and Jesse Thaler. "Energy Correlation Functions for Jet Substructure". In: *JHEP* 06 (2013), p. 108. DOI: 10.1007/JHEP06(2013)108. arXiv: 1305.0007 [hep-ph].

[113] Andrew J. Larkoski, Ian Moult, and Duff Neill. "Power Counting to Better Jet Observables". In: *JHEP* 12 (2014), p. 009. DOI: 10.1007/JHEP12(2014)009. arXiv: 1409.6298 [hep-ph].

[114] Ian Moult, Lina Necib, and Jesse Thaler. "New Angles on Energy Correlation Functions". In: (2016). arXiv: 1609.07483 [hep-ph].

[115] Huilin Qu and Loukas Gouskos. "ParticleNet: Jet Tagging via Particle Clouds". In: *Phys. Rev. D* 101.5 (2020), p. 056019. DOI: 10.1103/PhysRevD.101.056019. arXiv: 1902.08570 [hep-ph].

[116] "Identification of highly Lorentz-boosted heavy particles using graph neural networks and new mass decorrelation techniques". In: (2020). URL: http://cds.cern.ch/record/2707946.

[117] *Performance of deep tagging algorithms for boosted double quark jet topology in proton-proton collisions at 13 TeV with the Phase-0 CMS detector*. CMS Detector Performance Note CMS-DP-2018-046. 2018. URL: https://cds.cern.ch/record/2630438.

[118] James Dolen, Philip Harris, Simone Marzani, Salvatore Rappoccio, and Nhan Tran. "Thinking outside the ROCs: Designing Decorrelated Taggers (DDT) for jet substructure". In: (2016). arXiv: 1603.00027 [hep-ph].

[119] Chase Shimmin, Peter Sadowski, Pierre Baldi, Edison Weik, Daniel Whiteson, Edward Goul, and Andreas SÃÿgaard. "Decorrelated Jet Substructure Tagging using Adversarial Neural Networks". In: *Phys. Rev. D* 96.7 (2017), p. 074034. DOI: 10.1103/PhysRevD.96.074034. arXiv: 1703.03507 [hep-ex].

[120] J.A. Aguilar-Saavedra, Jack H. Collins, and Rashmish K. Mishra. "A generic anti-QCD jet tagger". In: *JHEP* 11 (2017), p. 163. DOI: 10.1007/JHEP11(2017)163. arXiv: 1709.01087 [hep-ph].

[121] Ian Moult, Benjamin Nachman, and Duff Neill. "Convolved Substructure: Analytically Decorrelating Jet Substructure Observables". In: *JHEP* 05 (2018), p. 002. DOI: 10.1007/JHEP05(2018)002. arXiv: 1710.06859 [hep-ph].

[122] *Performance of mass-decorrelated jet substructure observables for hadronic two-body decay tagging in ATLAS*. Tech. rep. ATL-PHYS-PUB-2018-014. Geneva: CERN, 2018. URL: https://cds.cern.ch/record/2630973.

[123] Albert M Sirunyan et al. "Search for Low Mass Vector Resonances Decaying to Quark-Antiquark Pairs in Proton-Proton Collisions at $\sqrt{s} = 13$ TeV". In: *Phys. Rev. Lett.* 119.11 (2017), p. 111802. DOI: 10.1103/PhysRevLett.119.111802. arXiv: 1705.10532 [hep-ex].

[124] J. Lin. "Divergence measures based on the Shannon entropy". In: *IEEE Transactions on Information Theory* 37.1 (1991), pp. 145–151.

[125] S. Kullback and R. A. Leibler. "On Information and Sufficiency". In: *Ann. Math. Statist.* 22.1 (1951), pp. 79–86. DOI: 10.1214/aoms/1177729694. URL: https://doi.org/10.1214/aoms/1177729694.

[126] E. Eichten, I. Hinchliffe, Kenneth D. Lane, and C. Quigg. "Super Collider Physics". In: *Rev. Mod. Phys.* 56 (1984), p. 579. DOI: 10.1103/RevModPhys.56.579.

[127] U. Baur, I. Hinchliffe, and D. Zeppenfeld. "Excited Quark Production at Hadron Colliders". In: *Int. J. Mod. Phys. A* 2 (1987), p. 1285. DOI: 10.1142/S0217751X87000661.

[128] Paul H. Frampton and Sheldon L. Glashow. "Chiral Color: An Alternative to the standard model". In: *Phys. Lett. B* 190 (1987), p. 157. DOI: 10.1016/0370-2693(87)90859-8.

[129] JoAnne L. Hewett and Thomas G. Rizzo. "Low-Energy Phenomenology of Superstring Inspired E(6) Models". In: *Phys. Rept.* 183 (1989), p. 193. DOI: 10.1016/0370-1573(89)90071-9.

[130] U. Baur, M. Spira, and P. M. Zerwas. "Excited quark and lepton production at hadron colliders". In: *Phys. Rev. D* 42 (1990), p. 815. DOI: 10.1103/PhysRevD.42.815.

[131] Lisa Randall and Raman Sundrum. "An Alternative to compactification". In: *Phys. Rev. Lett.* 83 (1999), p. 4690. DOI: 10.1103/PhysRevLett.83.4690. arXiv: hep-th/9906064 [hep-th].

[132] Schuyler Cullen, Maxim Perelstein, and Michael E. Peskin. "TeV strings and collider probes of large extra dimensions". In: *Phys. Rev. D* 62 (2000), p. 055012. DOI: 10.1103/PhysRevD.62.055012. arXiv: hep-ph/0001166 [hep-ph].

[133] Luis A. Anchordoqui, Haim Goldberg, Dieter Lust, Satoshi Nawata, Stephan Stieberger, and Tomasz R. Taylor. "Dijet signals for low mass strings at the LHC". In: *Phys. Rev. Lett.* 101 (2008), p. 241803. DOI: 10.1103/PhysRevLett.101.241803. arXiv: 0808.0497 [hep-ph].

[134] Tao Han, Ian Lewis, and Zhen Liu. "Colored Resonant Signals at the LHC: Largest Rate and Simplest Topology". In: *JHEP* 12 (2010), p. 085. DOI: 10.1007/JHEP12(2010)085. arXiv: 1010.4309 [hep-ph].

[135] R. Sekhar Chivukula, Arsham Farzinnia, Elizabeth H. Simmons, and Roshan Foadi. "Production of Massive Color-Octet Vector Bosons at Next-to-Leading Order". In: *Phys. Rev. D* 85 (2012), p. 054005. DOI: 10.1103/PhysRevD.85.054005. arXiv: 1111.7261 [hep-ph].

[136] CMS Collaboration. *Summary of the CMS Exotica results for 13 TeV data*. 2020. URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/SummaryPlotsEXO13TeV.

[137] Albert M Sirunyan et al. "Search for dark matter produced with an energetic jet or a hadronically decaying W or Z boson at $\sqrt{s} = 13$ TeV". In: *JHEP* 07 (2017), p. 014. DOI: 10.1007/JHEP07(2017)014. arXiv: 1703.01651 [hep-ex].

[138] Stefan Kallweit, Jonas M. Lindert, Philipp Maierhöfer, Stefano Pozzorini, and Marek Schönherr. "NLO electroweak automation and precise predictions for W+multijet production at the LHC". In: *JHEP* 04 (2015), p. 012. DOI: 10.1007/JHEP04(2015)012. arXiv: 1412.5157 [hep-ph].

[139] Stefan Kallweit, Jonas M. Lindert, Philipp Maierhöfer, Stefano Pozzorini, and Marek Schöherr. "NLO QCD+EW predictions for V+jets including off-shell vector-boson decays and multijet merging". In: *JHEP* 04 (2016), p. 021. DOI: 10.1007/JHEP04(2016)021. arXiv: 1511.08692 [hep-ph].

[140] Stefan Kallweit, Jonas M. Lindert, Stefano Pozzorini, Marek Schönherr, and Philipp Maierhöfer. "NLO QCD+EW automation and precise predictions for V+multijet production". In: *Proceedings, 50th Rencontres de Moriond, QCD and high energy interactions*. ARISF. ARISF, 2015, p. 121. arXiv: 1505.05704 [hep-ph]. URL: https://inspirehep.net/record/1372103/files/arXiv:1505.05704.pdf.

[141] J. M. Lindert et al. "Precise predictions for V + jets dark matter backgrounds". In: *Eur. Phys. J. C* 77 (2017), p. 829. DOI: 10.1140/epjc/s10052-017-5389-1. arXiv: 1705.04664 [hep-ph].

[142] A.M. Sirunyan et al. "Performance of reconstruction and identification of τ leptons decaying to hadrons and ν_τ in pp collisions at $\sqrt{s} = 13$ TeV". In: *JINST* 13.10 (2018), P10005. DOI: 10.1088/1748-0221/13/10/P10005. arXiv: 1809.02816 [hep-ex].

[143] R. A Fisher. "On the Interpretation of χ^2 from Contingency Tables, and the Calculation of P". In: *J. Roy. Statis. Soc.* 85 (1922), p. 87. ISSN: 09528385. DOI: 10.2307/2340521.

[144] Steve Baker and Robert D. Cousins. "Clarification of the use of CHI-square and likelihood functions in fits to histograms". In: *Nuclear Instruments and Methods in Physics Research* 221.2 (1984), pp. 437 –442. ISSN: 0167-5087. DOI: https://doi.org/10.1016/0167-5087(84)90016-4. URL: http://www.sciencedirect.com/science/article/pii/0167508784900164.

[145] CMS Collaboration. *CMS luminosity measurement for the 2017 data-taking period at* $\sqrt{s} = 13$ TeV. CMS Physics Analysis Summary. 2018. URL: https://cds.cern.ch/record/2621960.

[146] T. Junk. "Confidence level computation for combining searches with small statistics". In: *Nucl. Instrum. Meth. A* 434 (1999), p. 435. DOI: 10.1016/S0168-9002(99)00498-2. arXiv: hep-ex/9902006 [hep-ex].

[147] A. L. Read. "Presentation of search results: the CL_s technique". In: *J. Phys. G* 28 (2002), p. 2693. DOI: 10.1088/0954-3899/28/10/313.

[148] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. "Asymptotic formulae for likelihood-based tests of new physics". In: *Eur. Phys. J. C* 71 (2011), p. 1554. DOI: 10.1140/epjc/s10052-011-1554-0. arXiv: 1007.1727 [physics.data-an].

[149] Morad Aaboud et al. "Search for low-mass resonances decaying into two jets and produced in association with a photon using pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *Phys. Lett.* 795 (2019), p. 56. DOI: 10.1016/j.physletb.2019.03.067. arXiv: 1901.10917 [hep-ex].

[150] D. N. Spergel et al. "Wilkinson Microwave Anisotropy Probe (WMAP) three year results: implications for cosmology". In: *Astrophys. J. Suppl.* 170 (2007), p. 377. DOI: 10.1086/513700. arXiv: astro-ph/0603449 [astro-ph].

[151] P. A. R. Ade et al. "Planck 2013 results. XVI. Cosmological parameters". In: *Astron. Astrophys.* 571 (2014), A16. DOI: 10.1051/0004-6361/201321591. arXiv: 1303.5076 [astro-ph.CO].

[152] Mihailo Backovic, Kyoungchul Kong, and Mathew McCaskey. "MadDM v.1.0: Computation of Dark Matter Relic Abundance Using MadGraph5". In: *Phys. Dark Univ.* 5 (2014), p. 18. DOI: 10.1016/j.dark.2014.04.001. arXiv: 1308.4955 [hep-ph].

[153] Mihailo Backovic, Antony Martini, Olivier Mattelaer, Kyoungchul Kong, and Gopolang Mohlabeng. "Direct detection of dark matter with MadDM v.2.0". In: *Phys. Dark Univ.* 9 (2015), p. 37. DOI: 10.1016/j.dark.2015.09.001. arXiv: 1505.04190 [hep-ph].

[154] Tristan du Pree, Kristian Hahn, Philip Harris, and Christos Roskas. "Cosmological constraints on Dark Matter models for collider searches". 2016.

[155] Matthias Schlaffer, Michael Spannowsky, Michihisa Takeuchi, Andreas Weiler, and Chris Wymant. "Boosted Higgs Shapes". In: *Eur. Phys. J. C* 74 (2014), p. 3120. DOI: 10.1140/epjc/s10052-014-3120-z. arXiv: 1405.4295 [hep-ph].

[156] Massimiliano Grazzini, Agnieszka Ilnicka, Michael Spira, and Marius Wiesemann. "Effective Field Theory for Higgs properties parametrisation: the transverse momentum spectrum case". In: *52nd Rencontres de Moriond on QCD and high energy interactions*. 2017, p. 23. arXiv: 1705.05143 [hep-ph].

[157] Massimiliano Grazzini, Agnieszka Ilnicka, Michael Spira, and Marius Wiesemann. "Modeling BSM effects on the Higgs transverse-momentum spectrum in an EFT approach". In: *JHEP* 03 (2017), p. 115. DOI: 10.1007/JHEP03(2017)115. arXiv: 1612.00283 [hep-ph].

[158] You-Ying Li, Rosy Nicolaidou, and Stathes Paganis. "Exclusion of heavy, broad resonances from precise measurements of WZ and VH final states at the LHC". In: *Eur. Phys. J. C* 79.4 (2019), p. 348. DOI: 10.1140/epjc/s10052-019-6858-5. arXiv: 1904.03995 [hep-ph].

[159] J. M. Butterworth, A. R. Davison, M. Rubin, and G. P. Salam. "Jet Substructure as a New Higgs-Search Channel at the Large Hadron Collider". In: *Phys. Rev. Lett.* 100 (2008), p. 242001. DOI: 10.1103/PhysRevLett.100.242001. arXiv: 0802.2470 [hep-ph].

[160] Albert M Sirunyan et al. "Inclusive search for a highly boosted Higgs boson decaying to a bottom quark-antiquark pair". In: *Phys. Rev. Lett.* 120 (2018), p. 071802. DOI: 10.1103/PhysRevLett.120.071802. arXiv: 1709.05543 [hep-ex].

[161] *Inclusive search for a highly boosted Higgs boson decaying to a bottom quark-antiquark pair at* $\sqrt{s} = 13$ TeV *with* 137 fb^{-1}. Tech. rep. CMS-PAS-HIG-19-003. Geneva: CERN, 2020. URL: https://cds.cern.ch/record/2714916.

[162] D. de Florian et al. "Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector". In: (2016). arXiv: 1610.07922 [hep-ph]. URL: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHXSWG#SM_Higgs.

[163] Georges Aad et al. "Measurement of the Higgs boson mass from the $H \to \gamma\gamma$ and $H \to ZZ^* \to 4\ell$ channels with the ATLAS detector using 25 fb^{-1} of pp collision data". In: *Phys. Rev. D* 90.5 (2014), p. 052004. DOI: 10.1103/PhysRevD.90.052004. arXiv: 1406.3827 [hep-ex].

[164] Albert M Sirunyan et al. "Measurements of properties of the Higgs boson decaying into the four-lepton final state in pp collisions at $\sqrt{s} =$ 13 TeV". In: *JHEP* 11 (2017), p. 047. DOI: 10.1007/JHEP11(2017)047. arXiv: 1706.09936 [hep-ex].

[165] Nikolas Kauer and Giampiero Passarino. "Inadequacy of zero-width approximation for a light Higgs boson signal". In: *JHEP* 08 (2012), p. 116. DOI: 10.1007/JHEP08(2012)116. arXiv: 1206.4803 [hep-ph].

[166] Fabrizio Caola and Kirill Melnikov. "Constraining the Higgs boson width with ZZ production at the LHC". In: *Phys. Rev.* D88 (2013), p. 054024. DOI: 10.1103/PhysRevD.88.054024. arXiv: 1307.4935 [hep-ph].

[167] John M. Campbell, R. Keith Ellis, and Ciaran Williams. "Bounding the Higgs width at the LHC using full analytic results for $gg- > e^-e^+\mu^-\mu^+$". In: *JHEP* 04 (2014), p. 060. DOI: 10.1007/JHEP04(2014)060. arXiv: 1311.3589 [hep-ph].

[168] John M. Campbell, R. Keith Ellis, and Ciaran Williams. "Bounding the Higgs width at the LHC: Complementary results from $H \to WW$". In: *Phys. Rev.* D89.5 (2014), p. 053011. DOI: 10.1103/PhysRevD.89.053011. arXiv: 1312.1628 [hep-ph].

[169] Christoph Englert, Yotam Soreq, and Michael Spannowsky. "Off-Shell Higgs Coupling Measurements in BSM scenarios". In: *JHEP* 05 (2015), p. 145. DOI: 10.1007/JHEP05(2015)145. arXiv: 1410.5440 [hep-ph].

[170] Christoph Englert, Matthew McCullough, and Michael Spannowsky. "Combining LEP and LHC to bound the Higgs Width". In: *Nucl. Phys.* B902 (2016), pp. 440–457. DOI: 10.1016/j.nuclphysb.2015.11.017. arXiv: 1504.02458 [hep-ph].

[171] John M. Campbell, R. Keith Ellis, Elisabetta Furlan, and Raoul RÃűntsch. "Interference effects for Higgs boson mediated Z-pair plus jet production". In: *Phys. Rev.* D90.9 (2014), p. 093008. DOI: 10.1103/PhysRevD.90.093008. arXiv: 1409.1897 [hep-ph].

[172] John M. Campbell, R. Keith Ellis, and Ciaran Williams. "Bounding the Higgs width at the LHC: Complementary results from $H \to WW$". In: *Phys. Rev. D* 89 (5 2014), p. 053011. DOI: 10.1103/PhysRevD.89.053011. URL: https://link.aps.org/doi/10.1103/PhysRevD.89.053011.

[173] Christoph Englert and Michael Spannowsky. "Limitations and opportunities of off-shell coupling measurements". In: *Phys. Rev. D* 90 (5 2014), p. 053003. DOI: 10.1103/PhysRevD.90.053003. URL: https://link.aps.org/doi/10.1103/PhysRevD.90.053003.

[174] Stephen P. Martin. "Shift in the LHC Higgs Diphoton Mass Peak from Interference with Background". In: *Phys. Rev.* D86 (2012), p. 073016. DOI: 10.1103/PhysRevD.86.073016. arXiv: 1208.1533 [hep-ph].

[175] Stephen P. Martin. "Interference of Higgs diphoton signal and background in production with a jet at the LHC". In: *Phys. Rev.* D88.1 (2013), p. 013004. DOI: 10.1103/PhysRevD.88.013004. arXiv: 1303.3342 [hep-ph].

[176] Daniel de Florian, Nerina Fidanza, R. J. Hernandez-Pinto, Javier Mazzitelli, Yamila Rotstein Habarnau, and German F. R. Sborlini. "A complete $O(\alpha_S^2)$ calculation of the signal-background interference for the Higgs diphoton decay channel". In: *Eur. Phys. J.* C73.4 (2013), p. 2387. DOI: 10.1140/epjc/s10052-013-2387-9. arXiv: 1303.1397 [hep-ph].

[177] Lance J. Dixon and Ye Li. "Bounding the Higgs Boson Width Through Interferometry". In: *Phys. Rev. Lett.* 111 (2013), p. 111802. DOI: 10.1103/PhysRevLett.111.111802. arXiv: 1305.3854 [hep-ph].

[178] Albert M Sirunyan et al. "Measurements of the Higgs boson width and anomalous HVV couplings from on-shell and off-shell production in the four-lepton final state". In: *Phys. Rev.* D99.11 (2019), p. 112003. DOI: 10.1103/PhysRevD.99.112003. arXiv: 1901.00174 [hep-ex].

[179] M. Cepeda et al. "Higgs Physics at the HL-LHC and HE-LHC". In: (2019). arXiv: 1902.00134 [hep-ph].

[180] Alexander Conway and Hans Wenzel. "Higgs Measurements at a Muon Collider". In: (2013). arXiv: 1304.5270 [hep-ex].

[181] Tao Han, Zhen Liu, and Josh Sayre. "Potential Precision on Higgs Couplings and Total Width at the ILC". In: *Phys. Rev.* D89.11 (2014), p. 113006. DOI: 10.1103/PhysRevD.89.113006. arXiv: 1311.7155 [hep-ph].

[182] Mark Thomson. "Model-independent measurement of the $e^+ e^- \to HZ$ cross section at a future $e^+ e^-$ linear collider using hadronic Z decays". In: *Eur. Phys. J.* C76.2 (2016), p. 72. DOI: 10.1140/epjc/s10052-016-3911-5. arXiv: 1509.02853 [hep-ex].

[183] A. M. Sirunyan et al. "Observation of Higgs boson decay to bottom quarks". In: *Phys. Rev. Lett.* 121.12 (2018), p. 121801. DOI: 10.1103/PhysRevLett.121.121801. arXiv: 1808.08242 [hep-ex].

[184] Morad Aaboud et al. "Observation of $H \to b\bar{b}$ decays and VH production with the ATLAS detector". In: *Phys. Lett.* B786 (2018), pp. 59–86. DOI: 10.1016/j.physletb.2018.09.013. arXiv: 1808.08238 [hep-ex].

[185] Albert M. Sirunyan et al. "Measurements of properties of the Higgs boson decaying to a W boson pair in pp collisions at $\sqrt{s} = 13$ TeV". In: *Phys. Lett.* B791 (2019), p. 96. DOI: 10.1016/j.physletb.2018.12.073. arXiv: 1806.05246 [hep-ex].

[186] Morad Aaboud et al. "Measurement of VH, $H \to b\bar{b}$ production as a function of the vector-boson transverse momentum in 13 TeV pp collisions with the ATLAS detector". In: *JHEP* 05 (2019), p. 141. DOI: 10.1007/JHEP05(2019)141. arXiv: 1903.04618 [hep-ex].

[187] *Search for boosted resonances decaying to two b-quarks and produced in association with a jet at $\sqrt{s} = 13$ TeV with the ATLAS detector*. Tech. rep. ATLAS-CONF-2018-052. Geneva: CERN, 2018. URL: https://cds.cern.ch/record/2649081.

[188] CMS Collaboration. "Sensitivity projections for Higgs boson properties measurements at the HL-LHC". In: (2018).

[189] Yang-Ting Chien and Iain W. Stewart. "Collinear Drop". In: (2019). arXiv: 1907.11107 [hep-ph].

[190] Gilles Louppe, Kyunghyun Cho, Cyril Becot, and Kyle Cranmer. "QCD-Aware Recursive Neural Networks for Jet Physics". In: *JHEP* 01 (2019), p. 057. DOI: 10.1007/JHEP01(2019)057. arXiv: 1702.00748 [hep-ph].

[191] Junyoung Chung, Çaglar Gülçehre, KyungHyun Cho, and Yoshua Bengio. "Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling". In: *CoRR* abs/1412.3555 (2014). arXiv: 1412.3555. URL: http://arxiv.org/abs/1412.3555.

[192] Manqi Ruan. "Higgs Measurement at e^+e^- Circular Colliders". In: *Nucl. Part. Phys. Proc.* 273-275 (2016), pp. 857–862. DOI: 10.1016/j.nuclphysbps.2015.09.132. arXiv: 1411.5606 [hep-ex].

[193] Alain Blondel and Patrick Janot. "Future strategies for the discovery and the precise measurement of the Higgs self coupling". In: (2018). arXiv: 1809.10041 [hep-ph].

[194] A. Blondel, J. Gluza, S. Jadach, P. Janot, and T. Riemann, eds. *Theory report on the 11th FCC-ee workshop*. 2019. arXiv: 1905.05078 [hep-ph].

[195] David Curtin et al. "Long-Lived Particles at the Energy Frontier: The MATHUSLA Physics Case". In: (2018). arXiv: 1806.07396 [hep-ph].

Cristina Mantilla Suarez
cmantil1@jhu.edu

Education

2016 - 2020	**Johns Hopkins University** Department of Physics, Ph.D. Thesis: Probing new physics using Initial State Radiation Jets at the LHC Advisor: Petar Maksimovic
2010 - 2015	**Escuela Politecnica Nacional (Ecuador)** B.S. Physics Senior thesis: Measurement of the top quark mass with lepton kinematics Advisor: Edgar Carrera

Research Experience

PhD student, Johns Hopkins University August 2016 - Present

Probing new physics with boosted jets. Pioneered searches for exotic, light spin-1 and spin-0 particles decaying to quarks in CMS.

- Led a novel search for dark sector mediators decaying to quarks ($Z' \to qq'$) [2,4,5], that sets the most stringent constraints for masses less than 350 GeV.
  Performed signal sensitivity studies, processed the data, defined and validated the background estimation method, and extracted the Z' signal.
- Designed deep learning and jet substructure techniques to identify the radiation pattern of W/Z/Z' and Higgs jets and validated its tagging performance in data.
  Pioneered a novel and most optimal decorrelation scheme to allow the preservation of the jet mass shape [3].

Boosted Higgs to $\bar{b}b$. Contributing author to the first search for $h \to \bar{b}b$ decays using 2016 data [6] and full Run-2 data (in progress). Processed data samples and validated jet mass scale and resolution in data.

Boosted Higgs to WW^*. Designed a deep learning tagger that distinguishes leptons clustered inside jets to identify $h \to WW^*(\ell\nu qq')$ (validation in data in progress).

Boosted Higgs (inclusive). Led a novel sensitivity study to constrain the Higgs width at the LHC and HL-LHC using boosted Higgs decays [1]. This analysis introduces new approaches from machine learning and a modified jet reconstruction.

CMS Pixel Hit Reconstruction. Development and maintenance of a detailed offline simulation that accounts for radiation damage in the tracker detector.
Validated a method that improves the estimate of the reconstructed hit position for clusters with edge pixel hits.

CMS Phase-2 Inner Tracker Upgrade Performed preliminary performance measurements of small pitch planar pixel sensors at the Fermilab test beam facility.
Implemented the simulation of the new pixel readout in the offline reconstruction.

Triggering on dark photon signals at SpinQuest Optimized trigger logic to distinguish displaced dimuon decays and reduce background rate, using signal simulation and 2017 data.

Visitor Student, Fermilab August 2015 - August 2016

Dark Matter search with boosted jets Characterized events from a DM signal produced in association with a single boosted top quark. Validated tracking variables that identify bottom quarks inside a merged top jet.

Big Data approaches at LHC Developed an approach for a faster LHC data analysis based on Apache Spark techologies.

Pixel Phase I upgrade Tested functionality of pixel sensor modules and the boards that handle the communication with readout chips.

Summer Student, CERN June 2016 - September 2016
Alternative top mass measurement: Led first measurement of the top quark mass using unfolded lepton kinematics [7].

Selected Publications

1. O. Amram *et.al*, "Tag N' Train: A Technique to Train Improved Classifiers on Unlabeled Data", 2020, arXiv:2002.12376.
2. C. Mantilla Suarez *et.al*, "An approach to constraining the Higgs width at the LHC and HL-LHC", 2019, arXiv:1910.02082.
3. CMS Collaboration, 2019, "Search for low mass vector resonances decaying into quark-antiquark pairs in proton-proton collisions at $\sqrt{s} = 13$ TeV", submitted to PRD arXiv:1909.04114.
4. CMS Collaboration, 2019, "Machine learning-based identification of highly Lorentz-boosted hadronically decaying particles at the CMS experiment", CMS-PAS-JME-18-002 In preparation.
5. CMS Collaboration, 2017, "Search for light vector resonances decaying to a quark pair produced in association with a jet in proton-proton collisions at $\sqrt{s} = 13$ TeV", JHEP 01 (2018) 097.
6. CMS Collaboration, 2017, "Search for light vector resonances decaying to quarks at $\sqrt{s} = 13$ TeV", Phys. Rev. Lett. **118** (2017) 111802.
7. CMS Collaboration, 2017, "Inclusive search for the standard model Higgs boson produced in pp collisions at $\sqrt{s} = 13$ TeV using H$\rightarrow$ b$\bar{\text{b}}$ decays," Phys. Rev. Lett. 120, 071802.
8. CMS Collaboration, 2016, "Determination of the top quark mass from leptonic observables using $e\mu$ + jets final states selected in proton-proton collisions at $\sqrt{s} = 8$ TeV" CMS-PAS-TOP-16-002.

Talks at Conferences and Workshops

- Dark Photon detector status, **SpinQuest collaboration meeting** Fall 2019, Fermilab, IL, October 24, 2019.
- An approach to constrain the Higgs Width, lightning plenary talk at **BOOST 2019**, Boston, MA, July 21-26 2019.
- QCD in BSM searches at CMS, plenary talk at **QCD@LHC 2019**, Buffalo, NY, July 15-19 2019.
- CMS in 10 min, plenary talk at **New Perspectives 2019**, Fermilab, June 10 2019.
- ISR searches in CMS, lightning talk at **USCMS meeting 2019**, Washington, DC, June 5-8 2019.
- Diboson resonance searches at CMS, parallel talk at **SUSY 2019**, Corpus Christi, TX, May 20-24 2019.
- Diboson resonance searches at CMS, parallel talk at **Pheno 2019**, Pittsburgh, PA, May 6-8 2019.

- Machine Learning applications to jets @ CMS, parallel talk at **APS 2019**, Denver, CO, Apr 13-16 2019.
- Silicon Pixel Sensor Simulation in the CMS Monte Carlo Framework, invited talk at **Radiation effects at LHC**, CERN, CH, Feb 11-12 2019,
- Exotic Searches at LHC, plenary talk at **PIC 2018**, Bogota, CO. September 2018.
- Tagger performance in Data, plenary talk at **BOOST 2018**. Paris, FR. July 2018.
- Probing low mass resonances with boosted jets, lightning round at **USLUA LHC meeting**, Fermilab, US. Nov. 2017.
- Search for low mass resonances produced in association with an ISR jet in CMS. plenary talk at **DPF-APS Meeting 2017**. Fermilab, US. July 2017.
- Searches for light dijet resonances with the CMS experiment. plenary talk at **New Perspectives 2017**. Fermilab, US. June 2017.
- W/Z tagging techniques. invited plenary talk at **Jet Substructure Workshop** Fermilab, IL, US. November 2016.

Invited Seminars

- Constraining the Higgs total width with boosted Higgs, **JHU Particle Physics Seminar**, Baltimore, MD, Oct 30 2019.
- Dijet searches in CMS, **LPC Physics Forum**, Fermilab, June 20 2019.
- Probing new physics with boosted jets, **JHU Physics and Astronomy Advisory Council meeting**. Baltimore, MD, June 2018.

Poster Presentations

- Boosting hadronic resonances at CMS. **ESHEP 2019**. St. Petersburgh, Russia. Sep 2019.
- Search for boosted low mass resonances produced in association with a ISR jet at 13 TeV. **BOOST 2017**. Buffalo, US. July 2017.
- Measurement of the top quark mass from leptonic observables in pp collisions. **ICHEP 2016**. Chicago, US. August 2016.
- Search for dark matter in association with a boosted top quark in the hadronic final state. **Fermilab Users Meeting 2016**. Fermilab, US. June 2016.

Teaching and Outreach

Special Relativity and Waves, JHU	Fall 2018
General Physics II and Laboratory II/I, JHU	Fall 2016-Spring 2018
Optics and Atomic Physics, Linear Algebra I/II, EPN	Fall 2013-2015
Fermilab Saturday Morning Physics, On-site coordinator	2019
Fermilab Ask-a-Scientist physicist	2019
Letters to a Pre-Scientist volunteer	2019
QuarkNet videoconferences with high school students, moderator	2017-Present
LHC World Wide Data Day moderator	2018-2019
Johns Hopkins Physics Fair and 2018 DC Science Fair.	2017-2018
CMS Data Analysis School, facilitator of long and short exercises	2017-2019
CLAVEMAT tutor of math and calculus, directed to freshman students	2013-2015

Student Supervision

Graduate Students

- Jeff Krupa (MIT). Boosted Z' search with 2017 data, development and calibration of tagging variables for 2-prong identification. *2018-Present.*

Undergraduate and Summer Students

- Alex Bisnath (U. Chicago). Development of a deep learning algorithm to distinguish W(cs) vs Z(cc) decays. *2019*.
- Dylan Russell (Reed College). Optimization of a trigger algorithm to distinguish dark photon dimuon decays. *2019*.
- Minqi Ma, Evan Petroski (JHU). Development of a lepton-in-jet tagger. *2018*.

Honors and Awards

LPC Graduate Scholar, Fermilab	2019
URA Visiting Scholars Program, Fermilab	Fall 2017-2018
Distinguished Poster Award, ICHEP 2016	August 2016
LPC Guest and Visitors Program, Fermilab	Sep 2015
Dean's List, EPN	2013-2015

Professional Training

CERN-JINR, European School of High Energy Physics	2019
CERN, Latin American School of High Energy Physics	2015
CERN, Summer Student Program	2015
Utrecht, Summer School on Theoretical Physics	2014