

# UNIVERSITÀ DI PISA

Dipartimento di Fisica  
Corso di Laurea Magistrale in Fisica

## Measurement of the $|V_{cb}|$ CKM element from W hadronic decays at CMS

Relatore:  
Paolo Azzurri

Candidato:  
Piero Viscone



## Abstract

The precise determination of the  $|V_{cb}|$  Cabibbo-Kobayashi-Maskawa (CKM) element is of great interest as it currently is the CKM element with the largest relative uncertainty, playing a crucial role in present CKM unitarity tests.

Existing measurements of  $|V_{cb}|$  rely on semileptonic decays of B hadrons, but these determinations are hampered by non-perturbative Quantum Chromodynamics (NP-QCD) theoretical uncertainties, rendering further refinements challenging.

Most of all, there is a large ( $\sim 3\sigma$ ) observed tension between inclusive and exclusive B hadrons determinations, often referred to as the  $|V_{cb}|$  puzzle.

A determination of  $|V_{cb}|$  through W decays would allow us to tackle the  $|V_{cb}|$  puzzle, providing a direct measurement at the natural energy scale of the W boson, with an approach that is free from NP-QCD theoretical uncertainties. The  $|V_{cb}|$  element can be accessed by measuring the ratio between the branching fraction of the  $W \rightarrow cb$  decay over the total hadronic decay branching fraction

$$\frac{\Gamma(W \rightarrow cb)}{\Gamma(W \rightarrow qq)} = \frac{|V_{cb}|^2}{2}$$

Conducting inclusive measurements of hadronic W decays at a Hadron collider like the Large Hadron Collider (LHC) is impossible due to the large QCD background.

To address this problem, semileptonic decays of W pairs can be employed, characterized by the presence of one lepton in the final state. This distinctive signature enables the clean and efficient triggering of events allowing us to measure the  $W \rightarrow cb$  rate exploiting the hadronic decay of the second W boson. The primary source of W pairs at the LHC arises from the decay of top-antitop quark pairs, hence the analysis will be focused on the selection of semileptonic decays of top pairs.

At LHC, in all the RunII, there were produced almost 40 thousand  $W \rightarrow cb$  decays originating from semileptonic  $t\bar{t}$  decays, leading to a potential statistical uncertainty of 0.5% on  $|V_{cb}|$  in the ideal case of a perfect event selection. Since the difference between inclusive and exclusive determinations through B mesons decays is 7%, the direct measurement of  $|V_{cb}|$  at LHC is a promising approach to solve the  $|V_{cb}|$  puzzle.

In this work, prospects for a direct measure of  $|V_{cb}|$  at CMS with the RunII luminosity are provided. To enhance the precision of our determination, cutting-edge machine learning techniques, such as attention networks, are integrated into the analysis, to efficiently identify and classify events, separating the  $W \rightarrow cb$  signal from the backgrounds.

In the end, a template fit on the Neural network score using the Asimov dataset will be presented to estimate the uncertainties of such a measure.

This determination will be limited mainly by the b and c tagging systematic uncertainties, but the ratio of branching fractions should allow some cancellations.

The thesis is structured as follows:

**Chapter ??** provides an overview of the Standard model, with emphasis on the

CKM matrix and the production of top pairs at LHC.

**Chapter ??** is dedicated to the presentation of the current determination of  $|V_{cb}|$  with B hadrons and of the direct measure at LHC.

**Chapters ??** and ?? describe respectively the CMS detector and the algorithms employed to reconstruct the events.

**Chapter ??** delves into the statistical methods used in the analysis, in particular, the profile likelihood method and the neural networks.

In **Chapter ??** the selection of physics objects and events is discussed, along with the signal signatures and the backgrounds, and the respective Monte Carlo samples employed in the analysis.

In **Chapter ??**, the jet-parton assignment problem is tackled to provide an interpretation of the signal events.

Finally, in **Chapter ??**, the multivariate model which had the higher accuracy on the jet-parton assignment problem is adapted to perform the signal-background discrimination. The profile likelihood method is employed to assess the statistical and systematic uncertainties on  $|V_{cb}|$ .

---

# Contents

---

# *Chapter 1*

---

## Theoretical framework

---

### 1.1 The Standard Model of particle physics

The Standard Model of particle physics (SM) is a quantum field theory (QFT) that describes all known fundamental particles and forces besides gravity: the electromagnetic, the strong, and the weak force.

The QFT is a framework that combines quantum mechanics, classical field theory, and special relativity, allowing us to describe a particle as the excitation of a quantum field or, more formally, with an irreducible representation of the Poisson group that is characterized by the mass, spin, and additional quantum numbers of the particle. Such theories are characterized by a Lagrangian density  $\mathcal{L}$  that must be Lorentz invariant, must respect locality, and must be renormalizable, i.e. the perturbative expansions of the amplitudes must converge.

In the SM, additional local symmetries (Gauge symmetries) are imposed to describe the interactions between particles and, according to the Noether theorem [?], for each global symmetry a conserved charge arises.

#### **Particles of the Standard Model**

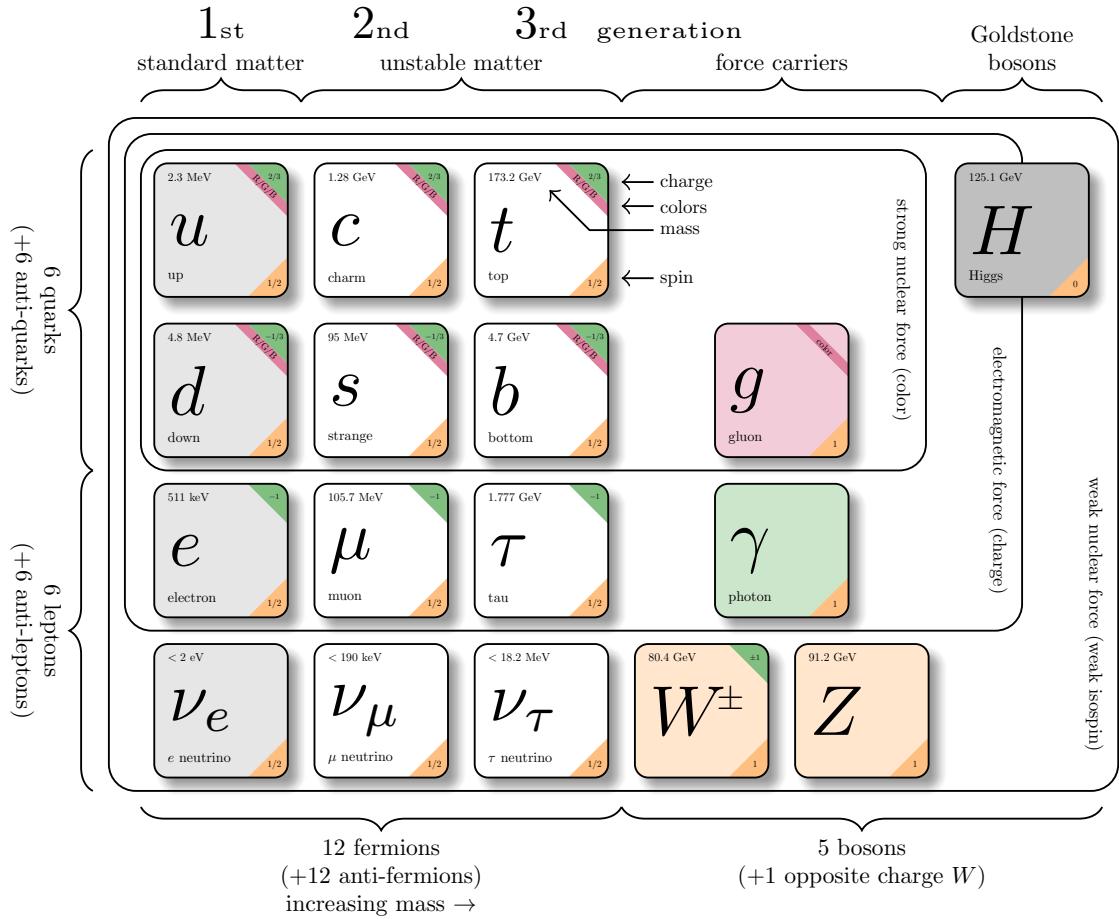
Particles are divided into two main categories: fermions and bosons. Fermions are particles with half-integer spin that obey the Fermi-Dirac statistics and the Pauli principle. Furthermore, each fermion has an antiparticle with the same mass but opposite quantum numbers.

Bosons are particles with integer spin that obey the Bose-Einstein statistics. In the SM there are the vector bosons (spin 1), the mediators of the forces, and the Higgs boson, the only scalar boson of the SM (spin 0). All the particles of the SM and their properties are summarized in Fig. ??.

**Fermions** Fermions are divided into leptons and quarks, both divided into 3 generations, for a total of 12 fermions.

*Quarks* are the constituents of the hadrons. They have a fractional electric charge and a color charge (or an anticolor charge for the antiquarks).

Quarks are divided into up-type quarks (up, charm, and top) with an electric charge of  $Q = 2/3$  and into down-type quarks (down, strange, and bottom) with an electric charge



**Figure 1.1:** Fundamental particles of the Standard Model [? ]

of  $Q = -1/3$ . One of the main properties of the QCD is color confinement: isolated particles must be color singlets, so free quarks can't be observed because they hadronize (see sec. ??). The only exception is the top quark that decays before the hadronization due to its short lifetime.

Furthermore, weak interactions don't couple identically with all the flavors: charged weak interactions mix the flavor of the down-type quarks according to the unitary CKM matrix, while neutral weak interactions have different couplings that depend on the weak isospin of the fermion.

*Leptons* are the electron ( $e$ ), the muon ( $\mu$ ), the tau ( $\tau$ ) and their respective neutrinos ( $\nu_e$ ,  $\nu_\mu$ ,  $\nu_\tau$ ). A charged lepton and its respective neutrino represent a generation of leptons with a corresponding family lepton flavor number  $L_{\ell_i} = n_{\ell_i} - n_{\bar{\ell}_i}$  that, in SM, is conserved together with the total lepton number  $L = n_\ell - n_{\bar{\ell}}$ . The leptons participate in all the interactions except the strong force.

### 1.1.1 Gauge theories and the QED

**Fermion free fields** The Lagrangian (density) for a free fermion field is the Dirac Lagrangian

$$\mathcal{L}_{\text{Dirac}} = i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi \quad (1.1)$$

where  $\bar{\psi} = \psi^\dagger\gamma^0$  is the adjoint Dirac spinor, and  $\gamma^\mu$  are the 4 Dirac matrices that obey the Clifford algebra  $\{\gamma^\mu, \gamma^\nu\} = 2\eta^{\mu\nu}$ .

The corresponding equation of motion has two solutions: the one with the positive energy represents the fermion field, and the one with the negative energy represents the anti-fermion field.

Dirac spinors are representations of the Lorentz group  $SO^+(1, 3)$  that can be decomposed into the  $SU(2)_L \otimes SU(2)_R$  group.

Using this different representation, Dirac spinors can be described in the so-called chiral basis  $\psi = (\psi_L, \psi_R)$  where  $\psi_{L/R}$  are called left/right Weyl spinors and are respectively the  $(\frac{1}{2}, 0)$  and the  $(0, \frac{1}{2})$  representations of the  $SU(2) \otimes SU(2)$  group.

The chirality operator  $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$  anti-commutes with all the gamma matrices and commutes with the Dirac Hamiltonian only if the mass is zero, so the chirality is a good quantum number only for massless particles.

The chirality plays an important role in the weak interaction, as described in sec. ??

**Gauge theories** The standard procedure to introduce a global symmetry in the Lagrangian and a Gauge boson field is the following:

1. Introduce a local symmetry based on a Lie group:

$$\psi(x) \rightarrow \psi'(x) = g(x)\psi(x) = e^{i\alpha_a(x)t^a}\psi(x) \quad (1.2)$$

where  $g(x)$  is a representation of the group and  $t^a$  its generators

2. Replace the partial derivatives  $\partial_\mu$  with the covariant derivative  $D_\mu$

$$D_\mu\psi(x) = (\partial_\mu - ig A_\mu^a(x)t^a)\psi(x) \quad (1.3)$$

where  $g$  is the coupling constant between the fermion field and the Gauge fields and  $A_\mu^a$  the Gauge fields

3. The covariant derivative should transform according to the Gauge group  $D_\mu\psi(x) \rightarrow g(x)D_\mu\psi(x)$  and, with some calculation, it is possible to find that the Gauge field transform as the following under the action of the Gauge group

$$A_\mu(a) \rightarrow g(x)A_\mu(x)g^{-1}(x) - \frac{i}{g} \left( \partial_\mu g(x) \right) g^{-1}(x) \quad (1.4)$$

where  $A_\mu = A_\mu^a(x)t^a$

4. The final Lagrangian now has a global invariance for the Gauge group and, according to the Noether theorem, there is a conserved charge. Furthermore, in this procedure, another vector massless field arises.

**Quantum electrodynamics** The simplest case of a Gauge theory is the quantum electrodynamics (QED) [? ] in which the Gauge group is U(1), a commutative group for which holds  $g(x)g(y) = g(y)g(x)$

The fundamental representation of the U(1) group is  $g(x) = e^{i\alpha(x)}$ , i.e. a local phase.

Imposing the U(1) invariance on the Dirac Lagrangian, we obtain the QED Lagrangian.

$$\mathcal{L}_{\text{QED}} = i\bar{\psi}\gamma^\mu(\partial_\mu + iqA_\mu)\psi - m\bar{\psi}\psi - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} \quad (1.5)$$

where  $A_\mu$  is the photon field,  $q$  the elementary electron charge and  $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$  the electromagnetic tensor

In the SM, all the fermions, except the neutrinos, have an electric charge and are involved in the electromagnetic interactions.

### 1.1.2 Quantum chromodynamics

Quantum chromodynamics (QCD) is the SU(3) Gauge theory that describes the strong interaction and involves gluons and quarks, bonding together into hadronic states. The quantum number related to strong interaction is the color charge. Applying the gauging technique described in the section ?? is possible to obtain the QCD Lagrangian.

$$\mathcal{L}_{\text{QCD}} = -ig_s\bar{\psi}\gamma^\mu\lambda_a\psi G_\mu^a - \frac{1}{4}G_a^{\mu\nu}G_{\mu\nu}^a \quad (1.6)$$

where  $G_{\mu\nu}^a = \partial_\mu G_\nu^a - \partial_\nu G_\mu^a - g_s f^{abc}G_\mu^b G_\nu^c$  is the corresponding of the electromagnetic tensor for the strong force,  $\lambda_a$  are the SU(3) generators, and  $f^{abc}$  the SU(3) structure constants. Each quark carries a color charge and each anti-quark an anticolor charge and, since SU(3) has 8 generators, 8 gluons exist (or just one gluon that can carry one of 8 possible couples of color-anticolor).

There are significant differences with the QED due to the non-abelianity of SU(3):

- **Gluon self-coupling:** gluons can interact with each other.
- **Asymptotic freedom:** the coupling strength decreases with the energy, i.e. at large energy scales the quarks become quasi-free.

The coupling constant of the strong interaction depends on the energy scale [? ]

$$\alpha_s(\mu^2) \propto \frac{1}{\ln\left(\frac{\mu^2}{\Lambda_{\text{QCD}}^2}\right)} \quad (1.7)$$

Where  $\Lambda_{\text{QCD}} \approx 200 \text{ MeV}$  and  $\mu$  is the energy scale.

When  $\mu \gg \Lambda_{\text{QCD}}$ , strong interactions can be described using perturbative techniques (pQCD) but, at low-energy scales, the QCD is non-perturbative and that leads to the third property of the QCD

- **Color confinement:** At low-energy scales, the strong coupling is so large that quarks are always bounded in color singlets and isolated quarks can't be observed. This is also why the strong force is a short-range interaction, while the electromagnetic force is a long-range interaction.

### 1.1.3 Electroweak unification

**Weak interactions** The weak interaction was first introduced by Enrico Fermi to explain the  $\beta$  decay through the 4-point Fermi interaction [? ] and is the only force that involves all the fermions and that violates the parity and the charge parity conservation.

The quantum number related to the weak interaction is the weak isospin, generated by the  $SU(2)$  gauge group that creates 3 gauge bosons: the  $W^\pm$  and the  $Z$  bosons.

Trying to build the weak Lagrangian only using the gauging technique described in the section ?? doesn't work in this case for different reasons:

- The  $W$  boson and the  $Z$  boson would have the same coupling to the fermions, while experimentally this was disproven. To address this problem, the weak interaction was unified with the electromagnetic interaction.

- The  $W/Z$  bosons are proven to be massive bosons, while gauge bosons are always massless. The massive term can't be added arbitrarily to the Lagrangian because such a theory would be non-renormalizable, so they have to be added through the Higgs mechanism as described in the section ??.

The non-zero  $W/Z$  mass is also the cause of why the weak interactions have a finite range, given that the weak interactions are suppressed by the weak propagator  $1/(q^2 - M_{W/Z}^2)$ .

- The weak interaction doesn't couple identically with all the quarks, but the down-type flavors are mixed by the Cabibbo-Kobayashi-Maskawa (CKM) matrix [? ], as described in section ???. The elements of the CKM matrix are complex, and this leads to the CP violation.

- To introduce the parity violation in the Lagrangian, the  $\gamma^\mu$  in the interaction term of the Lagrangian has to be replaced with  $\gamma^\mu (1 - \gamma^5)/2$ .

The term  $(1 - \gamma^5)/2$  is the left chirality projector, and this implies that weak interaction couples only with the left part of the spinor. This will turn out to be true only for charged weak interactions, while for neutral weak interactions, the mixing between the photon and the  $Z$  boson in the electroweak (EW) unification process will reintroduce the coupling with the right spinors.

**Electroweak unification** The electromagnetic and weak interactions were unified in the Glashow-Weinberg-Salam theory [? ? ? ] based on the  $SU(2) \otimes U(1)$  symmetry group. We start defining left-handed doublets and right-handed singlets

$$Q_{L,i} = \begin{pmatrix} u_{L,i} \\ d_{L,i} \end{pmatrix}, \quad u_{R,i}, \quad d_{R,i} \quad (1.8)$$

$$\mathcal{L}_{L,i} = \begin{pmatrix} \nu_{L,i} \\ \ell_{L,i} \end{pmatrix}, \quad \ell_{R,i} \quad (1.9)$$

so left-handed particles organize in  $SU(2)$  doubles while right-handed particles in  $SU(2)$  singlets. The gauge invariant Lagrangian for the group  $SU(2) \otimes U(1)$  is:

$$\mathcal{L} = i\bar{L}\gamma^\mu D_\mu L + \bar{R}i\gamma^\mu D_\mu R - \frac{1}{4}B^{\mu\nu}B_{\mu\nu} - \frac{1}{4}W_a^{\mu\nu}W_{\mu\nu}^a \quad (1.10)$$

where

- L and R are the left doublets and the right singlets
- $D_\mu = \partial_\mu - i\frac{g'}{2}\hat{Y}B_\mu - ig\hat{T}_iW_\mu^i$  is the covariant derivative
- $\hat{Y}$  is the U(1) generator (the hypercharge operator)
- $\hat{T}_i$  are the three SU(2) generators (weak isospin operators)
- $B^{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu$ , where  $B_\mu$  is the U(1) gauge field
- $W_{\mu\nu}^i = \partial_\mu W_\nu^i - \partial_\nu W_\mu^i - g\epsilon^{ijk}W_\mu^jW_\nu^k$  where  $W_\mu^i$  are the 3 SU(2) gauge fields and  $\epsilon^{ijk}$  the SU(2) structure constants.

The experimental observed W, Z and  $\gamma$  bosons can be expressed as a linear combination of the defined gauge fields.

$$W_\mu^\pm = (W_\mu^1 \pm iW_\mu^2)/\sqrt{2} \quad (1.11)$$

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} B_\mu \\ W_\mu^3 \end{pmatrix} \quad (1.12)$$

where the mixing angle  $\theta_W$  is the Weinberg angle, and it can be proven that  $\cos(\theta_W) = m_W/m_Z$ .

After the mixing, we can break the electroweak Lagrangian into different pieces:

$$\mathcal{L}_{\text{EWK}} = \mathcal{L}_{\text{kin}} + \mathcal{L}_{\text{QED}} + \mathcal{L}_{\text{wcc}} + \mathcal{L}_{\text{wnc}} + \mathcal{L}_{\text{gauge}} \quad (1.13)$$

- $\mathcal{L}_{\text{kin}}$  is the fermions kinetic term
- $\mathcal{L}_{\text{QED}}$  is the interaction term of the QED Lagrangian described in section ??
- $\mathcal{L}_{\text{wcc}}$  is the interaction term of the weak-charged currents

$$\mathcal{L}_{\text{wcc}} = \frac{g}{2\sqrt{2}} \{ W_\mu^\dagger \bar{L} \gamma^\mu (1 - \gamma_5) L + \dots \} \quad (1.14)$$

the charged weak interactions preserve the vector minus axial (V-A) structure, and this implies that the  $W^\pm$  bosons interact only with the left doublets.

Furthermore, the weak interactions mix the down-type quark flavors so the terms that involve quarks have to include a unitary matrix, the CKM matrix. (see sec.??)

$$\mathcal{L}_{\text{wcc}}^q = -\frac{g}{\sqrt{2}} (\bar{u}_L, \bar{c}_L, \bar{t}_L) \gamma^\mu W_\mu^+ V_{\text{CKM}} \begin{pmatrix} d_L \\ s_L \\ b_L \end{pmatrix} + \dots \quad (1.15)$$

This also does not allow the presence of flavour-changing neutral currents (FCNC).

- $\mathcal{L}_{\text{wnc}}$  is the interaction term of the weak neutral currents,

$$\mathcal{L}_{\text{wnc}} = \frac{e}{2 \sin\theta_W \cos\theta_W} Z_\mu \sum_f \bar{f} \gamma^\mu (v_f - I_f^3 \gamma_5) f \quad (1.16)$$

where  $I_f^3$  is the weak isospin of the fermion and  $v_f = I_f^3 (1 - 4|Q_f| \sin^2\theta_W)$ . Unlike the W boson, the Z boson interacts both with the left and the right part of the spinor with different couplings depending on the isospin and the charge of the fermion

- $\mathcal{L}_{\text{gauge}}$  is the kinematic part of the Lagrangian related to the Gauge bosons

### 1.1.4 The spontaneous symmetry breaking

The EW Lagrangian in Eq. (??) does not include both the fermions and the boson masses. In the SM, the W/Z masses are generated by the Higgs-Brout-Englert mechanism [? ? ]. Starting from the  $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$  Lagrangian,

$$\mathcal{L} = \mathcal{L}_{\text{kin}} + \mathcal{L}_{\text{QCD}} + \mathcal{L}_{\text{EWK}} \quad (1.17)$$

we add another term introducing a scalar complex doublet field

$$\mathcal{L}_H = (D^\mu \phi)^\dagger (D_\mu \phi) - \frac{1}{2} \mu^2 (\phi^\dagger \phi) - \frac{1}{4} \lambda (\phi^\dagger \phi)^2 \quad (1.18)$$

where  $D_\mu$  is the electroweak covariant derivative defined in Eq. (??) and  $\mu^2 < 0$ . The last 2 terms form the Higgs potential  $V(\phi)$ , represented in Fig. ??.

The minimum of  $V(\phi)$  is not unique, and the choice of a specific minimum for  $\phi$  leads to the spontaneous symmetry breaking (SSB) of the  $SU(3) \otimes SU(2) \otimes U(1)$  group into the  $SU(3) \otimes U(1)$  group.

A convenient choice is to pick the ground state of  $\phi$  as

$$\langle \phi \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ \sqrt{\frac{-\mu^2}{\lambda}} \end{pmatrix} \quad (1.19)$$

Developing the Lagrangian in Eq. (??) around this minimum, we can redefine the scalar doublet as

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} \quad (1.20)$$

and the potential  $V(\phi)$  expanded in powers of the new real Higgs field  $H(x)$  becomes

$$V(H) = -\frac{1}{2} \mu^2 H^2 - \frac{\mu^2}{2\nu} H^3 - \frac{\mu^2}{8\nu^2} H^4 \quad (1.21)$$

Replacing Eq. (??) into the Eq. (??) the W/Z bosons acquire a mass term with

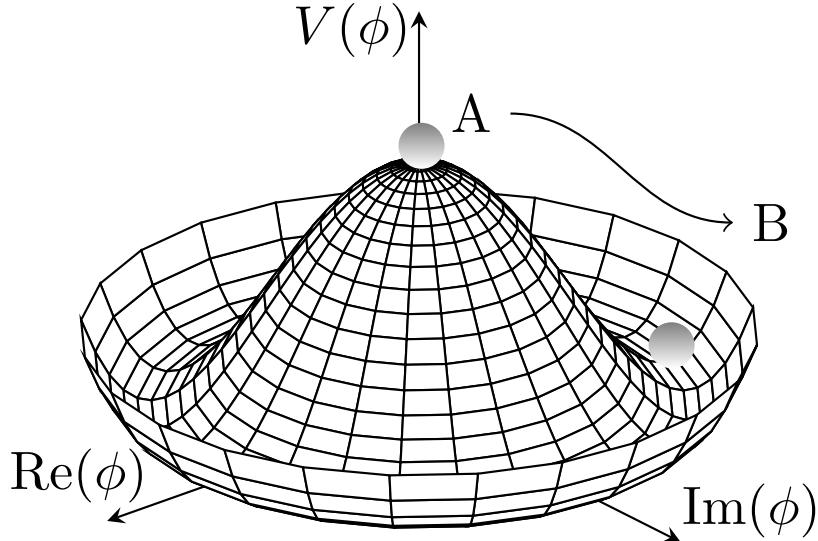
$$M_{W^\pm} = \frac{1}{2} g v \quad (1.22)$$

$$M_Z = \frac{M_{W^\pm}}{\cos(\theta_W)} \quad (1.23)$$

This also generates the Higgs boson mass  $m_H = \sqrt{2\lambda}\mu$ , the interactions between the Higgs boson and the vector bosons, and the Higgs self-interactions.

**Fermion masses** The Lagrangian  $\mathcal{L} = \mathcal{L}_{\text{kin}} + \mathcal{L}_{\text{QCD}} + \mathcal{L}_{\text{EWK}} + \mathcal{L}_H$  does not contain the fermion masses yet.

The mass terms of the Dirac Lagrangian (Eq. (??)) can't be added directly because the term  $-m\bar{\psi}\psi$  is not invariant under  $SU(2) \otimes U(1)$ . This problem can be addressed by



**Figure 1.2:** Illustration of the Higgs potential  $V(\phi)$  [? ]

introducing the Yukawa interaction between fermions and the Higgs field

$$\mathcal{L}_{\text{Yukawa}} = - (g_{ij}^d \bar{Q}_{L,i} \phi d_{R,j} + g_{ij}^u \bar{Q}_{L,i} (i\sigma^2 \phi^\dagger) u_{R,j} + g_{ij}^\ell \bar{\ell}_{L,i} \phi \ell_{R,j}) + h.c. \quad (1.24)$$

After the SSB, the masses of the fermions arise from the Yukawa interaction.

$$\mathcal{L}_{\text{masses}} = -M_{ij}^\ell \bar{\ell}_{L,i} \ell_{R,j} - M_{ij}^d \bar{d}_{L,i} d_{R,j} - M_{ij}^u \bar{u}_{L,i} u_{R,j} \quad (1.25)$$

$$M^{\ell,d,u} = \frac{v}{\sqrt{2}} g^{\ell,d,u} \quad (1.26)$$

The non-diagonal quark mass matrix  $M$  can be diagonalized through a unitary transformation, which is equivalent to a basis change.

$$M_{\text{diag}}^{u/d} = (V_L^{u/d})^\dagger M^{u/d} V_R^{u/d} \quad (1.27)$$

This transformation is the one that induces the mixing of mass eigenstates between the left-handed fermion and the W boson field. Indeed, the flavor mixing represented by the CKM matrix can be expressed in terms of these unitary transformations.

$$V_{CKM} = (V_L^u)^\dagger V_L^d \quad (1.28)$$

At this point, the SM is complete. A summary of all the particles of the SM and their quantum numbers is shown in Tab. ??

**Table 1.1:** Summary of the representation and quantum numbers SM fields. Bold numbers indicate the dimension of the representation under the respective gauge group.

Field name	Symbol	Representations		Quantum numbers		
		SU(3) <sub>C</sub>	SU(2) <sub>L</sub>	<b>T</b> <sub>3</sub>	<b>Y</b> /2	<b>Q</b> = <b>T</b> <sub>3</sub> + <b>Y</b> /2
Quark doublet	$Q_L = \begin{pmatrix} u_L \\ d_L \end{pmatrix}$	<b>3</b>	<b>2</b>	$\begin{pmatrix} +\frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}$	$+\frac{1}{6}$	$\begin{pmatrix} +\frac{2}{3} \\ -\frac{1}{3} \end{pmatrix}$
Up-quark singlet	$u_R$	<b>3</b>	<b>1</b>	$+\frac{2}{3}$	$+\frac{2}{3}$	$+\frac{2}{3}$
Down-quark singlet	$d_R$	<b>3</b>	<b>1</b>	$-\frac{1}{3}$	$-\frac{1}{3}$	$-\frac{2}{3}$
Lepton doublet	$L_L = \begin{pmatrix} \nu_L \\ e_L \end{pmatrix}$	<b>1</b>	<b>2</b>	$\begin{pmatrix} +\frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}$	$-\frac{1}{2}$	$\begin{pmatrix} 0 \\ -1 \end{pmatrix}$
Lepton singlet	$e_R$	<b>1</b>	<b>1</b>	0	-1	-1
Gluon field	$G_\mu^a$	<b>8</b>	<b>1</b>	0	0	0
Weak gauge field	$W_\mu^i = \begin{pmatrix} W_\mu^+ \\ W_\mu^- \\ W_\mu^3 \end{pmatrix}$	<b>1</b>	<b>3</b>	$\begin{pmatrix} +1 \\ -1 \\ 0 \end{pmatrix}$	0	$\begin{pmatrix} +1 \\ -1 \\ 0 \end{pmatrix}$
Hypercharge field	$B_\mu$	<b>1</b>	<b>1</b>	0	0	0
Higgs doublet	$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}$	<b>1</b>	<b>2</b>	$\begin{pmatrix} +\frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}$	$+\frac{1}{2}$	$\begin{pmatrix} +1 \\ 0 \end{pmatrix}$
Conjugate Higgs doublet	$\Phi^c = \begin{pmatrix} \phi^{0*} \\ \phi^- \end{pmatrix}$	<b>1</b>	<b>2</b>	$\begin{pmatrix} +\frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}$	$-\frac{1}{2}$	$\begin{pmatrix} 0 \\ -1 \end{pmatrix}$

## 1.2 The CKM matrix

As already discussed, the weak interactions mix the down-type quarks with the Cabibbo-Kobayashi-Maskawa (CKM) matrix [? ? ].

$$\mathcal{L}_{\text{wcc}}^q = -\frac{g}{\sqrt{2}}(\bar{u}_L, \bar{c}_L, \bar{t}_L)\gamma^\mu W_\mu^+ V_{\text{CKM}} \begin{pmatrix} d_L \\ s_L \\ b_L \end{pmatrix} + \dots \quad (1.29)$$

$$V_{\text{CKM}} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \quad (1.30)$$

The CKM matrix is a  $3 \times 3$  complex matrix that can be written in terms of 18 real parameters. The unitarity condition  $V^\dagger V = \mathbb{1}$  adds 9 constraints, and the 6 quark fields can absorb the phases defining a new common phase, so the number of parameters of the CKM matrix is four.

These parameters can be determined through global fits that combine experimental data from different experiments, imposing the constraints of the standard model (e.g. the unitarity), or can be measured independently [? ].

Measure all the parameters independently allows us to verify the unitarity of the CKM matrix that, if violated, can be a sign of new physics.

The  $|V_{ud}|$  is the element measured with the highest precision through the study of

$\beta$  decays. The other  $|V_{uq}|$  and  $|V_{cq}|$  elements are measured through leptonic and semileptonic meson decays, but these kinds of measures are highly affected by theoretical uncertainties due to the non-perturbative nature of the strong interaction at low energy that bonds the quarks in the meson. The  $|V_{tq}|$  elements are not affected by non-perturbative QCD uncertainties, but the QCD background of a hadron collider makes their measure challenging and, at this time, an electron-positron collider with sufficient energy in the center of mass to produce a top quark does not exist yet.

There are two common parametrizations of the CKM matrix: the standard parametrization and the Wolfenstein parametrization [? ]. The values of the parameters listed below are the ones obtained by the CKMfitter group through a global fit [? ? ? ].

**Standard parametrization** In the standard parametrization, the CKM matrix is expressed as the product of 3 rotation matrices with the addition of a phase.

$$V_{CKM} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & 0 & s_{13}e^{-i\delta} \\ 0 & 1 & 0 \\ -s_{13}e^{i\delta} & 0 & c_{13} \end{pmatrix} \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (1.31)$$

where  $c_{ij} = \cos \theta_{ij}$ ,  $s_{ij} = \sin(\theta_{ij})$  and  $\delta$  is the phase responsible for the CP violation in weak interactions.

The values obtained by the CKMfitter group are the following

$$\begin{aligned} \sin \theta_{12} &= 0.22500 \pm 0.00085, & \sin \theta_{13} &= 0.00369 \pm 0.00011, \\ \sin \theta_{23} &= 0.04182^{+0.00067}_{-0.00074}, & \delta &= 1.144 \pm 0.027. \end{aligned} \quad (1.32)$$

**Wolfenstein's parametrization** Defining new four real parameters,

$$s_{12} = \lambda = \frac{|V_{us}|}{\sqrt{|V_{ud}|^2 + |V_{us}|^2}}, \quad (1.33)$$

$$s_{23} = A\lambda^2 = \lambda \left| \frac{V_{cb}}{V_{us}} \right|, \quad (1.34)$$

$$s_{13}e^{i\delta} = A\lambda^3(\rho + i\eta) = \frac{A\lambda^3(\bar{\rho} + i\bar{\eta})\sqrt{1 - A^2\lambda^4}}{\sqrt{1 - \lambda^2}[1 - A^2\lambda^4(\bar{\rho} + i\bar{\eta})]}. \quad (1.35)$$

the Wolfenstein parameterization shows clearly the order of magnitude of the different elements in powers of  $\lambda$

$$V_{CKM} \simeq \begin{pmatrix} 1 - \frac{\lambda^2}{2} & \lambda & \lambda^3(\rho - i\eta) \\ -\lambda & 1 - \frac{\lambda^2}{2} & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4). \quad (1.36)$$

The CKMfitter group performed a global fit for this parametrization as well.

$$\begin{aligned} \lambda &= 0.22500 \pm 0.00067, & A &= 0.826^{+0.018}_{-0.015}, \\ \bar{\rho} &= 0.159 \pm 0.010, & \bar{\eta} &= 0.348 \pm 0.010. \end{aligned} \quad (1.37)$$

and the fit result of the magnitudes of the CKM elements is:

$$|V_{\text{CKM}}| = \begin{pmatrix} 0.97435 \pm 0.00016 & 0.22500 \pm 0.00067 & 0.00396 \pm 0.00011 \\ 0.22486 \pm 0.00067 & 0.97349 \pm 0.00016 & 0.04182^{+0.00085}_{-0.00074} \\ 0.00857^{+0.00020}_{-0.00018} & 0.04110^{+0.00083}_{-0.00072} & 0.999118^{+0.000031}_{-0.000036} \end{pmatrix}. \quad (1.38)$$

**Unitarity triangle** The unitarity constraints of the CKM matrix are:

$$\sum_i V_{ij} V_{ik}^* = \delta_{jk} \quad ; \quad \sum_j V_{ij} V_{kj}^* = \delta_{ik} \quad (1.39)$$

These constraints can be represented in the complex  $(\bar{\rho}, \bar{\eta})$  plane in a triangle with vertexes  $(0, 0), (1, 0), (\bar{\rho}, \bar{\eta})$  (Fig. ??). The area of the unitarity triangle is also a measure of the CP violation of the weak interactions.

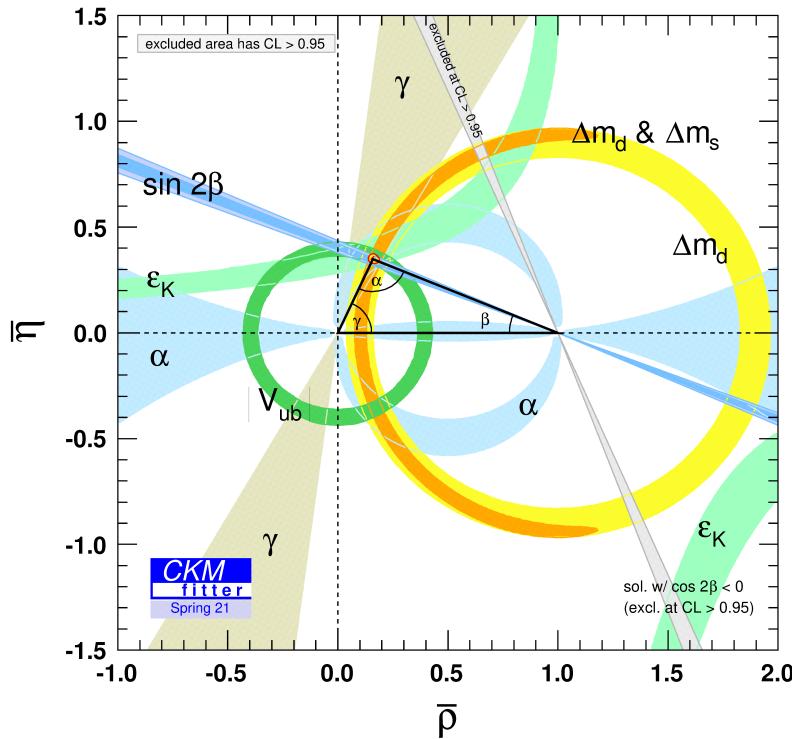


Figure 1.3: CKMfitter global constraints on the unitarity triangle [? ].

## 1.3 The Top quark

The top quark is the heaviest particle of the standard model, discovered at Fermilab in 1995 by the CDF and the D0 collaborations [? ]. Its lifetime,  $\sim 10^{-25} s$ , is less than the hadronization timescale, making it the only quark that decays freely [? ].

The top quark decays through weak interactions through the channel  $t \rightarrow Wq$  and, due to its mass of  $m_t \simeq 173 \text{ GeV}$ , the top quark is also the only fermion that can

decay into an on-shell W boson.

The final states that can be observed are:

- $t \rightarrow W^+ q_i^d \rightarrow q_1^d q_2 \bar{q}_3$  with a branching ration of  $\simeq 67.5\%$
- $t \rightarrow W^+ q_i^d \rightarrow q_1^d \ell^- \nu_\ell$  with a branching ratio of  $\simeq 32.5\%$

Where  $q_i^d$  is a down-type quark of flavor  $i = d, s, b$ . The branching fractions of the decay to the three different flavors is

$$\frac{\Gamma(t \rightarrow W q_i^d)}{\Gamma(t \rightarrow W q^d)} = |V_{ti}|^2 \quad (1.40)$$

Indeed, the top quark decays to a W boson and a b quark with a branching fraction  $> 99.8\%$ .

### 1.3.1 Top pairs production at LHC

In a hadron collider, top quarks are mainly produced in  $t\bar{t}$  pairs via gluon-gluon fusion ( $ggF$ )  $gg \rightarrow t\bar{t}$  or quarks annihilation ( $q\bar{q} \rightarrow t\bar{t}$ ).



**Figure 1.4:** Tree level Feynman diagrams of  $t\bar{t}$  production via gluon-gluon fusion (middle, right) and quark-antiquark annihilation (left)

The protons that collide in the LHC can be described by the parton model [? ].

Each proton is composed of three valence quarks (uud) and a sea of quarks, anti-quarks, and gluons generated by the QCD activity in the hadron.

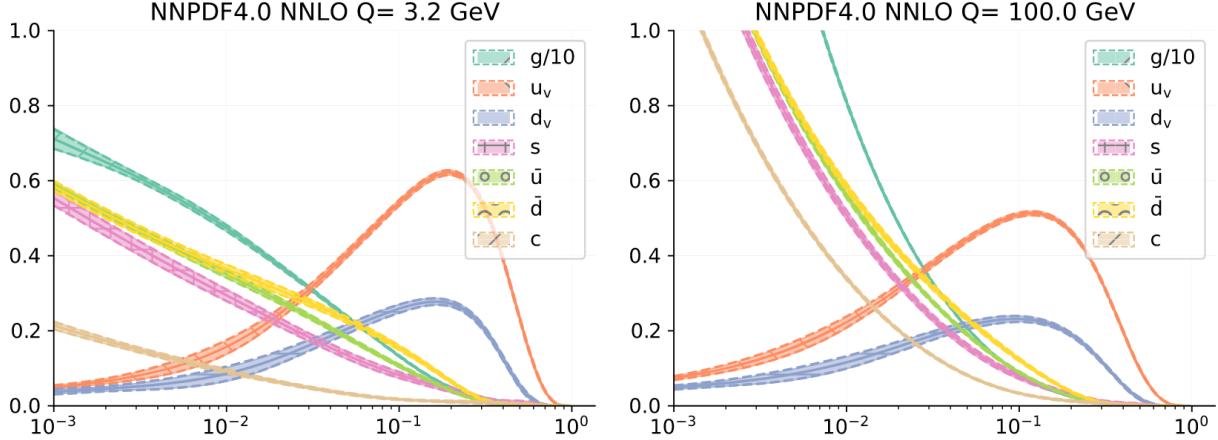
The elements that compose the proton are called “partons” and each parton carries a fraction  $x_i$  of the proton momentum.

The so-called Bjorken variable  $x_i$  follows a statistical distribution, the parton distribution functions (PDFs)  $f(x, Q^2)$ , that depends on the type of parton, and on  $Q^2$ , the energy in the center of mass of the partons involved in the hard scattering.

The equations that describe the evolution of the PDFs at the different energy scales are the Dokshitzer–Gribov–Lipatov–Altarelli–Parisi (DGLAP) equations [? ? ? ]. Still, the PDFs can’t be computed a priori, they have to be extracted from experimental data. At low energies, the PDFs are dominated by the presence of valence quarks, but the contribution of gluons and non-valence quarks increases with the proton energy as shown in Fig. ??.

To compute the inclusive  $t\bar{t}$  production cross-section at a hadron collider, the parton-level cross-section has to be convolved with the PDFs

$$\sigma_{t\bar{t}} = \sum_{i,j=q,\bar{q},g} \int dx_i dx_j f_i(x_i, Q^2) f_j(x_j, Q^2) \cdot \hat{\sigma}_{ij \rightarrow t\bar{t}}(x_i, x_j) \quad (1.41)$$



**Figure 1.5:**  $xf(x, Q^2)$  values for the NNPDF 4.0 NNLO PDFs set at  $Q = 3.2 \text{ GeV}$  (left) and  $Q = 100 \text{ GeV}$  (right) presented by the NNPDF collaboration [? ].

where the sums run over the type of partons (quarks, antiquarks, and gluons). The energy in the center of mass at LHC is  $\sqrt{s} = 13 \text{ TeV}$  and the predominant  $t\bar{t}$  production channel is the ggF with a branching ratio of  $\simeq 90\%$ .

The total  $t\bar{t}$  production cross-section computed at the  $\mathcal{O}(\alpha_s^4)$  order of accuracy using the program TOP++ [? ] and assuming  $m_t = 172.5 \pm 1.0 \text{ GeV}$  is [? ]

$$\sigma_{t\bar{t}} = 833.9^{+20.5}_{-30.0} {}^{+23.2}_{-22.5} \pm 21.0 \text{ pb} \quad (\pm \text{Scale} \pm m_t \pm (\text{PDF} + \alpha_s)) \quad (1.42)$$

**Additional jets** The emission of radiative gluons or additional quarks originating from gluon splitting can cause the formation of additional jets (see sec. ??).

These events are suppressed since there is an  $\alpha_s$  factor for each new additional vertex in the Feynman diagram.

## Chapter 2

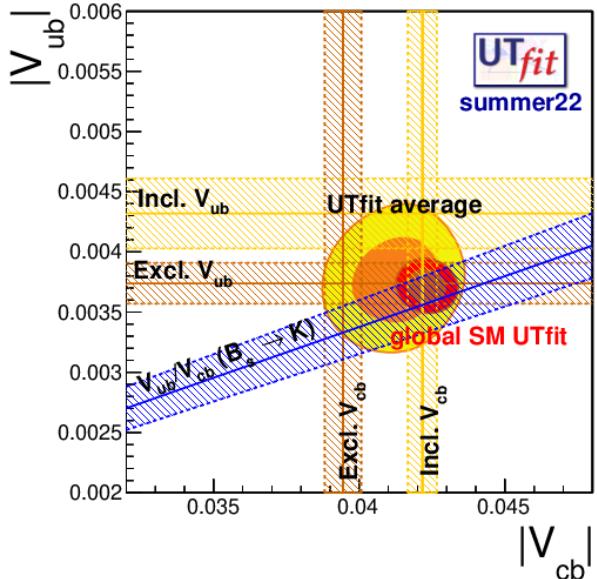
# The $|V_{cb}|$ CKM element

This work aims to access the  $|V_{cb}|$  element through the direct decay of a W boson at  $\sqrt{s} = 13\text{TeV}$  with the recorded luminosity of the RunII.

## 2.1 The $V_{cb}$ element

The  $V_{cb}$  element plays an important role in the unitarity of the CKM matrix. The radius of the unitarity-clock, the green circle centered in the origin in Fig. ??, is proportional to the ratio  $|V_{ub}/V_{cb}|$ , so  $V_{cb}$  normalize the whole triangle [? ].

$|V_{cb}|$  is measured from the semileptonic inclusive and exclusive  $b \rightarrow c\ell\nu$  decays but the parton level amplitude, which is proportional to  $|V_{cb}|^2$ , has to be corrected by introducing form factors to consider the contributions by strong interactions in the meson.



**Figure 2.1:**  $|V_{cb}|$  and  $|V_{ub}|$  global fit and averages of inclusive and exclusive measurements performed by the UTfit collaboration [? ]

These corrections inevitably lead to theoretical uncertainties that are of the same order as the experimental systematic uncertainties.

These issues also concern the  $|V_{ub}|$  measurements.

A summary of the single measurements of the  $|V_{cb}|$  element can be found in [? ].

### 2.1.1 $|V_{cb}|$ from semileptonic B meson decays

**Inclusive measurement** The inclusive channel  $B \rightarrow X_c \ell \nu$  is less model dependent than the exclusive one [? ]. The QCD contributions can be included by adding two terms in the amplitude: a perturbative term and a non-perturbative term.

$$\Gamma(B \rightarrow X_c \ell \nu) = \frac{G_F^2 m_b^5 |V_{cb}|^2}{192\pi^3} [1 + P(\alpha_s, \mu, m_b/m_c) + N(m_b, m_c, \mu_\pi^2, \mu_G^2, \rho_D^2, \rho_{LS}^3)] \quad (2.1)$$

The QCD corrections are computed using the so-called Operator Product Expansion (OPE). OPE allows us to express the widths and moments of the kinematic distributions as a double expansion in  $\alpha_S$  and  $\Lambda_{QCD}/m_b$ . These corrections are actually known up to the third order in  $\Lambda_{QCD}/m_b$  [? ].

This expansion depends on other parameters defined in the framework of the heavy quark effective field theory (HQET) that can be constrained through the study of other observables like the lepton energy, the transferred momentum or the hadronic mass [? ] so the final fit has to be performed on an order of  $\mathcal{O}(100)$  of observables.

The uncertainties of the inclusive measurements of  $|V_{cb}|$  are  $\simeq 2\%$ , mainly driven by theoretical uncertainties related to the determination of the HQET parameters [? ].

The average of the inclusive measurements is [? ]

$$|V_{cb}| \times 10^3 \text{ (incl)} = 42.16 \pm 0.50 \quad (2.2)$$

**Exclusive measurement** The two main exclusive channels studied for  $|V_{cb}|$  determination are the  $\bar{B} \rightarrow D^* \ell \bar{\nu}_\ell$  and the  $\bar{B} \rightarrow D \ell \bar{\nu}_\ell$ . The first channel was measured with an uncertainty of  $\simeq 2\%$  with comparable contributions from theory and experiment, while the latter was measured with an uncertainty of the  $\simeq 3\%$  but recently some questions about the form factor parametrization were raised.

The exclusive measurement requires knowing six form factors that can be computed using lattice QCD (LQCD) techniques. In the limit of heavy quarks, the six form factors collapse into a single one, and corrections to the  $m_b \rightarrow \infty$  limit can be obtained through the HQET framework [? ].

The average of the exclusive measurements is [? ]

$$|V_{cb}| \times 10^3 \text{ (excl)} = 39.44 \pm 0.63 \quad (2.3)$$

**The  $|V_{cb}|$  puzzle** The inclusive and exclusive measurements of the  $V_{cb}$  element have always been incompatible as shown in Fig. ??.

Recently, a revision of the techniques involved in the calculations of the QCD contributions has narrowed the gap, but the  $V_{cb}$  puzzle is still unresolved [? ]: indeed the current determinations show a 7% difference from each other.

Combining the measurements with a  $\chi^2$  fit, the best-fit value is  $|V_{cb}|_{\text{best}} = 41.10$  that leads to a chi-square of  $\chi^2 = 11.35$ , corresponding to a p-value  $p = 7.2 \cdot 10^{-4}$ , i.e. to a tension of  $\sim 3.3\sigma$ . The determination of  $|V_{cb}|$  through direct W decays would allow us to provide a new determination of  $|V_{cb}|$  at a new energy scale ( $\sqrt{s} = m_W$ ), avoiding complications and theoretical uncertainties due to non-perturbative QCD, tackling the  $|V_{cb}|$  puzzle with a novel strategy.

## 2.2 CKM measurements from hadronic W decays

### 2.2.1 LEP2 measurements

The Large Electron-positron Collider (LEP) was an electron-positron collider that reached a peak energy in the center of mass of 209 GeV. It was built at CERN and is located in the

tunnel that now hosts the LHC accelerator facility and started its physics program in 1989.

The LEP2 physics program started in 1996, running at a center of mass energy that was increased from 161 GeV to 209 GeV at the end of 1999, above the production threshold of W pairs. In this period, each of the four LEP experiments, ALEPH, DELPHI, L3, and OPAL have collected  $\sim 700\text{pb}^{-1}$  of data above the WW pair production threshold, corresponding to over 10000 WW events per experiment [? ].

In this period,  $\sim 10$   $W \rightarrow cb$  events per experiment were produced, making it impossible to provide a direct determination of  $|V_{cb}|$  from inclusive W decays.

However, with more than 13000 hadronic W decays per experiment, the LEP experiments were able to access other elements in the charm sector of the CKM matrix.

**Rate of charm production** ALEPH and OPAL performed a measure of the rate or  $W \rightarrow cX$  decays on both semileptonic and fully hadronic W pairs.

$$R_c^W = \frac{\Gamma(W \rightarrow cX)}{\Gamma(W \rightarrow q\bar{q})} = \frac{|V_{cd}|^2 + |V_{cs}|^2 + |V_{cb}|^2}{|V_{cd}|^2 + |V_{cs}|^2 + |V_{cb}|^2 + |V_{ud}|^2 + |V_{us}|^2 + |V_{ub}|^2} \quad (2.4)$$

This is a unitarity test of the second row of the CKM matrix, indeed, if the unitarity hypothesis holds, it has to be  $R_c^W = 0.5$ .

ALEPH performed the measurement extracting  $R_c^W$  with a binned maximum-likelihood fit on the shape of the most c-tagged jet[? ].

The charm jet tagger was a feed-forward neural network with 12 variables as input that uses mainly information from charm lifetime, jet-shape properties, reconstruction of D mesons, and lepton identification.

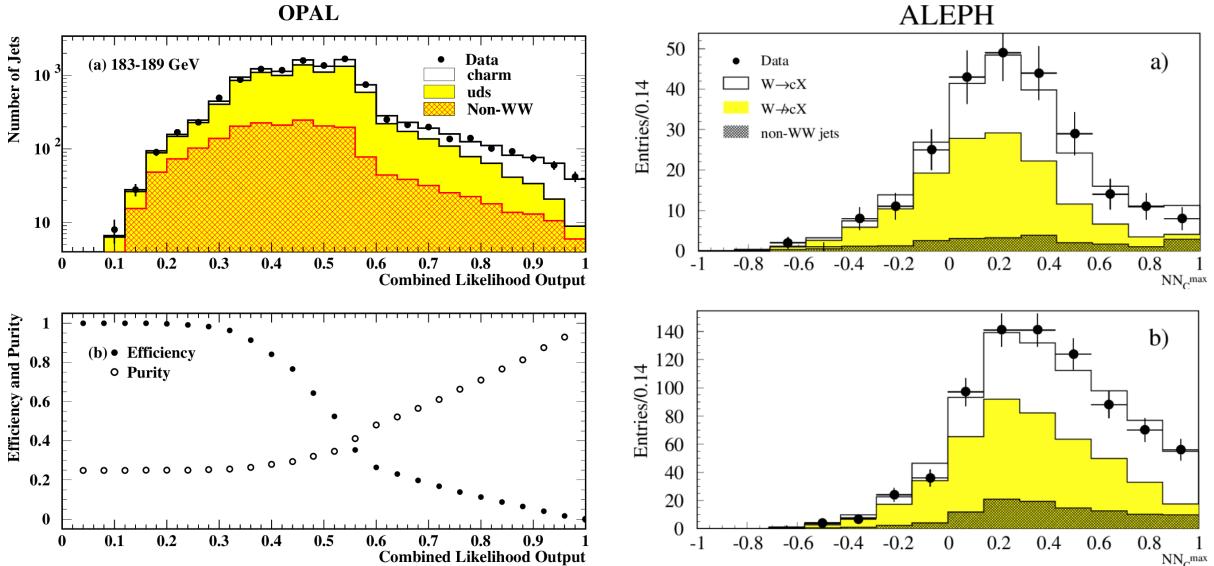
On the other hand, OPAL used a relative likelihood discriminant calculated for each jet to separate jets originating from charm quarks and jets from light quarks. This discriminant relies on jet and event shape properties, lifetime information, and lepton identification [? ].

The results of both experiments are consistent with the unitarity hypothesis with an uncertainty of  $\sim 10\%$

$$\begin{aligned} R_c^W &= 0.51 \pm 0.05 \text{ (stat)} \pm 0.03 \text{ (syst)} & (\text{ALEPH } [? ]) \\ R_c^W &= 0.481 \pm 0.042 \text{ (stat)} \pm 0.032 \text{ (syst)} & (\text{OPAL } [? ]) \end{aligned} \quad (2.5)$$

$R_c^W$  allow also an indirect measure of  $|V_{cs}|$ , subtracting the other  $|V_{cx}|$  CKM elements in quadrature from  $R_c^W$  obtained from other determinations.

$$\begin{aligned} |V_{cs}| &= 1.00 \pm 0.11 \text{ (stat)} \pm 0.07 \text{ (syst)} & (\text{ALEPH } [? ]) \\ |V_{cs}| &= 0.969 \pm 0.058 & (\text{OPAL } [? ]) \end{aligned} \quad (2.6)$$



(a) OPAL: Output of the combined likelihood used to tag charm hadrons at 183–189 GeV [?].

(b) ALEPH: Maximum c tagging NN score for the dijet pairs. In the top panel there are the semileptonic W pairs, in the bottom panel there are the fully hadronic [? ].

**Direct measure of  $|V_{cs}|$**  The  $V_{cs}$  CKM element was measured directly from W decays by the DELPHI collaboration, using semileptonic and fully hadronic decays of W pairs measuring  $r^{(cs)}$  [? ].

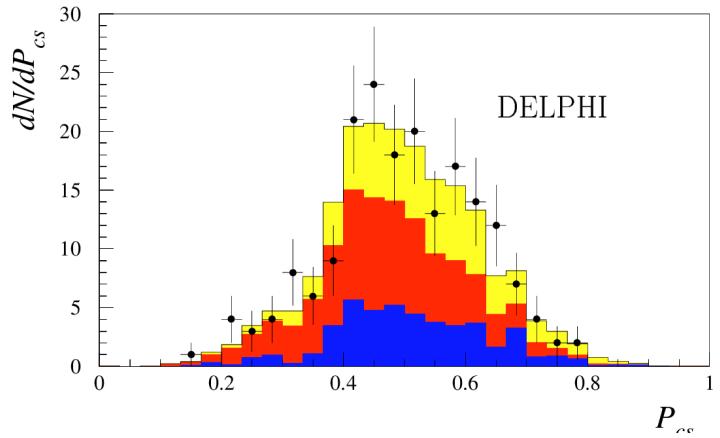
$$r^{(cs)} = \frac{\Gamma(W \rightarrow cs)}{\Gamma(W \rightarrow q\bar{q})} = \frac{|V_{cs}|^2}{2} \quad (2.7)$$

where, in the last step, the unitarity of the CKM matrix is assumed.

To discriminate the signal, DELPHI built a  $P_{cs}$  tagger that relies on information from the vertex and tracking detectors and on particle identification through RICH detectors.

The value of  $V_{cs}$  was extracted fitting the  $P_{cs}$  distribution of selected di-jet combinations by using the maximum likelihood method on a multinomial likelihood.

The direct determination of  $|V_{cs}|$  was performed on just 100 W pairs collected in 1996, leading to a large statistical uncertainty of  $\sim 30\%$



**Figure 2.3:**  $P_{cs}$  distribution. The yellow is  $W \rightarrow cX$  process, the red  $W \rightarrow uX$ , and the blue is the background [? ].

However, for the first time, DELPHI was able to provide a direct determination of  $|V_{cs}|$  dominated by the statistical uncertainties and not dependent on theoretical systematics.

$$|V_{cs}| = 0.94^{+0.32} \text{ (stat)} \pm 0.13 \text{ (syst)} \quad (\text{DELPHI [? ]}) \quad (2.8)$$

## 2.2.2 $|V_{cb}|$ with LHC data

At LHC, in all the RunII, more than 26 billion single W bosons were produced: of these, more than 17 billion W bosons decay hadronically, which means that more than 14 million cb quark pairs can be obtained from the inclusive decay of W bosons.

Despite that, contrary to any electron-positron collider, an inclusive measurement of  $W \rightarrow cb$  at a hadron collider would be an impossible task due to the large QCD background that would not even allow to trigger and record candidate events.

On the other hand, it is possible to exploit semileptonic decays of W pairs and use the lepton to trigger the event and the hadronic W decay to measure the  $W \rightarrow cb$  branching fraction.

As lepton  $\ell$ , we will only consider muons and electrons, due to the lower reconstruction efficiency of  $\tau$ s, that decay hadronically in the 64.8% of cases.

The value of  $|V_{cb}|$  can be extracted from the ratio of branching fractions in which the hadronic W decays into c and b quarks over the W hadronic branching fraction

$$\begin{aligned} R_{cb} &= \frac{\Gamma(WW \rightarrow \ell\nu cb)}{\Gamma(WW \rightarrow \ell\nu q\bar{q})} = \\ &= \frac{|V_{cb}|^2}{|V_{ud}|^2 + |V_{us}|^2 + |V_{ub}|^2 + |V_{cd}|^2 + |V_{cs}|^2 + |V_{cb}|^2} \end{aligned} \quad (2.9)$$

and if we assume the unitarity of the CKM matrix

$$R_{cb} = \frac{|V_{cb}|^2}{2} \quad (2.10)$$

This kind of measure will be limited mainly by the b and c tagging systematic uncertainties, but the ratio of branching fractions should allow some cancellations.

**WW production** At an energy of the center of mass of  $\sqrt{s} = 13$  TeV, the direct WW production cross section is  $\sigma_{pp \rightarrow WW} \sim 118$  pb, which means that in all the RunII, more than 16 million of W pairs were produced directly.

However, at LHC, the production cross-section of  $t\bar{t}$  pairs is  $\sim 7$  times greater than the direct WW production cross-section with  $\sigma_{pp \rightarrow t\bar{t}} \sim 832$  pb and, since the top quark decays in one b quark plus a W boson, we can exploit di-top decays as a WW factory.

### Decay of W pairs

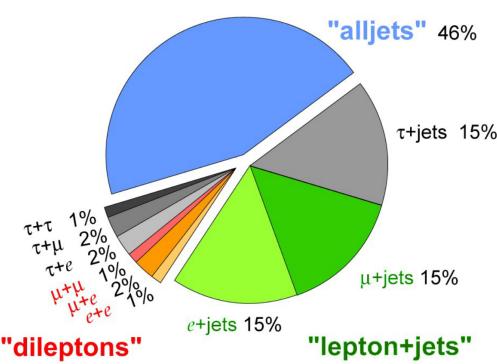


Figure 2.4: W pair decay branching fractions

Since the branching fraction of the W decay are:

- 32.6% Hadronic  $W \rightarrow q\bar{q}$
- 67.4% Leptonic  $W \rightarrow \ell\nu$

the branching fractions of the  $W^+W^-$  decay are:

- 45.5% Fully hadronic  $WW \rightarrow q\bar{q}q\bar{q}$ .
- 43.9% Semi leptonic  $WW \rightarrow q\bar{q}\ell\nu_\ell$
- 10.6% Di-leptonic  $WW \rightarrow \ell\bar{\nu}_\ell\ell\nu_\ell$

The branching ratio of the flavors of the  $q_i\bar{q}_j$  pairs generated by W decays is given by the respective CKM elements  $|V_{ij}|^2/2$ .

Since more than 110 million W pairs were produced in the RunII through di-top decays, around 40000  $W \rightarrow cb$  events were produced in the  $t\bar{t}$  semileptonic channel, while the amount of cb events produced by the semileptonic decay of directly produced W pairs is around 6000.

The amount of cb events provided by  $t\bar{t}$  semileptonic decays in the RunII allows us to provide a new determination of  $|V_{cb}|$  that, in case of a perfect selection, would reach a 0.5% statistical precision, that would be a huge improvement with respect the current 7% difference between the current inclusive and exclusive determination obtained through the semileptonic decay of B mesons.

However, the separation between signal and background, which relies mainly on the b/c-tagging capabilities of the experiment, must be evaluated, along with the b/c-tagging systematic uncertainties.

W → cb events				
W	WW		$t\bar{t}$	
Had	diHad	semiLept	diHad	semiLept
$1.4 \cdot 10^7$	$1.8 \cdot 10^4$		$1.2 \cdot 10^5$	
$1.4 \cdot 10^7$	$1.2 \cdot 10^4$	$5.8 \cdot 10^3$	$8.0 \cdot 10^4$	$4.0 \cdot 10^4$

Table 2.1: Amount of  $W \rightarrow cb$  events produced per experiment in all the RunII at LHC for each process.

## *Chapter 3*

---

# The CMS experiment at LHC

---

CERN is the largest international particle physics research center in the world. It is located across the Franco-Swiss border in Meyrin, near Geneva, and was founded in 1954, after the end of World War II, by twelve European countries.

Today, more than 100 countries, 500 institutes, and 13000 users collaborate to the CERN activities.

### 3.1 The Large Hadron Collider

The CERN's major facility is the Large Hadron Collider (LHC), a proton-proton (pp) and ions collider that, with its lengths of 26.7 km and its energy in the center of mass of  $\sqrt{s} = 13.6 \text{ TeV}$ , is the largest and the most powerful particle accelerator ever built in the world.

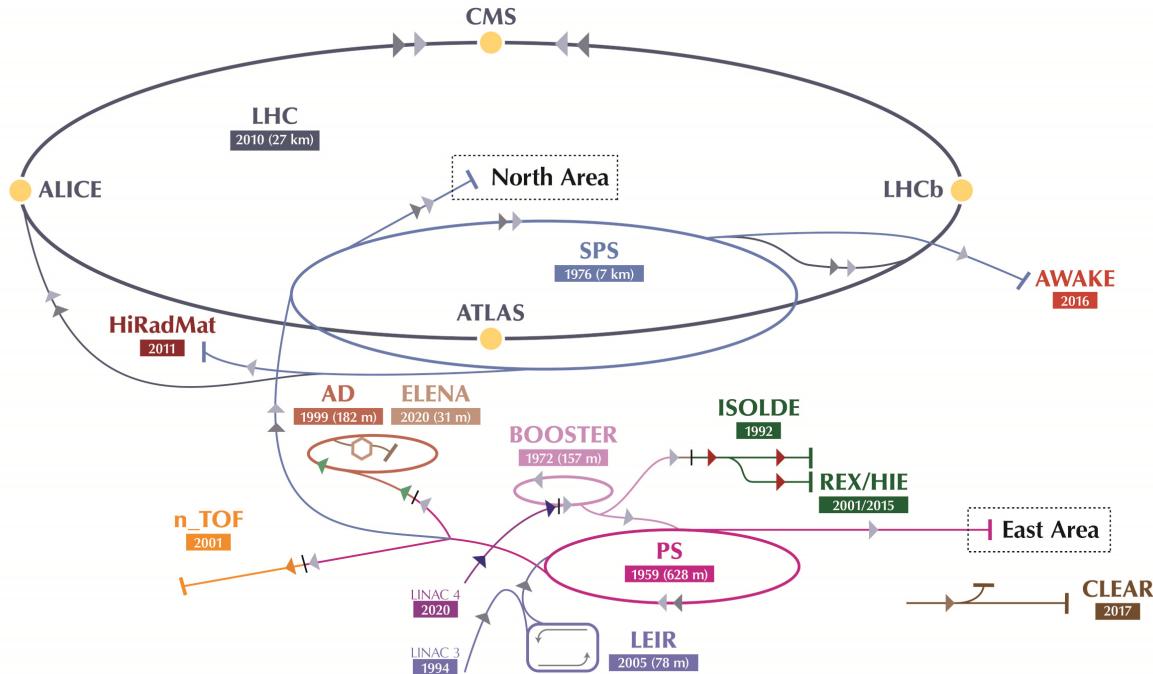
As shown in Fig. ??, the LHC is just the last stage of a chain of accelerators. The protons obtained from ionizing hydrogen are grouped in bunches through a quadrupole magnet and then are sent to a chain of accelerators that increase progressively the energy of the protons. Before entering the LHC, the proton beam is split into two beam lines that travel in opposite directions and then are accelerated by radiofrequency (RF) cavities that are tuned to oscillate at 400MHz. In addition to RF cavities, quadrupoles are used to keep the beam focused and dipoles to bend the beam [? ].

Once the beam is stabilized, the proton bunches collide in four different points where the experiments ATLAS, CMS, LHCb, and ALICE are located.

ATLAS (A Toroidal LHC ApparatuS) and CMS (Compact Muon Solenoid) are general-purpose detectors and are the ones that in the 2012 discovered the Higgs boson [? ? ], LHCb (LHC beauty) is a forward detector designed to study the physics of the B mesons and the matter-antimatter asymmetry and ALICE (A Large Ion Collider Experiment) is a detector devoted to the study of quark-gluon plasma and extreme phases of QCD matter through the analysis of lead ion collisions.

**Luminosity** The *instantaneous luminosity* is defined as the ratio between the event rate produced for a given process and its cross-section

$$\mathcal{L} = \frac{\partial N}{\partial t} \frac{1}{\sigma} \quad (3.1)$$



**Figure 3.1:** CERN’s accelerator complex. The accelerating chain is: LINAC2 (50 MeV) → Booster (1.4 GeV) → Proton Synchrotron (PS) (26 GeV) → Super Proton Synchrotron (450 GeV) → LHC (13/14 TeV). The intermediate accelerators also provide a proton beam to other smaller experiments and facilities. [? ]

and can be expressed also as a function of the beam parameters

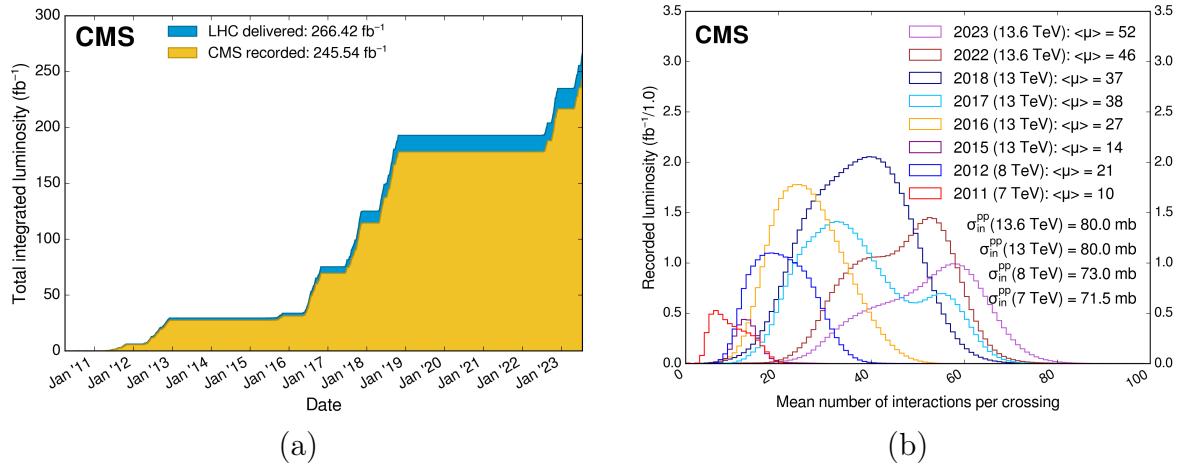
$$\mathcal{L} = \frac{N_p^2 f \gamma_r}{4\pi \epsilon_n \beta^*} F \quad (3.2)$$

where  $N_p$  is the number of protons per bunch,  $f$  the bunch frequency,  $\gamma_r$  the relativistic factor, and  $\epsilon_n, \beta^*, F$  are geometrical factors that take into account the shape of the beam and the crossing angle between the two beams in the interaction point (IP).

These parameters depend on the so-called “filling scheme”, the chosen pattern of filled and empty bunch crossings used in a single fill. A typical filling scheme is composed of long strings of consecutive bunches called a “train”, with the individual trains separated by gaps of varying lengths. Filling schemes also include some number of non-colliding bunch crossings that can be used to study effects from beam-induced background [? ].

The LHC was designed to achieve an instantaneous luminosity of  $\mathcal{L} = 10^{34} \text{ cm}^{-2} \text{s}^{-1}$  but, in 2018, the LHC was able to reach a peak luminosity of  $\mathcal{L} = 2.1 \cdot 10^{34} \text{ cm}^{-2} \text{s}^{-1}$  doubling the nominal design value. Increasing the luminosity is essential to observe rare events and to decrease the statistical uncertainties of every measurement, but it has a drawback: the number of simultaneous pp interactions, called pileup (PU), increases with the instantaneous luminosity, making the event reconstruction more difficult and generating underlying events, i.e. the hadronic activity that does not originate from the hard scattering. PU effects can be mitigated thanks to high-granularity detectors and advanced reconstruction algorithms.

The amount of recorded data is quantified by the *integrated luminosity*, the time integral of the instantaneous luminosity  $\mathcal{L}_I = \int dt \mathcal{L}(t)$ .



**Figure 3.2:** (a) cumulative integrated luminosity delivered and recorded by the CMS experiment; (b) PU distribution across different years at the CMS experiment [? ]

**Physics runs** The LHC program covers a period of  $\sim 30$  years, divided into two phases. The *Phase I* includes Run I (2011-2012), Run II (2015-2018), and Run III (2022-2025). During Phase I the center of mass energy was increased from 7 TeV to 13.6 TeV and the instantaneous luminosity reached a peak of  $\mathcal{L} = 2.1 \cdot 10^{34} \text{ cm}^{-2} \text{s}^{-1}$ .

At the end of 2025, the LHC will enter the *High Luminosity LHC* (HL-LHC) phase, a new phase of upgrades that will take the center of mass energy to 14 TeV and to an instantaneous luminosity of  $\mathcal{L} = 5 \cdot 10^{34} \text{ cm}^{-2} \text{s}^{-1}$ , providing an integrated luminosity during all the Phase II (2029-2040) of  $\mathcal{L}_I = 3000 \text{ fb}^{-1}$ , ten times more than in Phase I. During the long shutdown 3 (LS3: 2025-2029) the detectors will be upgraded, along the LHC, to be able to work in such a challenging environment that causes more radiation damage, especially in the forward subdetectors, and with a significant PU increase up to  $\langle \text{PU} \rangle \sim 200$  [? ].

## 3.2 The Compact Muon Solenoid experiment

The CMS experiment [?] is one of the two largest general-purpose detectors operating at CERN, located in Cessy(FR), 100 m underground at the LHC Point 5 (P5).

It is a cylindrical detector, 21.6 m long with a diameter of 14.6 m and a weight of 12500 tons, with coverage in pseudorapidity up to  $|\eta| < 3.0$ <sup>1</sup> extended up to  $|\eta| < 5.0$  by the high forward Cherenkov calorimeter.

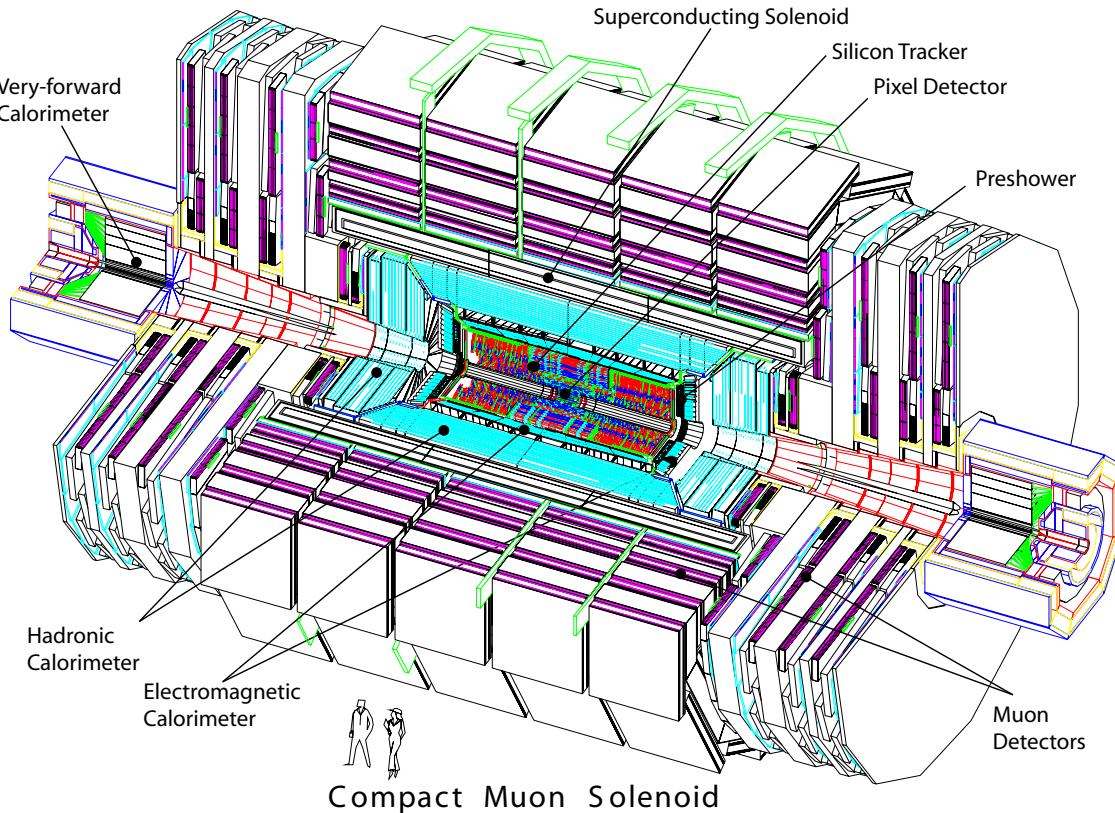
It consists of several subdetector layers, as shown in Fig. ??: from inside out, the silicon tracker, the electromagnetic calorimeter (ECAL) and the hadronic calorimeter (HCAL) contained inside the solenoid magnet and the muon system embedded in the flux-return yoke outside the magnet.

The key features of the CMS detector are the strong solenoidal magnetic field of 3.8T that is designed to optimize the  $p_T$  resolution of charged particles and a PbWO<sub>4</sub> electromagnetic calorimeter that improves the  $e/\gamma$  identification and resolution to maximize the

<sup>1</sup> $\eta = -\ln(\tan(\theta/2))$  where  $\theta$  is the azimuthal angle in spherical coordinates

sensitivity to the low mass  $\gamma\gamma$  resonances.

These choices have some drawbacks: the PbWO<sub>4</sub> crystals suffer radiation damage, becoming more and more opaque in time, and the solenoidal magnet constraint the size of the hadronic calorimeter to just 7 nuclear interaction lengths <sup>2</sup>, in contrast to the  $10\lambda$  of the ATLAS experiment, affecting the jet energy resolution [? ].



**Figure 3.3:** Exploded view of the CMS detector [? ]

### 3.2.1 Magnet

The solenoid magnet [?] is the core element of the CMS experiment that has constrained the design of the rest of the detector.

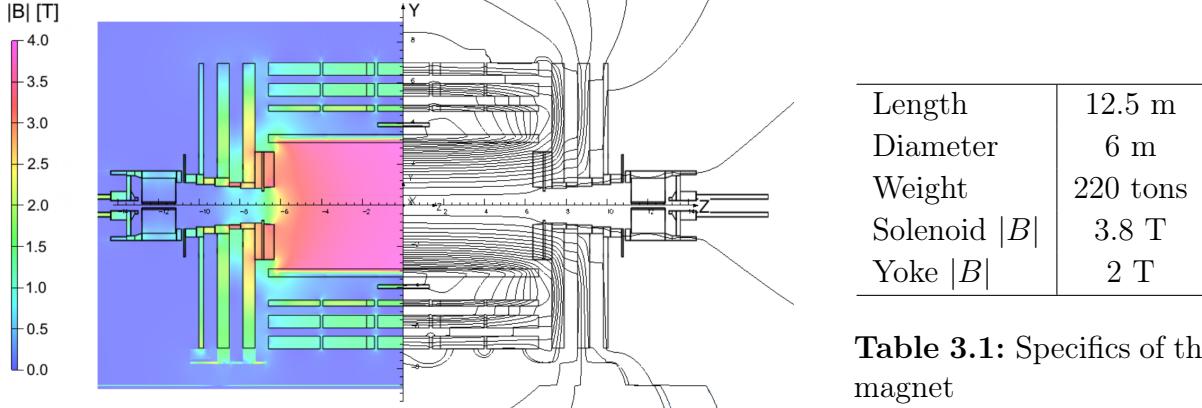
The purpose of the magnet is to bend the tracks of charged particles, due to the Lorentz force, and allow us to measure the transverse momentum  $p_T$  of the particle. The stronger the magnetic field, the better will be the  $p_T$  resolution.

The magnet's coil is made up of multiple layers of superconducting niobium-titanium (NbTi) wires wound together that are cooled using liquid helium down to 4.7 K, allowing the coil to work in the superconductive regime.

The solenoidal geometry generates a quasi-uniform magnetic field inside the volume of the magnet of 3.8 T. The solenoid is surrounded by the steel *return yoke*, composed of five three-layered dodecagonal barrel wheels and three endcap disks at each end. The

<sup>2</sup>The nuclear interaction length  $\lambda$  of a certain material is the mean distance that hadronic particles travel in the material before interacting with a nucleus

yoke contributes to only 8% of the central magnetic flux density [? ]. Its main role is to increase the field homogeneity inside the solenoid, closing the magnetic circuit and preventing magnetic leakage. It also hosts the muon system, embedding it in a 2 T magnetic field.

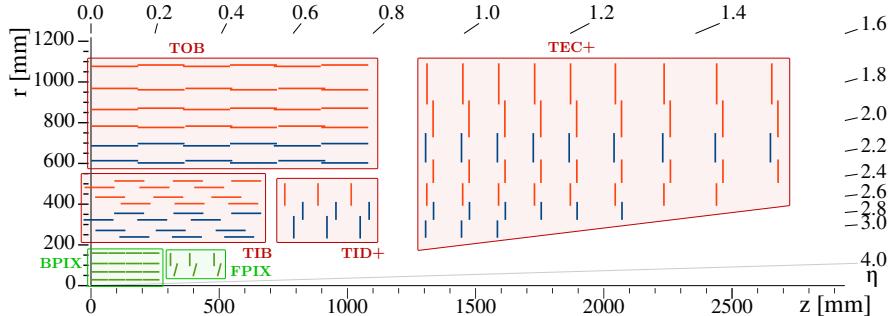


**Figure 3.4:** Field lines and magnitude of the magnetic field of the CMS solenoid magnet [? ].

**Table 3.1:** Specifics of the magnet

### 3.2.2 Tracker

The silicon tracker [? ? ] is the closest subdetector to the IP, covering the region  $|\eta| < 2.5$ , and allows us to reconstruct the tracks of charged particles and their transverse momentum, measuring the curvature of the track caused by the magnet.



**Figure 3.5:** Schematic view of a quarter of the CMS tracker in the  $r - \eta$  plane. Pixel modules are shown in green, single-sided strip modules in red and double-sided strip modules in blue [? ].

All the modules that compose the tracker are silicon semiconductor sensors and are divided into pixel modules that contain about 66 million pixel cells each with a size of  $100 \times 150 \mu\text{m}^2$ , and strips modules that contain strips of thickness from  $80 \mu\text{m}$  to  $205 \mu\text{m}$  and a length that increases with the pseudorapidity.

The *pixel tracker* [? ] is the nearest component to the beam pipe. Its scope is not only limited to collecting the particle hits to reconstruct the tracks, but also to identifying and reconstructing the primary and secondary vertices. The vertex reconstruction is crucial to mitigate PU effects and to identify the decay of displaced particles, enhancing the b-tagging capabilities. The high granularity of the pixel modules

allows achieving these goals. Furthermore, being closest to the IP, the inner tracker has to be designed to be more radiation tolerant than the outer tracker.

The pixel tracker is composed of different parts:

- **Barrel pixel (BPIX)**: 4 concentric layers of pixel modules.
- **Forward pixel (FPIX)**: 6 disks of pixel modules

	Radius [mm]	z position [mm]	Number of pixel modules	Dose [Mrad]
<b>BPIX</b>				
L1	29	-270 to +270	96	100
L2	68	-270 to +270	224	47
L3	109	-270 to +270	352	22
L4	160	-270 to +270	512	13
<b>FPIX</b>				
D1 inner	45–110	$\pm 338$	88	21-106
D1 outer	96–161	$\pm 309$	136	13-28
D2 inner	45–110	$\pm 413$	88	21-106
D2 outer	96–161	$\pm 384$	136	13-28
D3 inner	45–110	$\pm 508$	88	21-106
D3 outer	96–161	$\pm 479$	136	13-28

**Table 3.2:** Summary of average r, z positions, number of modules and dose for the four BPIX layers and the six FPIX rings [? ].

The *strip tracker* [? ] is composed of  $\sim 20000$  strip modules. Depending on the positions, the geometry of the sensors and the number of strips vary: In the barrel region, the sensors are rectangular, while the endcap sensors are trapezoidal.

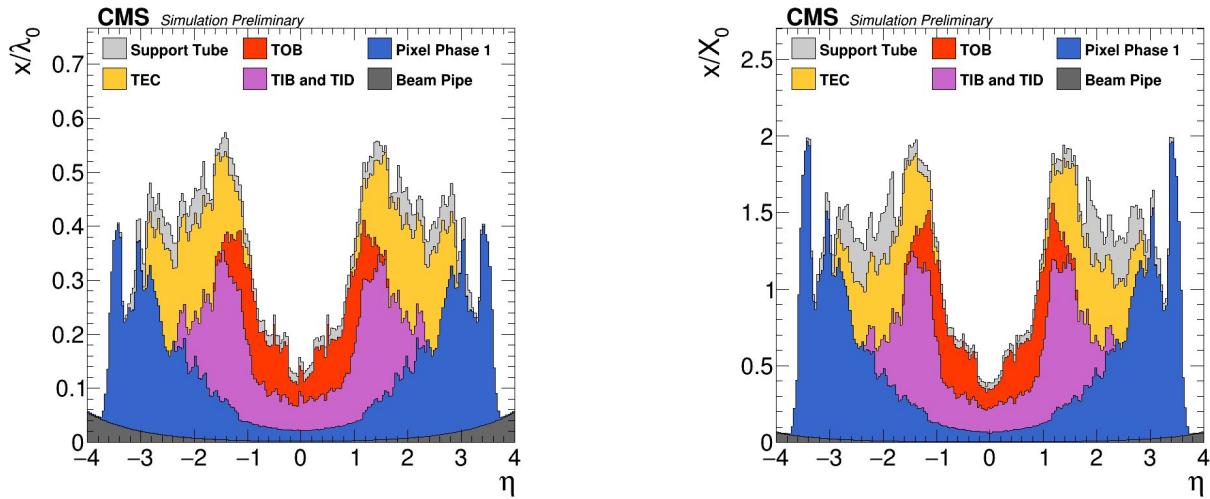
Some modules are double-sided (DS) strip modules, i.e. two single-sided (SS) strip modules mounted back to back with a strip inclination of  $5.7^\circ$  against each other (stereo configuration) to resolve the ambiguity on the z coordinate.

- **Tracker inner barrel (TIB)**: 2 layers of DS + 2 layers of SS strip modules
- **Tracker inner disk (TID)**: 2 layers of DS + 1 layers of SS strip modules
- **Tracker outer barrel (TOB)**: 2 layers of DS + 4 layers of SS strip modules
- **Tracker endcap (TEC)**: 3 layers of DS + 4 layers of SS strip modules

The design of the tracker has to consider the multiple scattering, bremsstrahlung and photon conversion that can occur in the tracker itself and that can alter the trajectory and the measurement of the energy deposited in the calorimeter. This implies that the tracker should minimize the material budget without sacrificing the tracking capabilities.

Layer	Radius[mm]	Type	Modules	Pitch[mm]	Strips
TIB1	250	DS	336	80	768
TIB2	340	DS	456	80	768
TIB3	430	SS	552	120	512
TIB4	520	SS	648	120	512
TOB5	610	DS	504	122/183	768/512
TOB6	696	DS	576	122/183	768/512
TOB7	782	SS	648	183	512
TOB8	868	SS	720	183	512
TOB9	965	SS	792	122	768
TOB10	1080	SS	888	122	768
TID1	277	DS	144	81-112	768
TID2	367	DS	144	113-143	768
TID3	447	SS	240	124-158	512
TEC1	277	DS	144	81-112	768
TEC2	367	DS	288	113-143	768
TEC3	447	SS	640	124-158	512
TEC4	562	SS	1008	113-139	512
TEC5	677	DS	720	126-156	768
TEC6	891	SS	1008	163-205	512
TEC7	991	SS	1440	140-172	512

**Table 3.3:** Summary of average r, type, number of modules, pitch, and number of strips for the strip tracker [? ].



**Figure 3.6:** Stacked histogram of the thickness of the subparts of the tracker traversed by a particle originated in the IP in function of  $\eta$  ((left) in nuclear interaction length units<sup>3</sup>, (right) in radiation length units) [? ].

### 3.2.3 Electromagnetic calorimeter

The CMS electromagnetic calorimeter (ECAL) [? ? ] is the subdetector dedicated to the identification and measurement of the energy of particles that interact primarily via the electromagnetic interaction.

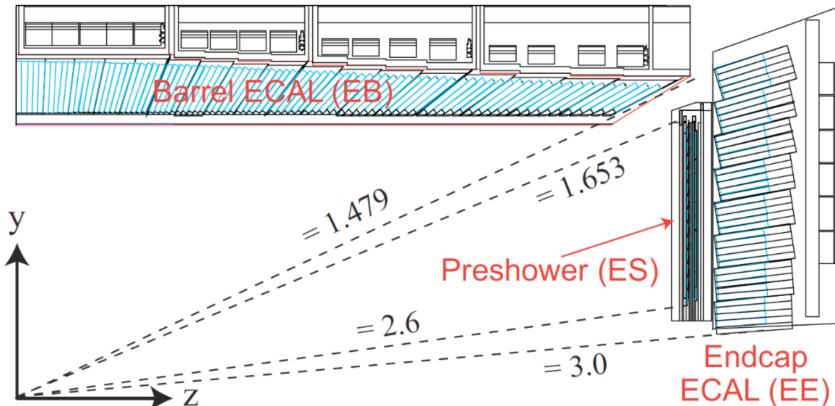
One of the main goals of the CMS collaboration was the Higgs boson discovery and, to

<sup>3</sup>The radiation length  $X_0$  of a given material is the length for which, for an electromagnetic particle that travels into the material, holds  $E(z) = E_0 e^{-z/X_0}$  where  $E(z)$  is the energy of the particle that is a function of the length traveled into the material and  $E_0$  the initial energy.

maximize the discovery sensitivity on the channel  $H \rightarrow \gamma\gamma$ , the ECAL should have met some requirements as high granularity and hermeticity, fast response ( $\sim 25\text{ns}$ ), particle identification/isolation/energy at trigger level and linear response in a large energy range (5 GeV to 5 TeV).

To comply with these requirements, the ECAL was built using PbWO<sub>4</sub> crystals: a high-density material ( $\delta = 8.28\text{g/cm}^3$ ) with a short radiation length ( $X_0 = 0.85\text{cm}$ ) and a small Moliere radius <sup>4</sup> ( $R_M = 2.19\text{cm}$ ).

The crystals emit optical light by scintillation in a short time (80% of the light in 25ns, which is the time separation between two different bunch crossing) but, the light yields of only  $\sim 100\gamma/\text{MeV}$  force the usage of an active photodetector and, due to the presence of the magnetic field, the silicon photomultipliers (SiPM) were the most logical choice. The ECAL is made of a barrel that covers the  $|\eta| < 1.479$  region and two endcaps that cover the region  $1.479 < |\eta| < 3.0$ .



**Figure 3.7:** Schematic view in the r-z plane of a quarter of the CMS ECAL [? ].

The ECAL endcaps are composed of 14648 crystals of size  $3 \times 3 \times 22\text{cm}^3 (\sim 25X_0)$  and the ECAL barrel is composed of 61200 crystals of size  $2.2 \times 2.2 \times 23\text{cm}^3 (\sim 26X_0)$ . The crystals are not aligned to the IP to avoid cracks aligned with particle trajectories. Furthermore, a pre-shower silicon detector made of 4288 strip modules is placed in front of the endcaps to improve the  $\gamma - \pi^0$  separation <sup>5</sup>.

The energy resolution of the PbWO<sub>4</sub> crystals measured with an electron beam without magnetic field and radiation damage was [? ]

$$\frac{\sigma_E}{E} = \frac{2.8\%}{\sqrt{E[\text{GeV}]}} \oplus \frac{0.127\text{GeV}}{E[\text{GeV}]} \oplus 0.3\% \quad (3.3)$$

where the first term is the stochastic term, the second is the noise term and the third is the constant term that contains the calibration errors and the leakage of the energy.

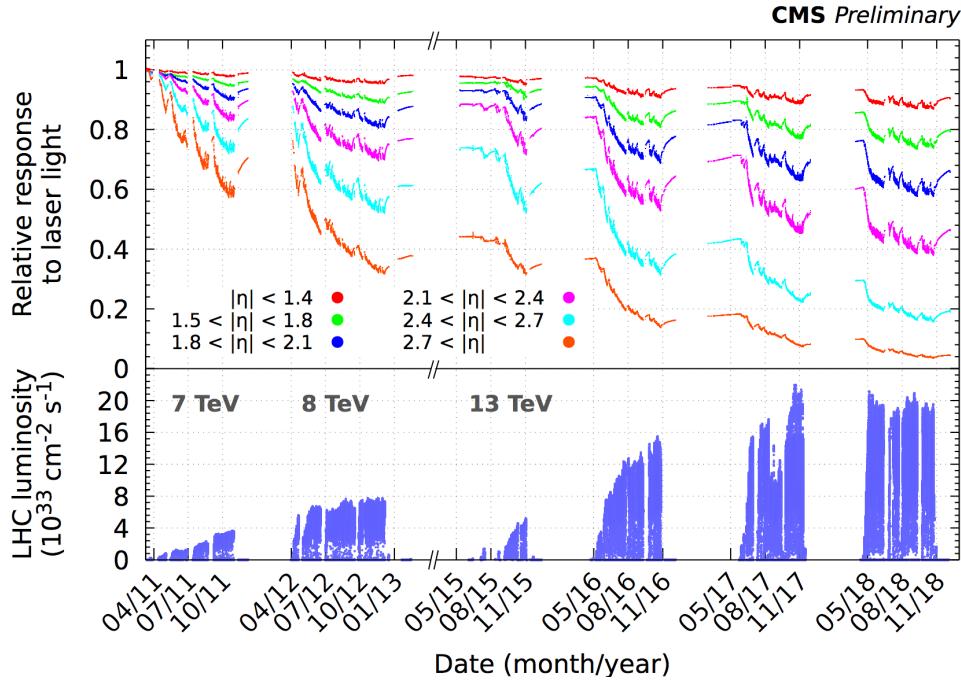
In the CMS detector, the ECAL energy resolution is worse mainly due to the presence of the tracker between the ECAL and the IP and due to the transparency variations induced

<sup>4</sup>The Moliere radius  $R_M$  is the mean radius of the cylinder that contains 90% of the shower, which tends to disperse due to multiple scattering

<sup>5</sup>The  $\pi^0$  decay  $\pi^0 \rightarrow \gamma\gamma$  can cause the misidentification of a pion into a photon.

by radiation damage and the aging of photomultipliers, but, during the Run1, the ECAL was able to achieve the  $\sim 1\%$  energy resolution for high-energy electrons [? ].

The radiation damage of PbWO<sub>4</sub> crystals is constantly monitored by a laser system in various  $\eta$  intervals as shown in Fig. ??.



**Figure 3.8:** ECAL relative response to the laser monitoring system from 2011 to 2019 [? ].

### 3.2.4 Hadronic calorimeter

Hadrons usually don't deposit all their energy in the ECAL, especially neutral hadrons, therefore the ECAL is surrounded by the hadronic calorimeter HCAL that allows the measurements of the jets' energy and the missing energy.

The CMS HCAL [? ?] is divided into four subdetectors: the hadron barrel (HB), the hadron endcaps (HE), the hadron outer (HO), and the hadron forward (HF).

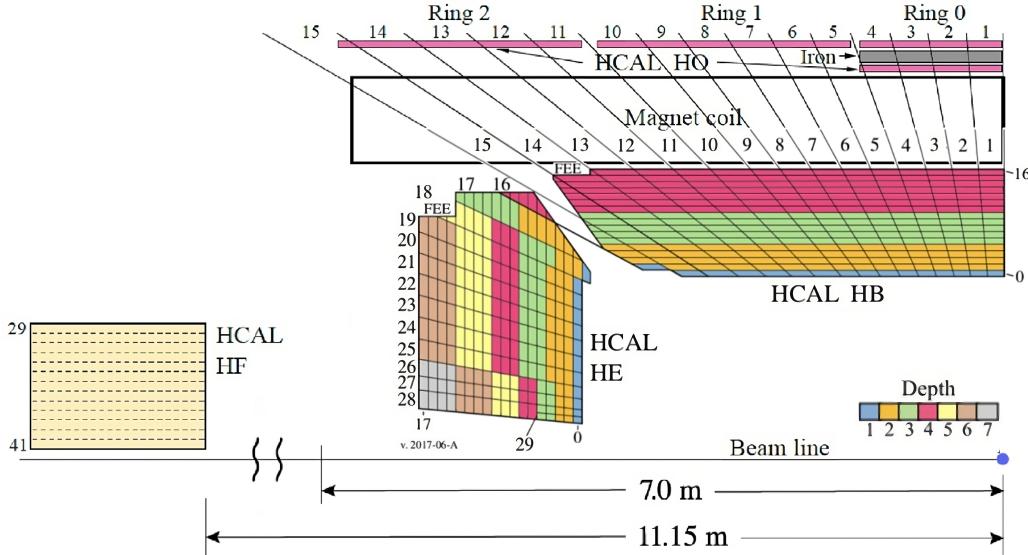
The HB is a sampling calorimeter made of brass absorbers and plastic scintillators that act as active material. The brass was chosen due to its short interaction length and because it is non-magnetic.

It covers the pseudorapidity region  $|\eta| < 1.3$  and it extends in a radius  $1.77m < R < 2.95m$ . The barrel is divided into sections of size  $(\Delta\eta, \Delta\phi) = 0.087 \times 0.087$  and is composed of eight 50.5 mm thick and six 56.5 mm thick brass plates alternating with the scintillating tiles, that are connected through optical cables to SiPM sensors to collect and amplify the light.

The total thickness of the HB in nuclear interaction length units at  $|\eta| = 0$  is  $5.82\lambda_I$  that increases up to  $10.6\lambda_I$  at  $|\eta| = 1.3$ .

In addition to this, the ECAL barrel adds  $\sim 1.1\lambda_I$  thickness of material.

Also the HE is a sampling calorimeter made of brass absorbers and plastic scintillating



**Figure 3.9:** Schematic view in the r-z plane of a quarter of the CMS HCAL and depth segmentation after the Phase I update [? ].

tiles. It is divided into 18 sectors in  $\phi$  and into 14 sectors in  $\eta$  (the size increases with  $\eta$ ) and is placed 4006.5 mm from the interaction point. It covers the  $1.3 < |\eta| < 3.0$  region and including the ECAL endcaps, its length in interaction lengths units is  $\sim 10\lambda_I$

The HO is placed outside the solenoid magnet, and its purpose is to catch the tails of the hadronic showers that are not contained by the HB. The HO is composed by five scintillator rings with a radius of 4.07 m that cover the region  $|\eta| < 1.26$ . The central ring that covers the thinner HCAL region is equipped also with a 19.5 cm thick iron absorber and an additional scintillator, as shown in fig Fig. ??.

The HF is a sampling calorimeter is located 11150 mm away from the interaction point and covers the region  $3 < |\eta| < 5.2$ . It is composed of 1.65 mm steel absorber plates and 0.6 mm thick quartz fibers that emit light by Cherenkov effect. It is also used for the online monitoring of luminosity.

The combined energy resolution of the ECAL and the HCAL was measured using a pion beam from 2 GeV to 350 GeV [? ]

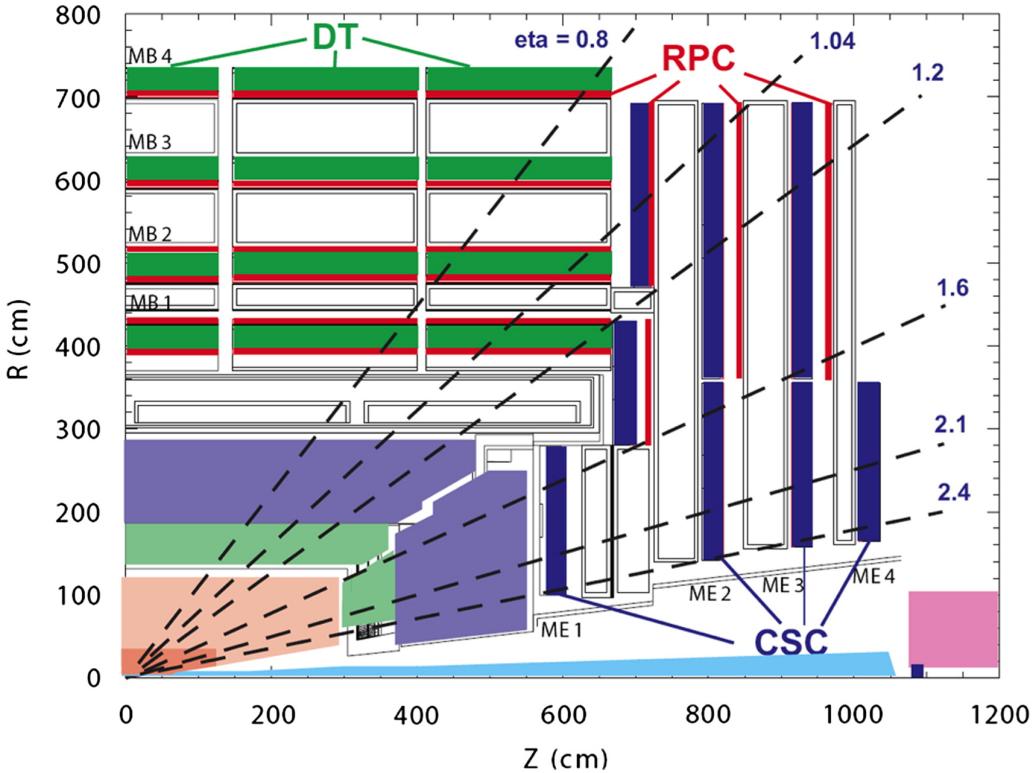
$$\frac{\sigma_E}{E} = \frac{84.7\%}{\sqrt{E}} \oplus 7.4\% \quad (3.4)$$

The energy resolution is mainly affected by the sampling fluctuations and the noise term is negligible.

### 3.2.5 Muon system

The CMS muon system [? ] is embedded in the iron return yoke and its scope is to collect muon hits that are used, in combination with information from the tracker, to enhance the identification of muons and momentum resolution of high-energy muons. It is divided into five separate wheels in the barrel, containing four layers of detectors, and

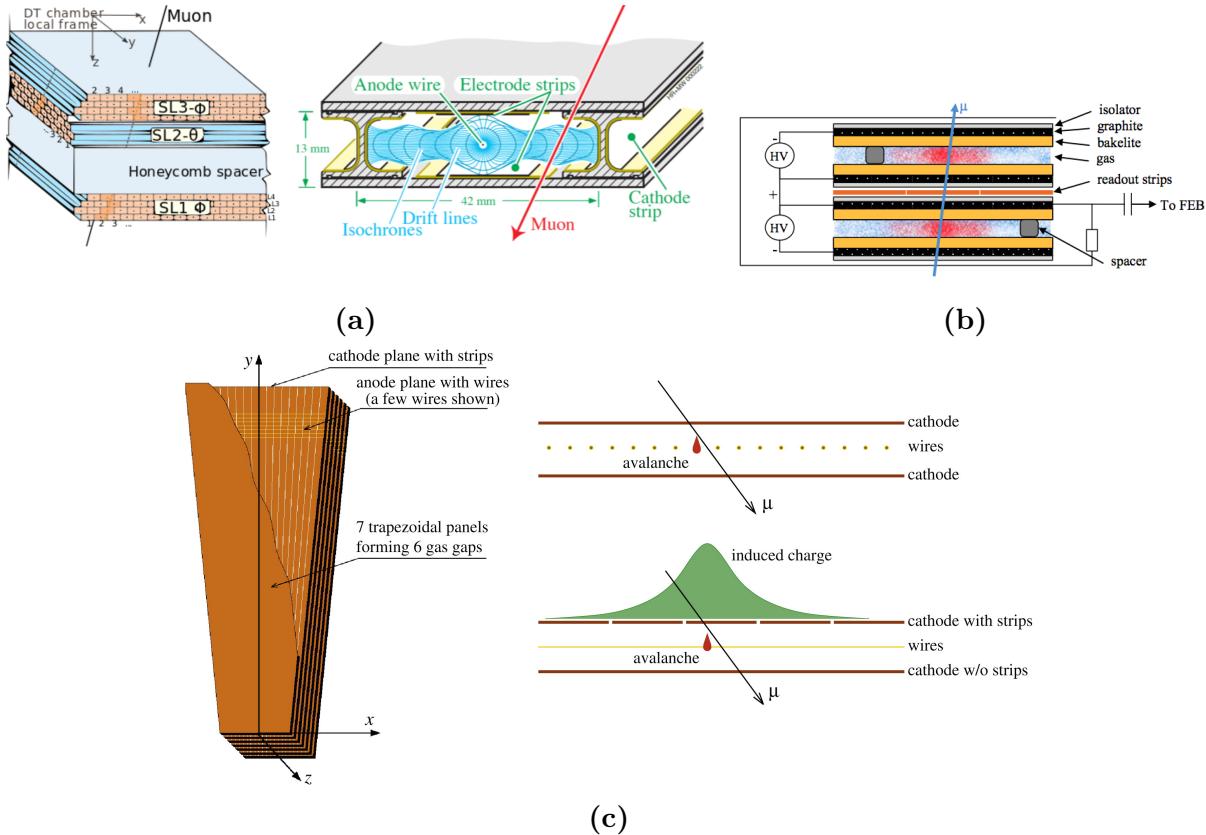
four independent endcaps for each side.



**Figure 3.10:** Schematic view in the  $r - z$  plane of a quarter of the muon system [? ].

Three different gaseous detector technologies are employed:

- **Drift tubes chambers (DT):** The 250 DT chambers are divided into five wheels and cover the region  $|\eta| < 1.2$ . Each wheel is divided into 12 sectors covering  $\Delta\phi = 30$  around the interaction point, and each sector is organized into four muon barrels (MB). Each chamber consists of 8 or 12 aluminum layers with up to 90 tubes.
- **Cathode strip chambers (CSC):** CSCs consist of arrays of anode wires crossed with copper cathode strips within a gas volume. The 486 CSCs are organized in 4 layers, placed in the endcap disks, and extend the muon system coverage up to  $0.8 < |\eta| < 2.4$
- **Resistive plate chambers (RPC):** The RPCs are located both in the barrel and in the endcap region, up to  $|\eta| < 1.6$ . RPCs consist of two charged parallel high resistive plastic plates filled with gas. The read-out is performed by aluminum strips separated from the graphite coating. The scope of the RPCs is to provide, in combination with the DTs and the CSCs, a fast signal to exploit for the muon trigger system.



**Figure 3.11:** Schematics view of the (a) drift tubes chamber [? ], (b) resistive plates [? ], (c) cathode strips chamber [? ].

### 3.2.6 Trigger and data acquisition

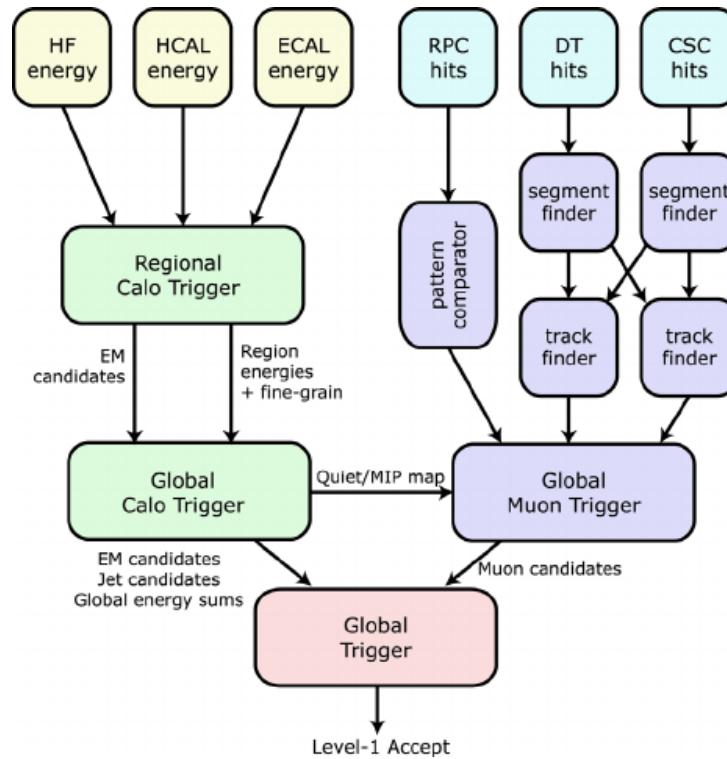
At LHC, bunch crossing occurs each 25ns. Considering the raw event size of 1 MB [? ], this would lead to a data throughput of  $\sim 1$  TB/s at the nominal luminosity, that, not only will saturate the bandwidth of the data acquisition system but also would make it impossible to save on tape all the recorded data.

The scope of the CMS trigger system is to select only the events relevant to the CMS physics program. The trigger system is divided into the level 1 low-level trigger (L1) and the high-level trigger.

The CMS L1 Trigger [? ] operates as a hardware-based trigger system exploiting FPGA boards. It relies on custom algorithms that have to make decisions within  $3.2\mu s$ , the maximum time that the raw data can occupy the buffers.

The L1 trigger collects raw data from the front-end electronics of the calorimeters and the muon system that converts the analog signals to digital.

The algorithms roughly estimate the energy deposit and the trajectory of the particles in the subdetectors to select EM, jet, and muon candidates. Thereafter, the information from the muon system and from the calorimeters are combined into the global trigger that chooses whether to discard the event or to pass it to the HLT trigger for further filtering. The L1 trigger reduces the output rate from the bunch crossing rate of 40MHz to 100kHz.



**Figure 3.12:** Architecture of the Level-1 Trigger [? ].

The HLT system [? ] is a software-based trigger implemented on a computer farm composed of 16000 CPUs that decreases the output rate from the 100 kHz of the L1 trigger to 1 kHz. It employs the same algorithms used by the offline reconstruction but adapted to the timing constraint of the HLT trigger. It has access to the full detector readout, and events are accepted or rejected as soon as there is enough information to decide. The raw data of the events that are accepted by the trigger are saved on tape and the events are reconstructed offline when the computing resources are available.

# *Chapter 4*

---

## Event simulation and reconstruction

---

### 4.1 Event reconstruction

The event reconstruction is the identification of all the particles and their kinematics from each pp collision, starting from the raw data that is just a collection of hits and energy clusters.

CMS is a highly segmented detector, designed to give a different signature for each kind of particle. All the charged particles produce tracks, electrons, and photons produce a shower in the ECAL, hadrons start showering in the ECAL and are absorbed in the HCAL, while muons, due to their mass, are minimum ionizing particles that penetrate all the detector, leaving tracks in the tracker and the muon system (Fig. ??).

The informations of the different subdetectors are linked together to improve the resolution of the detector and to consider other signatures (e.g. reconstructing a photon that has converted an electron-positron pair in the tracker). This approach is called Particle Flow (PF) [? ], and each reconstructed particle is called a PF candidate.

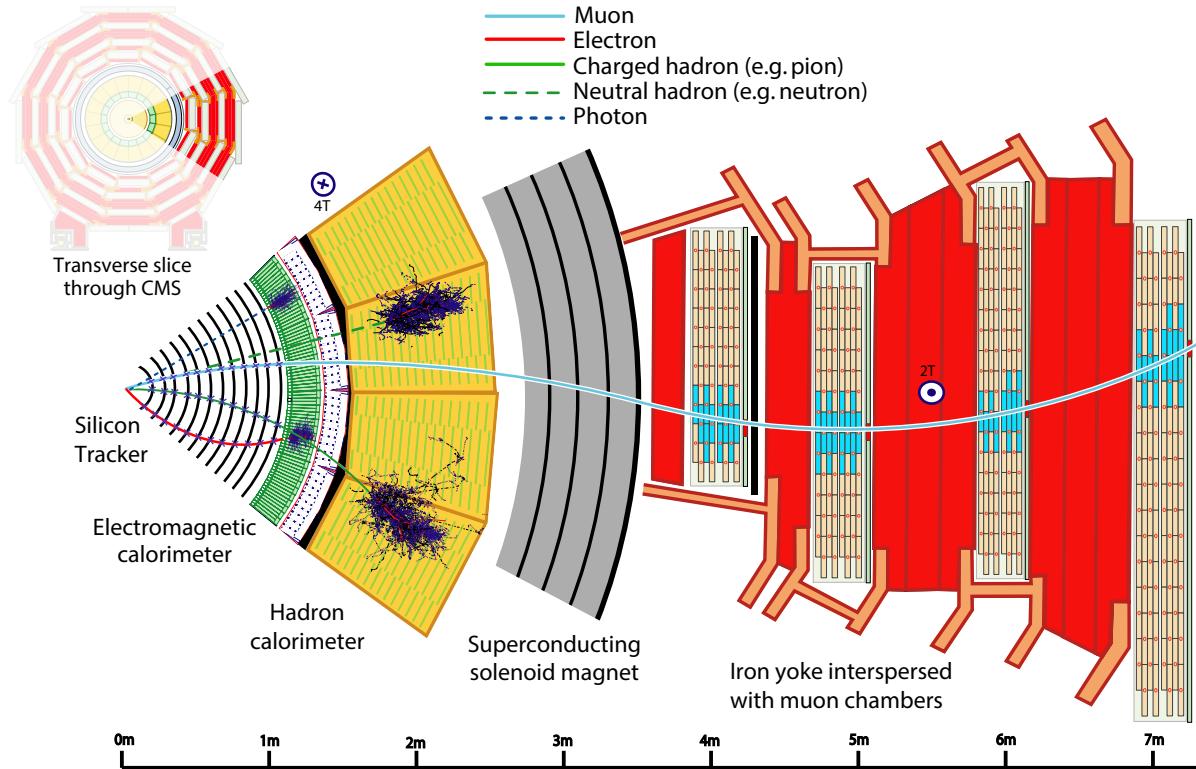
#### 4.1.1 Tracking

The algorithm used to reconstruct the trajectories of charged particles and their momentum is the Combinatorial Track Finder (CTF) [? ], an adaptation of the Kalman Filtering (KF) [? ] algorithm. At the beginning, a few hits in consecutive layers of the tracker are used as seed of the KF algorithms and then a final fit is performed to estimate the position of the vertex, the direction, and the momentum of each charged particle.

**Seeding and iterations** This process is iterated in multiple steps, starting with searching for tracks that are easier to find (e.g. particle with large  $p_T$  produced in the IP), removing the associated hits to reduce the combinatorial multiplicity, and searching for other kinds of tracks using different seeding criteria (Table ??).

**Track finding** The track finding is divided into four steps:

1. Using the parameter of the track candidate (given by the seed at the beginning), determine which successive layer of the tracker can be intersected by the trajectory,



**Figure 4.1:** Signatures of different types of particles in a slice of the CMS detector. The muon and the charged pion are positively charged, and the electron is negatively charged. [? ]

Iteration	Name	Seeding	Targeted Tracks
1	InitialStep	pixel triplets	prompt, high $p_T$
2	DetachedTriplet	pixel triplets	from b hadron decays, $R \leq 5\text{cm}$
3	LowPtTriplet	pixel triplets	prompt, low $p_T$
4	PixelPair	pixel pairs	recover high $p_T$
5	MixedTriplet	pixel+strip	triplets displaced, $R \leq 7\text{cm}$
6	PixelLess	strip triplets/pairs	very displaced, $R \leq 25\text{cm}$
7	TobTec	strip triplets/pairs	very displaced, $R \leq 60\text{cm}$
8	JetCoreRegional	pixel+strip pairs	inside high $p_T$ jets
9	MuonSeededInOut	muon-tagged	tracks muons
10	MuonSeededOutIn	muon detectors	muons

**Table 4.1:** Seeding configuration of ten iterations with different target tracks.  $R$  is the target distance of the track origin from the beam axis [? ].

considering its uncertainties.

The trajectories are assumed to be perfect Helix, neglecting the multiple scattering, allowing us to use fast analytical methods.

2. Look for a compatible silicon module <sup>1</sup> in the layer such that its boundary is less than  $3\sigma$  from the trajectory. The trajectory and its uncertainties are propagated to the sensor surface.

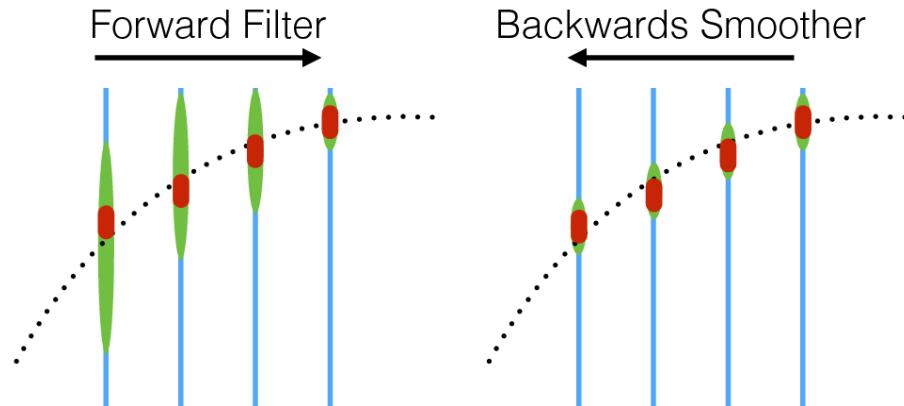
<sup>1</sup>Silicon modules usually overlap, so in this step the algorithms look for groups of modules, also to ease the computational effort.

3. A  $\chi^2$  test is used to check which of the hits are compatible with the track candidate, considering both the hit and the trajectory uncertainties.
4. Update the track candidate and iterate

**Track fitting** To determine the track parameters, after all hits of each track are assigned by the track finder, a new KF fit is initialized at the location of the innermost hit. The fit proceeds through all the hits, from inside outwards.

This filter is followed by a smoothing stage that consists of a second KF initialized with the result of the first one and is run backward. The track parameters are obtained from a weighted average of the track parameters of these two filters.

During the track fitting stage, the effects of interaction with materials and inhomogeneities of the magnetic field are considered.



**Figure 4.2:** Visualization of the forward filter and the backwards smoother filter. The red points indicate the measurements and their uncertainties on each layer, while the green points indicate the predictions [? ].

## Muon tracks

The reconstruction of muon tracks combines the information from the tracker and the muon system. Three types of tracks are defined:

- **Standalone muons:** Tracks built using only the hits in the muon system.
- **Tracker muons:** Tracks built using only the hits in the tracker that match the track segments in the muon system.
- **Global muons:** Tracks fitted combining the tracker and the muon system hits, improving the  $p_T$  resolution of the tracking detector.

Usually, there is no reason to not use global muons, but low  $p_T$  muons can be reconstructed as tracker muons because of the large multiple scattering in the return yoke.

## Electron tracks

The initial identification of electrons is performed in two ways: The first consists of estimating the electron momentum from the  $\phi$  spreading of the ECAL cluster and finding a track compatible with the cluster. The second method consists of finding ECAL

clusters compatible with tracks extrapolated from the tracker.

Electrons, traversing the tracker, radiate energy by Bremsstrahlung. For this reason, the KF doesn't perform well, so, tracks that are identified as electrons are refitted with the Gaussian sum filter (GSF) algorithm [? ] that approximates the energy loss of electrons, described by the Bethe-Heitler distribution [? ], as a Gaussian mixture.

At each layer, a set of helical track segments is computed for a different value of the energy loss, and it's weighted with the corresponding probability of occurring. These informations are used to decrease the energy of the particle and update the trajectory uncertainties at each layer.

## Vertex reconstruction

Once all the tracks are reconstructed, those that come from the same vertex are clustered together and a fit is performed to determine the vertex position.

The fit is done using the simulated annealing algorithm [? ] that estimates the vertex position by minimizing a free energy function and associates to each track the probability of belonging to the i-th vertex.

### 4.1.2 Calorimeter clustering

The clustering in calorimeters is necessary to reconstruct photons (in the ECAL) and neutral hadrons (in the HCAL), objects that don't leave any signature in the tracker system.

The ECAL is also used to improve the identification of electrons, matching the tracks with the ECAL energy clusters and looking for bremsstrahlung photons, and both the ECAL and the HCAL to improve the reconstruction of charged hadrons, especially the high energetic ones.

Cluster seeds are built from cells with an energy larger than a given threshold and larger than the energy of the neighboring cells.

Starting from the seeds, topological clusters are built by aggregating neighboring cells that have an energy larger than twice the noise level.

Assuming that the energy deposit in the M cells contained in the topological cluster arises from N Gaussian energy deposits, one for each seed in the topological cluster, a Gaussian mixture model is used to reconstruct the cluster. The parameters of the Gaussian are the location parameters in the  $(\eta, \phi)$  plane and the amplitude, while the width has a fixed value, different for each subdetector.

The energy and the position of the seeds are used as initial values for the parameters of the Gaussian mixture that are estimated through an analytical maximum likelihood fit.

	ECAL		HCAL		Preshower
	barrel	endcaps	barrel	endcaps	
Cell E threshold (MeV)	80	300	800	800	0.06
Seed # closest cells	8	8	4	4	8
Seed E threshold (MeV)	230	600	800	1100	0.12
Seed ET threshold (MeV)	0	150	0	0	0
Gaussian width (cm)	1.5	1.5	10.0	10.0	0.2

**Table 4.2:** Seeding and clustering parameters for the ECAL, the HCAL, and the preshowers [? ].

### 4.1.3 ParticleFlow linking

The PF elements are initially built separately in each subdetector and each of them can be linked with other PF elements in their k-nearest neighborhood [? ] (kNN) in the  $(\eta, \phi)$  plane.

The linking algorithm assigns a distance between each possible pair of elements that quantifies the quality of the link. All elements that are linked together form a PF block.

**Track to track linking** Charged particles that originate from a common secondary vertex and have an invariant mass greater than 0.2 GeV are linked together to reconstruct a nuclear interaction or a displaced decay.

**Track-Calorimeter linking** A track is linked to a calorimeter cluster if the position of the track in the  $(\eta, \phi)$  plane ( $(x, y)$  for the endcaps), extrapolated from the hit in the outermost layer, is contained in the union of the cluster areas of the ECAL and HCAL. The cluster areas are enlarged to take into account the uncertainties in the position, multiple scattering, and the presence of cracks between cells.

The link distance is defined as the distance between the track and the cluster in the  $(\eta, \phi)$  plane for the barrel and in the  $(x, y)$  plane for the endcaps, keeping only the link with the smallest distance for each element.

Furthermore, photons emitted tangent to electron tracks (with  $|\Delta\eta| < 0.5$ ) are collected as Bremsstrahlung photons. They have a significant probability of converting into electron-positron pairs, so a dedicated conversion finder algorithm [? ], based on the iterative tracker procedure (discussed in sec ??), is used to link  $e^+e^-$  tracks to the corresponding photons.

In this step, the clusters associated with electrons and Bremsstrahlung photons are clustered into superclusters.

**Cluster to cluster linking** When the cluster position of the more granular calorimeter is contained in the cluster area of the less granular calorimeter, the link is established. Also in this case, the distance is defined as the distance of the cluster positions in the  $(\eta, \phi)$  plane for the barrel and in the  $(x, y)$  plane for the endcaps and just the link with the smallest distance is kept.

**Track-muon linking** As explained in sec ??, the PF elements in the tracker are linked with the ones in the muon system through spatial matching to allow the global fitting of

the muon tracks.

#### 4.1.4 ParticleFlow reconstruction

In each PF block, muon candidates are identified and reconstructed first. The following steps deal with the reconstruction of electrons, photons, and hadrons. At each step, the PF blocks used to reconstruct the candidate are removed. In particular, the electron reconstruction is performed before the photon reconstruction to remove elements associated with Bremsstrahlung photons.

At the end, when all blocks have been processed, the event is revisited in a post-processing step.

**Muons** To avoid the misidentification of hadrons in muons (that can occur for high-energy hadrons that reach the muon system), isolated global muons are selected by considering all the tracks and energy deposits within  $\Delta R < 0.3$  from the muon direction, with the constraint that: the scalar sum of the  $p_T$  and the  $E_T$  of other elements have to be less than the 10% of the muon  $p_t$ ; there are at least three track segments in the muon system; and the calorimeter deposit match the  $p_t$  estimate from the track.

Muons originating from the semileptonic decay of hadrons require more stringent criteria.

**Electrons and photons** Because electrons emit Bremsstrahlung photons and photons convert to electron pairs, the reconstruction of electrons and photons is performed in the same step.

For each supercluster, if it matches a GSF track, an electron candidate is built, if not, a photon candidate is built.

All the ECAL clusters in the PF block linked to the supercluster, to the electron track, or tracks belonging to a photon conversion are associated with the candidate.

Electron candidates must satisfy additional identification constraints on the amount of energy radiated, the distance between the track and the ECAL cluster in the  $(\eta, \phi)$  plane, the ratio of the energy released in the ECAL and the HCAL, the number of hits and goodness of fit of the track reconstructed both with the KF and the GSF. The final identification of the electron is performed by a Boosted decision tree (BDT) that is trained on all these inputs.

Photons must satisfy isolation criteria that just require the isolation from other tracks and clusters.

**Hadrons** The remaining PF blocks in the ECAL that are not linked with a track are identified as non-isolated photons originating from  $\pi^0$  decays, while the ones in the HCAL as neutral hadrons.

The remaining HCAL clusters are linked to tracks and, eventually, to other remaining ECAL clusters to identify charged hadrons.

It is relevant to note that, outside the tracker coverage, charged and neutral hadrons cannot be distinguished anymore.

**Missing transverse momentum** The neutrinos and other hypothetical exotic particles do not interact with the detector material and the only quantity that we can measure is the missing transverse energy <sup>2</sup>.

The missing transverse momentum is defined as the opposite vector to the vectorial sum of the transverse momentum of all the particles.

$$\vec{p_T}^{\text{miss}} = - \sum_i^{N_{\text{part}}} \vec{p_{T,i}} \quad (4.1)$$

**Post-processing** The misidentification of high  $p_T$  particles or the presence of cosmic rays can lead to a large reconstructed missing energy  $\vec{p_T}^{\text{miss}}$ . The post-processing phase consists of selecting high  $p_T$  particles, computing the correlation between the  $p_T$  of the particle and the reconstructed missing energy, and changing the identification and the reconstruction of the particles if this at least halves the missing energy amplitude.

### 4.1.5 Jets

To infer the kinematics of the parton produced in the hard-scattering, hadrons that originated from the hadronization of the parton have to be clustered together into jets. The clustering algorithm for jets must satisfy two requirements:

- **Collinear safety:** Collinear splitting of the momentum of components of the jet must not change the jet.
- **Infrared safety:** Soft emission must not change the jet.

The current algorithm used by the CMS collaboration is the anti- $k_T$  algorithm [? ]. The anti- $k_T$  introduces the custom distances  $d_{ij}$  between hadronic PF candidates, and  $d_{iB}$  between the PF candidate and the beam spot

$$d_{ij} = \min(p_{Ti}^{-2}, p_{Tj}^{-2}) \cdot \frac{\Delta R_{ij}^2}{R^2} \quad (4.2)$$

$$d_{iB} = p_{Ti}^{-2} \quad (4.3)$$

$R$  is a parameter that regulates the dimension of the jet cone. In CMS, jets are constructed using  $R=0.4$  (AK4 jets) but, to study boosted objects that cause jets to merge,  $R=0.8$  (AK8 jets a.k.a. fatjets) is another choice.

The algorithm performs as follows

This algorithm tends to cluster soft particles with hard ones instead of clustering soft particles with each other and tends to group all the components in cones.

The energy scale and resolution of the jet, and consequently also the missing energy, is corrected afterward, analyzing the composition of the jets [? ].

---

<sup>2</sup>We can measure only the transverse component because, at a hadron collider, it is impossible to know the initial state due to the composite nature of the protons

---

**Algorithm** anti- $k_T$  algorithm

---

```

1: Compute all the  $d_{ij}$  and  $d_{iB}$ 
2: while # candidates > 0 do
3:   Select the couple of PF candidates with the minimum  $d_{ij}$ 
4:   Sum their Lorentz vector and replace the two PF candidates with this pseudojet.
5:   Recompute  $d_{ij}$  and  $d_{iB}$ 
6:   if  $d_{ij} < d_{iB} \forall j$  then
7:     The i-th candidate is defined as a jet and removed.
8:   end if
9: end while

```

---

Jet can also contain contributions from pileup interaction that occurred in the current bunch crossing (in-time pileup) or the previous bunch crossing (out-of-time pileup).

In-time pileup contributions from charged hadrons, that are over 50% of the in-time pileup contribution, can be avoided by subtracting the contribution of charged hadrons originating from pileup vertices while out-of-time pileup is determined by fitting the pulse shape of the calorimeters signals and subtracted to the jet energy afterward.

The remaining contributions from pileup are parametrized as a function of  $p_T$ ,  $\eta$ , and the jet area.

## Flavour tagging

To fully characterize the jet, we must understand which is the flavor of the parton that generated the jet.

The identification of b (c) jets (b/c tagging) can be performed by studying the b and c hadrons that compose it.

Heavy flavor hadrons have a relatively large lifetime that makes them travel the detector for  $\sim 0.3 - 0.5 mm$ , which allows us to use displaced tracks and secondary vertices as a signature to recognize heavy flavor jets.

Despite that, using a single observable to determine the jet flavor is not efficient enough and a multivariate analysis is required.

The model used in the Run II analysis by the CMS collaboration is DEEPJET [?], a neural network (see sec. ??) trained using labeled MC samples of AK4 jets of different flavors.

DEEPJET uses  $\sim 650$  observables as input, divided into global variables that describe the jet kinematics, charged and neutral PF candidates features (the first 25 ordered by the displacement significance), and secondary vertices features (up to 4 secondary vertices sorted by the flight distance significance).

The PF candidate features and the secondary vertices features are fed as input to three different convolutional neural networks (CNN) [?] and, afterward, to three different recurrent neural networks (RNN) composed of LSTM layers [?] that treat the constituents of the jet as a sequence. The three outputs and the global features are then fed to a dense feed-forward neural network that can classify the jet as b-tagged, c-tagged, or gluon/light-tagged, assigning to each class a score.

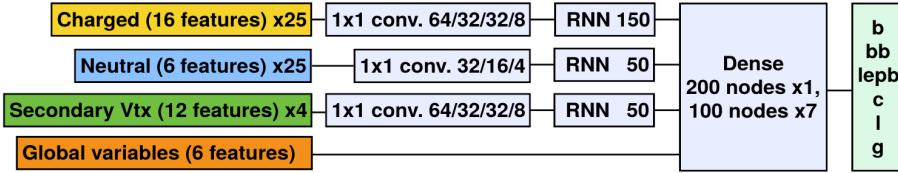


Figure 4.3: Illustration of the DEEPJET architecture. [? ]

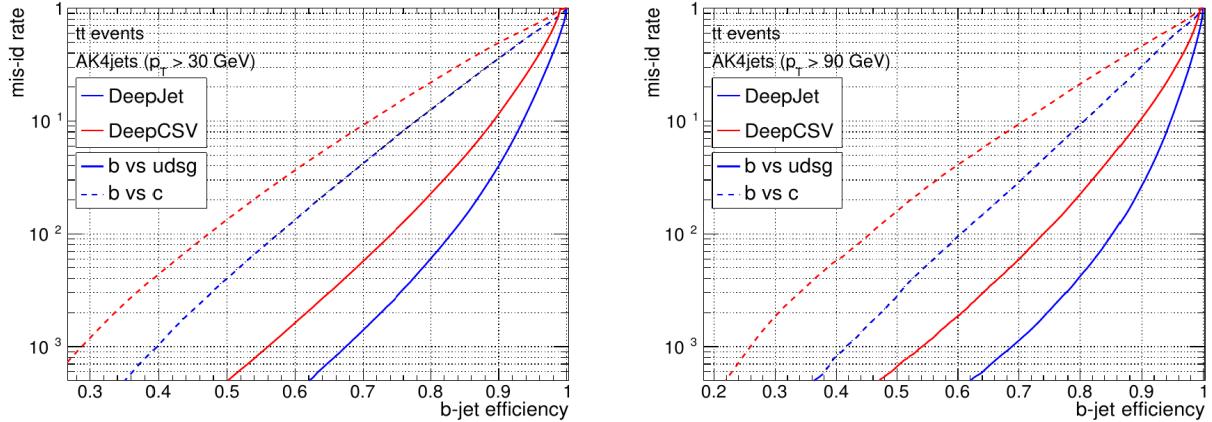


Figure 4.4: Performance comparison between DEEPJET and DEEPCSV, another model based on a simple feed-forward network. [? ]

## 4.2 Event simulation

To optimize the analysis, estimate the efficiencies, evaluate the systematic uncertainties, and interpret the data, we need Monte Carlo simulations (MC): labeled data that reproduces the real data as faithfully as possible.

The first stage of the simulation is the generation of the matrix elements, usually done with software like MADGRAPH [? ] or POWHEG [? ]: after the user has defined all the Feynman diagrams to compute and the physics model, the four momenta of all the particles are computed at the parton level. The next step, performed by PYTHIA [? ] in the CMS software, is the hadronization of colored particles, done exploiting the phenomenological Lund string model [? ].

The second step, performed by the GEANT4 framework [? ], is the simulation of the detector response and is the most computationally intensive step, to the point that other approaches to the simulation are being considered [? ? ].

GEANT emulates the propagation of particles through the CMS detector, simulating the particle-matter interactions and the readout of each subdetector. In these steps, the electronic noise, the digitization, the Landau fluctuations, and all the inefficiencies of the detectors like the non-uniformity of the materials are simulated.

The events are then reconstructed exactly like the real data, with the same algorithms. Despite the complexity of the full simulation, the MC can't reproduce exactly the real data so, for each analysis, the user has to apply the so-called scale factors (SF) to the

MC, a set of event weights that represents the ratios of efficiencies between MC and data  $SF = \epsilon_{data}/\epsilon_{MC}$ .

# *Chapter 5*

---

## Analysis methods

---

### 5.1 Neural networks

In this work, neural networks are used as a tool to classify different events in signal or background, giving to the input of the NN different features of the events (like the 4-vectors of the reconstructed particles in the event).

Therefore, in this section, the discussion will be focused on the classification problem.

A neural network (NN) is a machine learning <sup>1</sup> model that can solve a wide range of problems and make predictions, learning from the data.

The learning of an NN can be supervised, if the target of our dataset is known, or, vice versa, unsupervised.

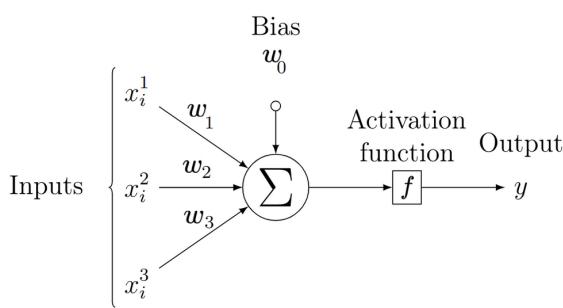
In the context of supervised learning, there are mainly two classes of problems that a neural network can address:

- **Regression:** Given a set of inputs  $\mathbf{x} \in \mathbb{R}^N$ , predict a set of output  $\mathbf{y} \in \mathbb{R}^M$
- **Classification:** Given a set of inputs  $\mathbf{x} \in \mathbb{R}^N$ , predict a set of output  $\mathbf{y} \in \mathbb{Z}_M$

---

<sup>1</sup>Machine learning is the process by which a computer program uses data to infer and build a predictive model.

## Perceptron



**Figure 5.1:** Visualization of a perceptron with three dimensional inputs  $x_{ij}$ . Each feature input is multiplied by a weight and everything is summed, along with the bias  $w_0$ . The result is passed to the activation function  $f$ , which, in this case, is the Heaviside theta function

The simplest NN is the Rosenblatt perceptron [? ] (Fig. ??), an NN with one single neuron that addresses the binary classification problem.

The characteristics of the neuron are:

- The dimension of the input. Each data entry is N-dimensional and each dimension is called a "feature".
- Each neuron has N weights, plus one bias, that are tuned through the learning procedure, looking at the data.
- The activation function, i.e. the function is applied do the dot product  $\mathbf{x} \cdot \mathbf{w} + w_0$ . In the case of the perceptron,  $f$  is the Heaviside theta function.

Given a set of inputs  $\mathbf{x} \in \mathbb{R}^N$  and a set of weights  $\mathbf{w} \in \mathbb{R}^{N+1}$ , the output will be

$$y = \begin{cases} 1 & \text{if } \sum_{i=1}^N w_i x_i + w_0 > 0 \\ 0 & \text{Otherwise} \end{cases} \quad (5.1)$$

The weights  $w_i$  are obtained through the learning procedure.

Given  $\hat{y}_i$  the truth label of the input  $\mathbf{x}_i$ , and  $x_{ji}$  the i-th features of the j-th data point:

---

### Algorithm Perceptron learning

---

```

1: Initialize the weights randomly
2: for T times do
3:   for  $i = 0$  to  $\text{len}(\mathbf{x})$  do
4:     Compute  $y_i(\mathbf{x}_i) = f(\mathbf{w} \cdot \mathbf{x}_i + w_0)$ 
5:     Update the weights  $w_j = w_j + r(\hat{y}_i - y_i)x_{ji}$ 
6:   end for
7: end for
```

---

Each iteration is called an "epoch". The number of epochs T and the so-called "learning parameter"  $r$  are the first two examples of hyperparameters that are chosen by the user.

The perceptron is a linear classifier. It means that it can not solve non-linearly separable problems, like an XOR gate, and can not address multiclass classification problems.

**Feedforward neural networks** Neurons can be connected in layers to build a feed-forward neural network in which each layer is fully connected to the neurons of the previous

one. So, the output of the  $j$ -th neuron in the  $k$ -th layer is

$$y_j^k = f^k \left( \sum_i w_{ji}^k y_i^{k-1} + w_{j0}^k \right) \quad (5.2)$$

where  $f$  is the so-called activation function of the neuron.

The universal approximation theorem [?], asserts that a neural network with one single hidden layer and enough number of neurons can approximate any function.

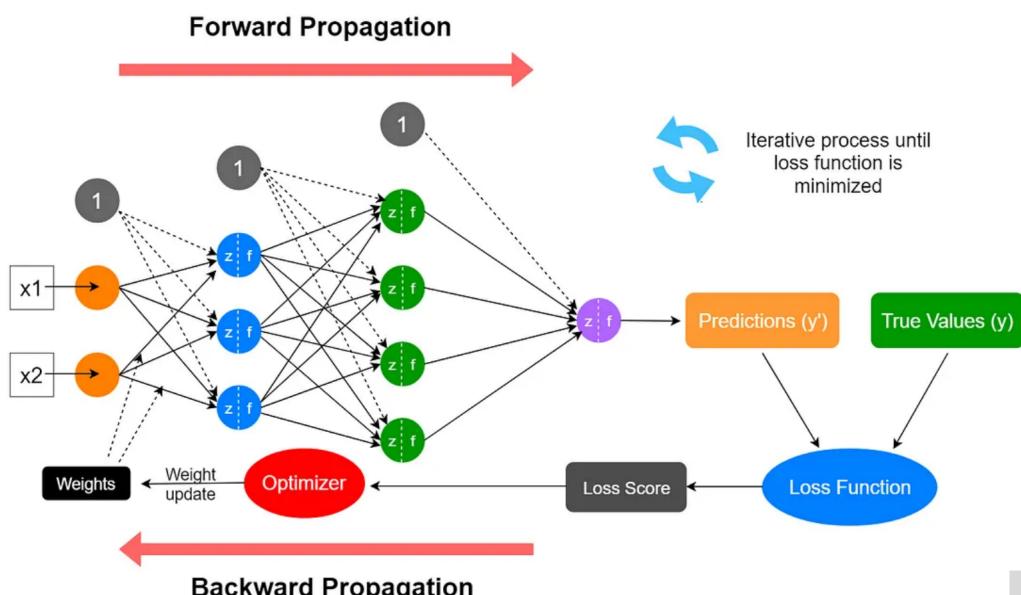
Despite that, it could be needed an arbitrarily large number of neurons to approximate one function using a single layer, so, to simplify the learning process, more layers can be added.

A neural network can work in two regimes: evaluation and training.

The training regime is the phase in which each weight of the network is tuned learning from the data, and is divided into different steps that will be discussed in detail in the next paragraphs:

1. Architecture definition: the number of neurons, layers, and activation functions have to be defined.
2. Weight initialization: the weights are initialized randomly.
3. Evaluation of the output: the evaluation phase in which the NN makes the predictions.
4. Loss evaluation: The loss-function value is computed.
5. Backpropagation: The gradients of the loss with respect to each weight are computed and the weights are updated.
6. Iterate from point 3.  $T$  times.

After that the NN is trained, it can be used in the evaluation mode, feeding the data to it and collecting the output to obtain the predictions.



**Figure 5.2:** Sketch of the training process of a binary classifier NN with 2 input features.

### 5.1.1 Loss function

Under the hood, a neural network is just an optimizer that tries to find the optimal function that minimizes a specific metric.

In other words, it is just a tool to perform a multidimensional fit, specifying the objective function in terms of sum and composition of activation function <sup>2</sup>.

The metric that the NN tries to minimize during the learning phase is the so-called "Loss function", a continuous function that compares the prediction of the NN with the labels and has its minimum, ideally, when all the data are predicted correctly.

For example, in the case of a regression problem, one of the most used loss functions is the MSE  $L = \sum(y_i - \hat{y}_i)^2$  <sup>3</sup>.

In the context of classification problems, the most used loss function is the cross-entropy loss.

Assuming that each class belongs to one output neuron, we can use the so-called one-hot encoding, which is a binary representation method where a categorical variable is transformed into a binary vector with a 1 indicating the presence of a particular category and zeros elsewhere <sup>4</sup>.

The output of the NN will be a vector as long as the total number of classes, and such that the sum of its elements is equal to 1. This can be achieved using the softmax activation function in the output layer (see the paragraph on activation functions).

The one-hot encoding representation is used, instead of the integer encoding that just indexes the classes, to avoid the assumption of natural ordering and to avoid predictions halfway between categories.

In the one-hot encoding representation, the cross-entropy loss is

$$\mathcal{L} = - \sum_i^N w_{\hat{y}_i} \hat{\mathbf{y}}_i \cdot \log \mathbf{y}_i \quad (5.3)$$

where  $\hat{\mathbf{y}}_i$  and  $\mathbf{y}_i$  are respectively the i-th label and prediction, both in the one-hot encoding representation, while  $w_{\hat{y}_i}$  is a weight related to each class that is used to compensate the class imbalance in the training dataset, avoiding the NN to be biased towards one specific class.

The cross-entropy loss is related to the likelihood of the multinomial distribution. Assuming m classes, the negative log-likelihood of the multinomial distribution, for a single data point, is

$$-\log \mathcal{L}_{MN_i}(\mathbf{p}_i; \mathbf{n}_i) = - \sum_j^m n_j \log p_j + \text{const.} \quad (5.4)$$

<sup>2</sup>This is called "induction bias" and is determined by the architecture of the NN

<sup>3</sup>which is the same metric that is minimized when a least squares fit is performed

<sup>4</sup>For example, if we have 3 classes, the 2nd class is represented with the vector [0,1,0]

where  $n_j$  is the number of predicted entries in the  $j$ -th class, and  $p_j$  is the probability that the data belongs to the  $j$ -th class.

In the one-hot encoding representation  $n_j = \delta_{j\hat{y}_i}$  and  $p_j$  is the score  $y_{\hat{y}_i}$ , where  $\hat{y}_i$  is the  $i$ -th label in the integer representation.

$$\mathcal{L}_{MN_i} = -\hat{y}_i \cdot \log y_i \quad (5.5)$$

Since each data point is independent from the others, the total log-likelihood is the sum of the single log-likelihoods

$$\mathcal{L}_{MN} = - \sum_i \hat{y}_i \cdot \log y_i \quad (5.6)$$

## Optimizers and backpropagation

To minimize the loss function, neural networks exploit the technique of gradient descent using different algorithms called "optimizers".

The simplest optimizer is the *stochastic gradient descent* (SGD) that, for each data point, computes the gradient of the loss function and updates the weight in the opposite direction of the gradient, propagating it along the network in a procedure called *back-propagation*<sup>5</sup>.

---

### Algorithm Stochastic gradient descent

---

```

1: for i=0 to T do
2:   for j=0 to len( $\mathbf{x}$ ) do
3:      $\mathbf{w} = \mathbf{w} - \eta \nabla_{\mathbf{w}} \mathcal{L}(y_j(\mathbf{w}), \hat{y}_j)$ 
4:   end for
5: end for
```

---

where  $\eta$  is the *learning rate* hyper-parameter.

The learning rate should be tuned carefully because a small learning rate will make the NN slow to reach the minimum of the loss, while a huge learning rate will make the gradient descent procedure very unstable, jumping randomly across the loss hypersurface.

To mitigate the fluctuation of the SGD and to improve the computational efficiency, we can compute the gradient on a batch of data and then update the weights

---

### Algorithm Mini-batch stochastic gradient descent

---

```

1: for i=0 to T do
2:   for k=0 to  $n_{batch}$  do
3:      $\mathbf{w} = \mathbf{w} - \eta \nabla_{\mathbf{w}} \sum_j^{batch\ size} \mathcal{L}(y_j^k(\mathbf{w}), \hat{y}_j^k)$ 
4:   end for
5: end for
```

---

The gradient is computed and back-propagated using the *automatic differentiation*. In principle, when the NN is evaluated, a computational graph is built. The nodes of the

---

<sup>5</sup>That is just a fancy name to call a composite derivative.

graphs are just the operators applied to the tensors <sup>6</sup> and, in the backpropagation phase, the gradients of each node are computed and propagated all the way along the graph using the chain rule.

Furthermore, NNs perform a lot of simple arithmetic operations that can be performed simultaneously in parallel, so they can be accelerated using graphing processing units (GPU).

The list of optimizers is long and each of them has its pros and cons. The SGD has a slower convergence compared to the other optimizers due to its stochastic nature, it is really sensitive to the learning rate, and, especially in high-dimensional and nonconvex problems, it gets stuck very easily in local minima.

To address these problems, one of the most used optimizers is ADAM (ADAptive Moment estimator) [? ]. It is able to adapt the learning rate of each parameter individually computing the moving average of gradients and square gradients, ensuring a fast convergence.

The drawback of ADAM is that it has some convergence issues due to its large variance in the first epochs of training in the adaptive learning rate, showing huge fluctuations in the loss, and being too sensitive to the choice of hyperparameters.

Trying to stem these issues, recently, a new variant of Adam was proposed: RAdam (Rectified Adam) [? ].

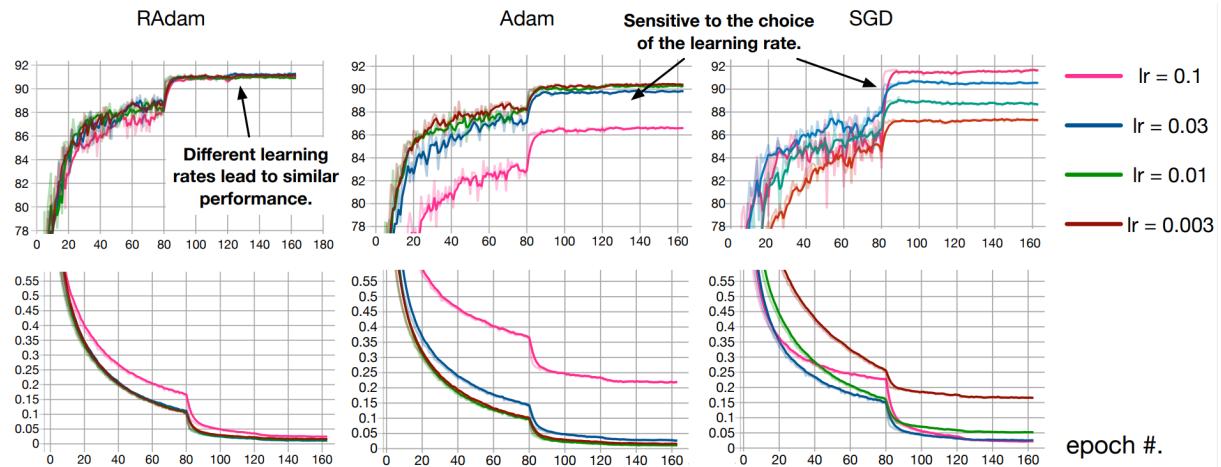
RAdam tries to tackle Adam's convergence issues by introducing a term to rectify the variance of the adaptive learning rate, trying to keep it small in the first epochs.

<b>Algorithm</b>	Radam Optimizer
<b>Initialize:</b>	$\theta_0, m_0 \leftarrow 0, v_0 \leftarrow 0$
<b>For</b> $\theta_t = [1, \dots, T]$ <b>do</b>	<b>Hyperparameters:</b> $\alpha, \beta_1, \beta_2, T$
$g_t \leftarrow \nabla_{\theta} f_t(\theta_{t-1})$	▷ Gradient Computation
$m_t \leftarrow \beta_1 m_{t-1} + (1 - \beta_1) g_t$	▷ First Order Moment
$v_t \leftarrow \beta_2 v_{t-1} + (1 - \beta_2) g_t^2$	▷ Second Order Moment
$\rho_\infty \leftarrow 2/(1 - \beta_2) - 1$	
$\rho_t = \rho_\infty - 2t\beta_2^t/(1 - \beta_2)$	
<b>If</b> $\rho_t \geq 5$	▷ Check if the variance is tractable
$\rho_u = (\rho_t - 4) \times (\rho_t - 2) \times \rho_\infty$	
$\rho_d = (\rho_\infty - 4) \times (\rho_\infty - 2) \times \rho_t$	
$\rho = \sqrt{(1 - \beta_2) \times \rho_u / \rho_d}$	▷ Variance rectification term
$\alpha_1 = \rho \times \alpha / (1 - \beta_1^t)$	▷ Rectified learning rate
$\theta_t \leftarrow \theta_{t-1} - \alpha_1 m_t / (\sqrt{v_t} + \epsilon)$	▷ Update parameters with rectification
<b>Else</b>	
$\alpha_2 = \alpha / (1 - \beta_1^t)$	▷ Bias correction
$\theta_t \leftarrow \theta_{t-1} - \alpha_2 m_t$	▷ Update parameters without rectification
<b>return</b> $\theta_T$	

As the SGD, all the other optimizers can work on minibatch, evaluating the gradients on a group of data points.

---

<sup>6</sup>Tensors are just matrices that contain the input data, the weights, and all the intermediate representations of the network.



**Figure 5.3:** Evolution in the epochs of the accuracy (top) and loss (bottom) with RAdam, Adam, and SGD with different learning rates on the CIFAR10 problem [? ? ].

### 5.1.2 Architecture and complexity

The architecture of a neural network, specifically the number of layers and neurons, plays a crucial role in determining how well the network can learn and generalize from data. A neural network can be shallow (few layers with many neurons) or deep (many layers with few neurons in each layer).

Deep networks tend to capture more complex patterns but suffer the vanishing gradient problem (the magnitude of the gradient decreases propagating back in the network).

A high number of neurons improves the capacity to represent functions and makes the network more robust to the noise, embedding the data in a high-dimensional space.

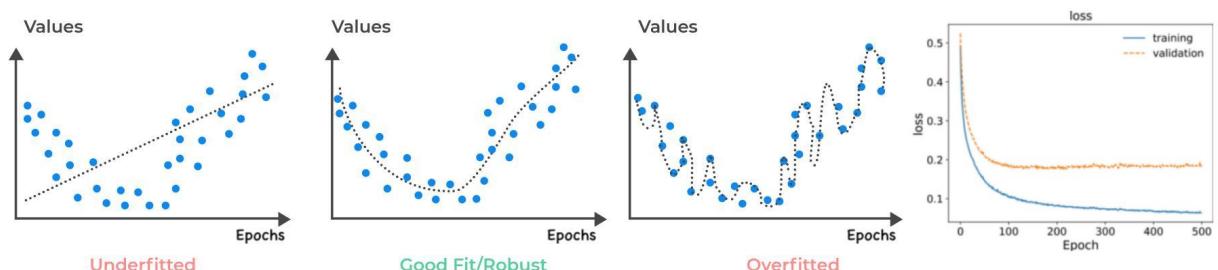
Complex NNs are more prone to the *overfitting*. Overfitting is the behavior of an NN that occurs when the model performs well on the training dataset but doesn't perform in the same way on other data.

This occurs because the NN starts to learn the noise and the fluctuations of the training dataset, worsening its generalization capabilities and losing its predictive power.

During the training of an NN, overfitting is the most important thing to take under control, and, to do it, the loss is evaluated, for each epoch, not only for the data on which the network is training, but also for other data, called the "validation dataset".

In the case of overfitting, we will see the validation loss that will start to increase, separating from the training loss.

On the other hand, if the network is too simple, we will fall into the underfitting, i.e. the objective function represented by the NN is too simple to fit the data.



**Figure 5.4:** Representation of underfitting, of a good fit, and of overfitting. The right plot shows the behavior of the validation loss in case of overfitting

Several techniques can be used to prevent the overfitting:

- Diminish the number of neurons and the complexity of the network
- Increase the size of the training dataset
- Add a penalty term to the loss function, like the L2 regularization

$$\mathcal{L}' = \mathcal{L} + \lambda \sum \|\mathbf{w}\|^2 \quad (5.7)$$

a term that penalizes scenarios in which the weights assume large values (that usually is one consequence of the overfitting)

- **Early stopping:** stop the learning when the validation loss starts to separate from the training loss.
- **Dropout:** the dropout [? ] is the probability that, in a single epoch during the training, one neuron will just be ignored and its connection with the rest of the network will be dropped out.

By randomly disabling neurons, dropout prevents any single neuron from becoming overly specialized on a specific feature of the training data.

Furthermore, introducing this stochastic behavior can also accelerate the training, preventing the network from getting stuck in local minima of the loss.

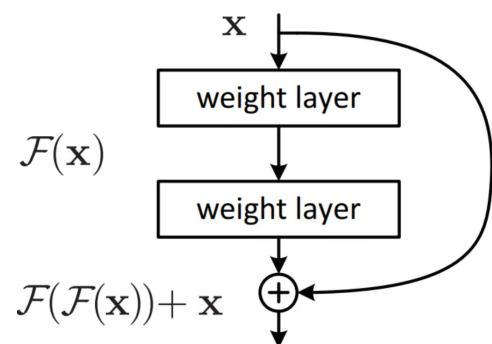
The dropout rate is another hyperparameter that should be tuned carefully by the user.

## Residuals

To mitigate the effect of the vanishing gradient, some "skip connections" can be used to back-propagate the gradients better through the network, ensuring that the output of a layer is added together with the output of a later layer.

If the dimensionality at the beginning of the connection is different from the dimensionality at the end of the connection, the residual can be multiplied by a linear projection to align the dimensions.

These networks are called "residual networks" or ResNets [? ].



**Figure 5.5:** Visualization of the skip connection in a residual block of a ResNet.

## Activation functions

The activation functions in a neural network must possess some properties:

- Continuity and differentiability: to compute the gradients of the loss with respect to the weights.
- Non-linearity: the activation functions must be non-linear because the composition of linear functions is still a linear function

- Their derivative must be bounded to avoid the so-called "explosive gradient", the opposite phenomena of the vanishing gradient.

The activation functions of the output layer are different from the others.

In a regression problem, the activation function of the output layer is linear, while, in the case of classification, using the one-hot encoding, the used activation function is the *softmax* function

$$\text{softmax}(\mathbf{x})_j = \frac{e^{x_j}}{\sum_{k=1}^K e^{x_k}} \quad \text{for } j = 1, \dots, K \quad (5.8)$$

where  $K$  is the number of classes.

The softmax function ensures that the output of the network is a vector of length  $K$ , such that the sum of its elements is always 1. This allows us to treat the output of the network as a probability.

In the hidden layers, instead, we have more choices. The first activation functions ever used were the logistic function and the hyperbolic tangent, which are related to each other, and have a finite codomain.

These two functions were chosen due to the interpretation that we can give to them, like in the perceptron <sup>7</sup>: when the function is around its lower bound, the neuron is off, and when the function is around its upper bound, the neuron is on.

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} = 2\sigma(2x) - 1 \quad (5.9)$$

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (5.10)$$

But the derivative of these functions is significantly different from zero just in a narrow range, slowing the learning of the network once the function saturates.

Another activation function that was proposed was the ReLU (Rectified linear unit)[?], one of the most used activation function

$$\text{ReLU}(x) = \max(0, x) \quad (5.11)$$

The ReLU allows a better propagation of the gradients and is very efficient from the computational point of view since is a piecewise linear function and the gradient is a constant.

Furthermore, the ReLU introduces a sparse representation in the network: when the network is initialized, only about half of the neurons have a non-zero output, and this has a regularization effect similar to the dropout.

Despite that, the ReLU suffers the so-called "dying ReLU problem": some units can be pushed into some states in which they become inactive to all possible inputs, making an always zero gradient. In this state, the neuron enters an endless dead state.

To solve the dying ReLU problem, a lot of ReLU-like functions were proposed. In

---

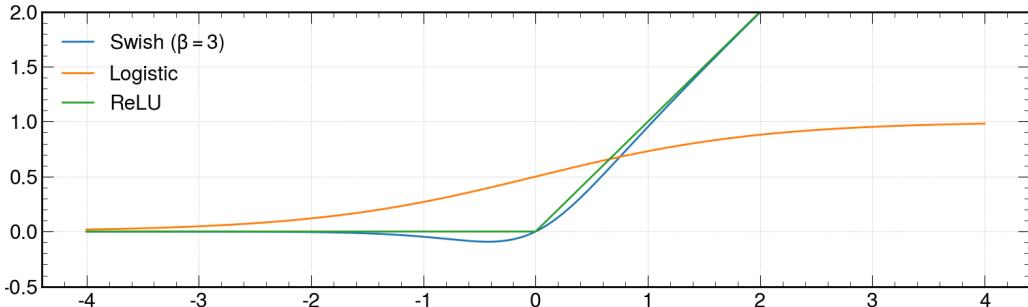
<sup>7</sup>in fact the logistic function can be seen as the "smoothed version" of the Heaviside theta function

this work, the used activation function was the recently proposed Swish [? ].

$$\text{Swish}(x; \beta) = \frac{x}{1 + e^{-\beta x}} \quad (5.12)$$

where  $\beta$  is a learnable parameter.

The Swish function has a smooth gradient that can avoid the dying ReLU problem, ensuring the neurons continue to produce non-zero outputs.

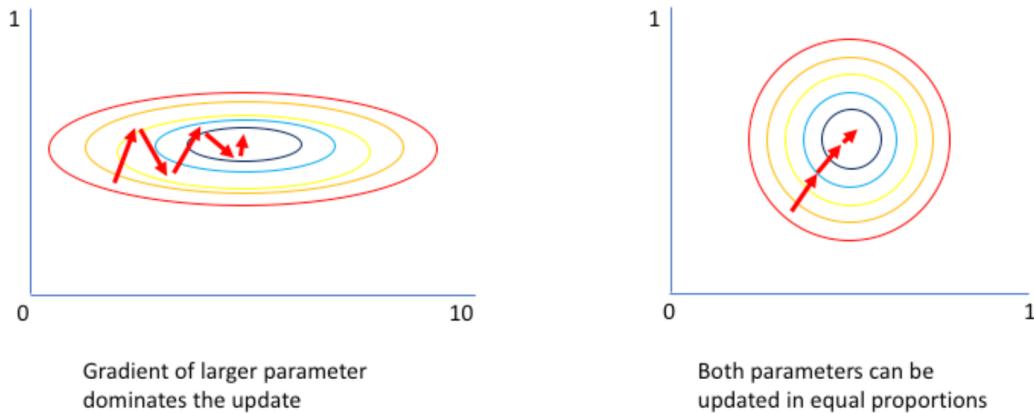


**Figure 5.6:** Logistic, ReLU, and Swish activation functions.

Anyway, despite all the motivations that were provided, the best activation function does not exist and, in general, in the field of neural networks, every choice is justifiable only through heuristics.

## Input normalization

The features of the input data have different distributions with different ranges and scales. The parameters associated with each feature will also have different scales and this can lead to a loss function whose gradient will be biased toward certain parameters, as shown in Fig. ??.



**Figure 5.7:** Loss curve levels in case of non-normalized inputs (left) and normalized inputs (right) [? ].

To make the learning of the network more stable and faster, we can normalize all inputs to a standard scale, allowing us to use larger learning rates.

Furthermore, this also avoids the saturation of a large number of activation functions in the first epochs which can lead to a slow learning due to the small gradients in the

saturation regime.

For specific activation functions, like the ReLU-like, having inputs with zero mean will cause  $\sim 50\%$  of the neurons to be inactive, creating a sparse representation (see paragraph on activation functions).

One of the most used input normalization techniques is the *Batch norm* [?], which consists of normalizing each input feature with its mean and standard deviation computed across all the batch

$$\mathbf{x}'_i = \frac{\mathbf{x}_i - \mu_i}{\sigma_i + \epsilon} \quad (5.13)$$

where  $\mathbf{x}_i$  is the i-th feature of the input tensor and  $\epsilon$  a small number added to prevent numerical issues.

Despite that, the mean and the standard deviations are not robust estimators, and, if a feature has significant outliers, the batch norm can create distributions not centered around zero.

To overcome this issue, we can replace the mean and the standard deviation with more robust estimators like the median, we can just clip the outlier to a threshold value, or we can apply a function that will shrink them.

In this work, the applied normalization to positive value features that have significant outliers is the following

$$\mathbf{x}'_i = \frac{\log(1 + \mathbf{x}_i) - \mu_i^{\log}}{\sigma_i^{\log} + \epsilon} \quad (5.14)$$

where  $\mu_i^{\log}$  and  $\sigma_i^{\log}$  are the mean and the standard deviation computed on  $\log(1 + \mathbf{x}_i)$

**Internal normalization** During the training, the distribution of each layer's weights changes, as the parameters of the previous layers change. This slows down the training by requiring lower learning rates.

This behavior is called "internal covariate shift", and we can address the problem by introducing a normalizing layer in the network.

We could use the batch norm or, instead, the layer norm [?], which is the same thing as the batch norm but the mean and the standard deviation are computed across the features instead of across the batch.

In some architectures (like RNN) or domains (natural language processing), the layer norm was proven to work better than the batch norm.

## Weight initialization

The initialization of the weights is strictly related to the chosen normalization and the chosen activation functions.

Large or small initialization values will lead respectively to the problem of exploding and vanishing gradient.

An appropriate initialization should impose a zero mean of the weights and a variance that should stay constant across all the layers.

To do it, if we use activation functions that are linear around zero (like the tanh), we can use the Xavier initialization [? ].

$$W \sim U \left[ -\frac{\sqrt{6}}{\sqrt{n_j + n_{j+1}}}, \frac{\sqrt{6}}{\sqrt{n_j + n_{j+1}}} \right] \quad (5.15)$$

where  $n_j$  is the number of units in the j-th layer.

If we use ReLU-like functions, instead, we should use the He initialization [? ].

$$W \sim \mathcal{N} \left( 0, \frac{2}{n^l} \right) \quad (5.16)$$

the derivation of these two distributions is simple and is provided in the respective papers, and consists only of imposing the variance of the output of a unit to be constant across all layers.

### Hyperparameter optimization

In this section, we have introduced many hyperparameters like the learning rate, the size of the network, etc. These parameters have to be tuned to obtain the best performances.

The simplest way to perform a hyperparameters search is to perform a *grid search*, which consists of creating a grid of parameters and repeating the training for each combination of them.

However, this simple algorithm is feasible only for small networks because the number of trainings required depends on a combinatorial factor.

Clever ways to find the optimal set of hyperparameters can involve Bayesian optimization methods, but, if the model is too big and the time and the computing resources are very scarce, the only remaining tool is manual optimization in which the user tries to find a good set of hyperparameters by trying with just a couple of trainings.

#### 5.1.3 Attention mechanism

Attention is a mechanism that enables models to focus on specific parts or elements of the input data while performing a task. It is particularly important in fields like natural language processing (NLP) and computer vision.

Attention networks belong to the class of sequence-to-sequence (seq2seq) models. One of the most common seq2seq tasks is the language translation in which the model takes as input one sentence (i.e. a sequence) in a language and translates it into another language (i.e. another sequence). This means that the inputs of the network are 3 dimensional: one is the batch dimension, one is the sequence dimension, and the last is the feature dimension.

In particle physics, seq2seq models enable us to treat each event as a sequence composed of its particles and to maintain the event structure.

## Scaled dot product

One particular method to implement the attention is the "scaled dot product" [? ].

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (5.17)$$

The scaled dot product has 3 inputs [? ]: the "Query" (Q) of dimensions  $(N, S_2, d_k)$  and "Key" (K) of dimension  $(N, S_1, d_k)$ , and the "Value" (V) of dimension  $(N, S_2, d_k)$ , where  $N$  is the batch size,  $S_1, S_2$  the length of the sequences and  $d_k$  the number of neurons of the previous layer in the network.

$Q, K, V$  are linear embeddings of the layer inputs, so  $Q, K, V = W_{Q,K,V} \cdot X_{Q,K,V}$  where  $W$  are the weights learned during the training.

The  $\text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)$  is the matrix of the "Attention-scores" and has dimension  $(N, S_2, S_1)$ , so it exploits the dot product between the Query and Key tensors to compute and assign a score to all the pairs in the two sequences, performing a weighted average on the Value tensor.

The inputs of the scaled dot product are called Key, Query, and Value as an analogy to the search engine systems: a search engine will compare the user Query with the Key associated with all the pages available in the database and will return the Value, i.e. the page, which associated key best match the Query.

Attention networks are a rare example of explainability. By studying attention scores we can understand what the network is learning. Usually, instead, neural networks are black boxes that tend to learn representations that are difficult to understand by humans.

In particle physics, this approach is particularly feasible for two reasons

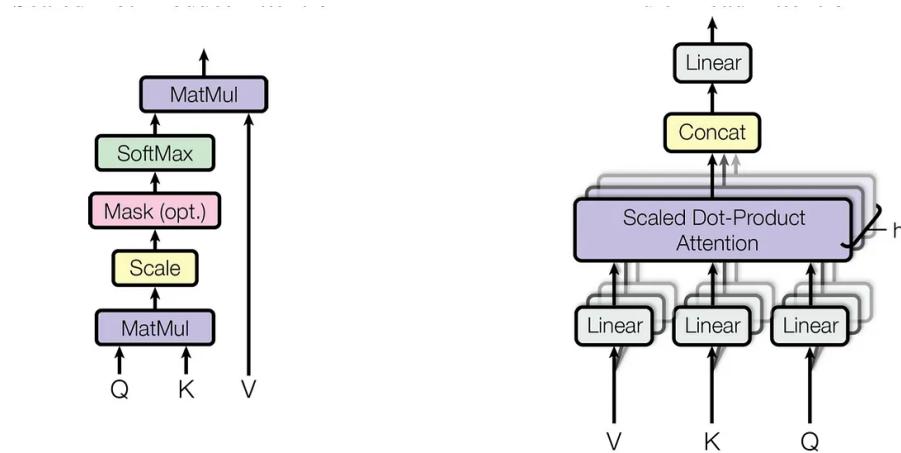
- There is a natural way to handle arrays of different lengths: we can just fix the maximum length of the sequence and, for the absent particles, just impose the respective attention scores to zero.  
This allows us to maintain the ragged structure of the event.
- Attention networks are invariant for permutation of the elements in the sequence, so we are not forced to introduce an ordering between the particles

**Multihead attention** One set of  $W_Q, W_K, W_V$  matrices is called an *attention head*, but we can stack multiple attention heads together, allowing the model to learn different representations of the sequence separately.

	I	love	you
je	0.94	0.02	0.04
t'	0.11	0.01	0.88
aime	0.03	0.95	0.02

**Figure 5.8:** Attention scores between sentences in English and French. The network has learned to recognize which words have the same meaning also if they are in a different order.

The different heads can be computed in parallel and their output, at the end, is concatenated.



**Figure 5.9:** (Left) scaled dot product, (right) multi-head attention [? ].

**Self-attention** With a "self-attention layer" we refer to a scaled dot product layer in which  $Q, K, V$  are linear embeddings of the same input  $\mathbf{X}$ .

It enables the model to capture contextual information and relationships between elements in the sequence, computing attention scores for each element with respect to all the others. These attention scores determine how much each element should focus on the others and the resulting weighted sum provides a contextually enriched representation of the sequence.

**Cross-attention** Cross-attention is an extension of the self-attention mechanism that allows a model to attend to elements in a target sequence based on the elements in another sequence (called context sequence).

The Key and Value vectors are linear embeddings of the context sequence while the Query is a linear embedding of the target sequence.

Cross-attention allows models to establish relationships between elements in different sequences, also of different lengths, making it suitable for tasks that involve aligning information from one sequence with another.

**Self-attention pooling** Pooling is a technique used to downsample or reduce the dimensionality of feature maps, or in our case a sequence, retaining important information. As already discussed, we are interested in classifying an event as a signal or background, so we have to reduce the dimensionality of the sequences that contain all the particles in the events, to just one.

Self-attention pooling [?] is a pooling layer inspired by self-attention

$$\text{Self-Att. Pool} = \text{Softmax}(\mathbf{W} \mathbf{X}^T) \mathbf{X} \quad (5.18)$$

where  $\mathbf{X}$  is the input tensor of size  $(N, S, F)$  and  $\mathbf{W}$  is a set of trainable weight of size  $(S', F)$ , where  $S'$  is the target size of the output sequence.

Therefore, self-attention pooling is a weighted average of the sequence features where the weights are the attention scores learned by the self-attention mechanism.

## 5.2 Signal extraction

### 5.2.1 Likelihood construction

The signal extraction procedure relies on a maximum likelihood fit of a binned distribution of one or more observables.

The likelihood is constructed as the product of Poisson distributions, one for each bin, that contains the parameter of interest (POI) to fit.

Given  $C$  channels<sup>8</sup>, each that contains  $P_c$  processes, represented with histograms with  $B_c$  bins, once we have the signal and the background MC predictions  $s_{cb}$ ,  $b_{cb}$ , and the observed bin content  $n_{cb}$ , the likelihood is [? ]

$$\mathcal{L}(\mu) = \prod_c^C \prod_b^{B_c} \frac{(\mu s_{cb} + b_{cb})^{n_{cb}}}{n_{cb}!} e^{-(\mu s_{cb} + b_{cb})} \quad (5.19)$$

where index  $c$  stands for "channel",  $p$  for "process",  $b$  for "bin". In this case, the POI is the signal strength  $\mu$ , which is estimated by finding the value that maximizes the likelihood. The optimization problem is solved using the MIGRAD algorithm [? ], implemented in the MINUIT2 library, that performs a line search in the direction of the gradient and updates the covariance matrix (inverse of the Hessian matrix) with each step.

To improve the numerical stability, the software implementation exploits the negative log-likelihood  $NLL = -\log \mathcal{L}$  instead of the likelihood in the form of the Eq. (??)

$$\hat{\mu} = \text{argmin}_\mu -\log \mathcal{L}(\mu)$$

The predictions for each process are obtained through MC simulations, that are rescaled to match the total number of expected observed events. If  $N_{cp}^{MC}$  is the total number of entries in the histogram relative to the process  $p$  in the channel  $c$ , the respective entries have to be rescaled by a factor  $w_{cp}$

$$w_{cp} = \frac{\mathcal{L}_{\mathcal{I}} \sigma_p \epsilon_{cp}}{N_{cp}} \quad (5.20)$$

where  $\mathcal{L}_{\mathcal{I}}$  is the integrated luminosity,  $\sigma_p$  the cross-section of the process  $p$  and  $\epsilon_{cp}$  the fraction of events of the process  $p$  that pass the preselection defined by the channel  $c$  that is estimated using MC events.

---

<sup>8</sup>A channel is defined by a set of cuts and selections that have the scope to isolate the signal and reject the background or viceversa

## Systematic uncertainties

To incorporate in the model all the relevant uncertainties (i.e. luminosity and rate uncertainties, uncertainties related to the detector resolution, etc.) we add to the likelihood the so-called nuisance parameters  $\vec{\theta}$ .

The likelihood becomes [? ]

$$\mathcal{L}(\mu, \vec{\theta}) = \prod_c^C \prod_b^{B_c} Poiss(n_{cb} | \lambda_{cb}(\mu, \vec{\theta})) \prod_k \pi_k(\theta_k) \quad (5.21)$$

where  $\pi_k$  is the prior probability distribution of the nuisance  $\theta_k$ .

The expectation value  $\lambda_{cb}$  of the Poisson distribution now depends on the nuisance parameters that can be classified into normalization nuisances, shape nuisances, and statistical nuisances.

$$\lambda_{cb}(\mu, \vec{\theta}) = \sum_p^{P_c} M_{cp}(\mu) N_{cp}(\vec{\theta}_n) y_{cpb}(\vec{\theta}_s) + E_{cb}(\vec{\theta}_{MC}) \quad (5.22)$$

The term  $M_{cp}(\vec{\mu})$  is the physical model that contains the POI to fit,  $N_{cp}(\vec{\theta}_n)$  contains the factors relative to normalization nuisances,  $y_{cpb}(\vec{\theta}_s)$  is the predicted content of each bin (that depends on shape nuisance parameters) and  $E_{cpb}(\vec{\theta}_{MC})$  represents bin per bin the statistical error of the Monte Carlo predictions.

In the absence of systematic uncertainties, we have  $N_{cp} = 1$  and  $E_{cb} = 0 \quad \forall c, p, b$  and  $M_{cp} = \mu$  if  $p$  is a signal process, otherwise  $M_{cp} = 1$ ; while  $y_{cpb}$  is just the predicted bin content for each process.

All the nuisance parameters  $\theta_k$  are defined in units of standard deviations <sup>9</sup>.

Given that we know the central values and the standard deviations of all the variations, following the maximum entropy principle [? ], we can use the normal distribution as a prior for all the nuisance parameters.

$$\pi_k(\theta_k) = \mathcal{N}(\theta_k | 0, 1) = \frac{1}{\sqrt{2\pi}} e^{-\theta_k^2/2} \quad (5.23)$$

- **Normalization nuisances** are multiplicative corrections that affect the normalization of one process (e.g. the cross-section uncertainties) or of all processes (e.g. the luminosity uncertainties).

This type of uncertainty does not change the shape of the histogram but changes the number of expected events.

In  $\lambda_{cb}$  [eq. ??], they are represented by the term

$$N_{cp}(\vec{\theta}_n) = k_{cp}^{\theta_k^{(n)}} \quad (5.24)$$

where  $k_{cp}$  the relative  $+1\sigma$  variation from the nominal event yield for the process  $p$  in the channel  $c$  estimated through theory calculation of external measurements, while  $\theta_k$  is the relative normalization nuisance parameter.

---

<sup>9</sup>it means that  $\theta = \pm 1$  correspond to a 68% CL interval.

- **Shape nuisances:** Shape-changing nuisances are produced by uncertainties that cause a different variation of event yields for each different bin.

The model parameters are varied of  $\pm 1\sigma$  around their central value, creating an "Up" and a "Down" variation.

The bin content  $y_{cpb}$  has to be modified by multiplying a continuous and differentiable function of  $\vec{\theta}_s$ , so, given  $\delta_b^\pm$  the differences between the  $\pm 1\sigma$  bin content variations of the bin and the nominal one, the up and down variations are interpolated with a spline in a procedure called "vertical morphing" [? ]

$$f_b(\theta_k^{(s)}) = \begin{cases} \frac{1}{2}((\delta_b^+ - \delta_b^-)\theta_k + \frac{1}{8}(\delta_b^+ + \delta_b^-)(3\theta_k^6 - 10\theta_k^4 + 15\theta_k^2)) & \text{for } |\theta_k| \leq 1 \\ \theta_k\delta^+ & \text{for } \theta_k > 1 \\ -\theta_k\delta^- & \text{for } \theta_k < -1 \end{cases} \quad (5.25)$$

The term  $y_{cpb}(\vec{\theta}_s)$  is then defined as

$$y_{cpb}(\theta_s) = \max \left( 0, y_{cpb} + \sum_k f_{cpb}(\theta_k^{(s)}) \right) \quad (5.26)$$

In this way, we can clip  $y_{cpb}$  to 0 preventing the height of the bin from being negative. The vertical morphing procedure is shown and described in detail in Fig. ??

- **MC statistical nuisances:** The MC events used to predict the expected signal and background yield in each bin are limited.

To model the statistical MC uncertainties, we could add a nuisance parameter per bin per process, following the Barlow-Beeston approach [? ].

But, if we have enough statistics in each bin, we can just use a nuisance parameter per bin that scales the sum of the processes yield, following the Barlow-Beeston lite approach [? ].

The term  $E_{cb}$  in  $\lambda_{bc}$  is the term related to the statistical uncertainties is

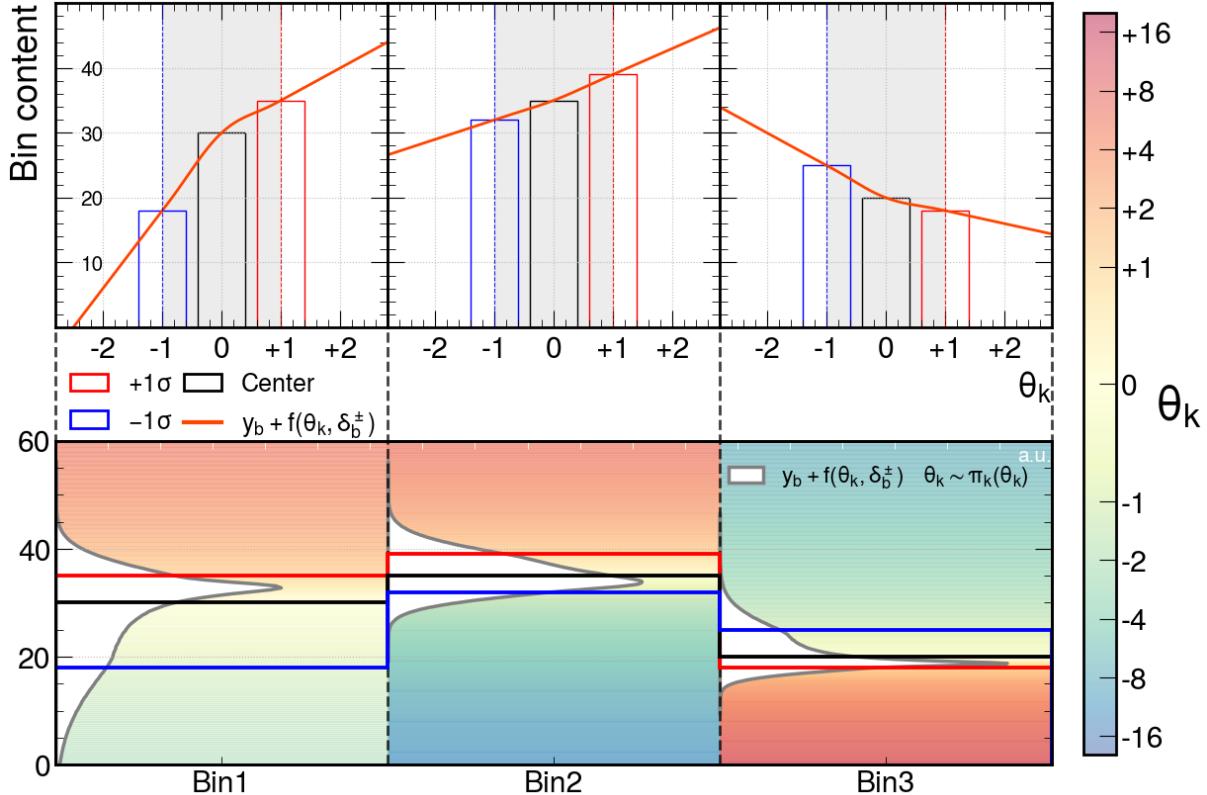
$$E_{cb}(\vec{\theta}_{MC}) = \theta_b \left( \sum_p \left( e_{cpb} N_{cp}(\vec{\theta}_n) M_{cp}(\mu) \right)^2 \right)^{1/2} \quad (5.27)$$

where  $N_{cp}$  and  $M_{cp}$  are the normalization factor and the physical model defined before.  $e_{cpb}$  is the Poissonian standard deviation of the b-th bin in the process p. Considering that the  $\hat{\sigma}_{Poiss} = \sqrt{y_b}$  and that the bin height has to be rescaled to match the expected number of observed events,  $e_{cpb}$  is

$$e_{cpb} = w_{cp} \sqrt{N_{cpb}^{MC}} = \frac{\mathcal{L}_I \sigma_p \epsilon_{cp}}{N_{cp}^{MC}} \sqrt{N_{cpb}^{MC}} \quad (5.28)$$

where  $N_{cpb}$  is the number of MC events in the b-th bin.

The term  $E_{cb}$  scales, linearly with the nuisance parameter, the event yield of the bin with the root square sum of the Poisson uncertainties of each process.



**Figure 5.10:** In the lower panel is represented an example of a histogram with 3 bins. The black line represents the histogram, the blue and the red lines represent the  $\pm 1\sigma$  shape variations. In the top panel, each bin is represented with its  $\pm 1\sigma$  variations and, on top of them, the  $y_b + f(\theta_k, \delta_b^\pm)$  interpolation is shown with a red line. In the background of the lower panel, a color map shows, for each value of the bin height, the corresponding value of the nuisance parameter  $\theta_k$ , while the rotated vertical white shadow represents, for each bin, the probability distribution in arbitrary units of the expectation value of the Poisson distribution  $\lambda_b = y_b + f(\theta_k, \delta_b^\pm)$  drawing  $\theta_k$  from the normal prior.

**Smoothing** For some systematics, up and down variations can lead to significant statistical fluctuations in the less populated bins due to low MC statistics, which, usually, are the high signal purity regions of the background histogram.

To overcome this inconvenience, a smoothing procedure is applied.

---

**Algorithm** Smoothing algorithm

---

- 1: Compute  $r_b^{up} = y_b^{up}/y_b$  and  $r_b^{down} = y_b^{down}/y_b$
  - 2: Compute  $\delta r_b = (r_b^{up} - r_b^{down})/2$
  - 3: Define  $\hat{r}_b^{up} = 1 + \delta r_b$  and  $\hat{r}_b^{down} = 1 + \delta r_b$
  - 4: Given  $x_b$  is the position of the bin centers, apply the LOWESS smoothing algorithm to the points  $(x_b, \hat{r}_b^{up})$  and  $(x_b, \hat{r}_b^{down})$ , finding the functions  $f^{up}(x)$  and  $f^{down}(x)$
  - 5: Redefine  $y_b^{up} = y_b f^{up}(x_b)$  and  $y_b^{down} = y_b f^{down}(x_b)$
- 

The LOWESS algorithm [? ] (Locally WEighted Scatterplot Smoothing) performs different linear regressions on a sliding window iteratively, weighting each point accordingly with its distance from the fitted line in each window.

## 5.2.2 Profile Likelihood

According to Wilks' theorem [?], the distribution of the likelihood ratio is asymptotically distributed like a  $\chi^2$ .

We have to fit just the POI  $\mu$ , so we can define the profile likelihood

$$-2\Delta \ln(\mathcal{L}) = -2 \ln \lambda(\mu) = -2 \ln \left( \frac{\mathcal{L}(\mu, \tilde{\boldsymbol{\theta}}(\mu))}{\mathcal{L}(\hat{\mu}, \hat{\boldsymbol{\theta}})} \right) \sim \chi_1^2(\mu) \quad (5.29)$$

$\hat{\mu}$  and  $\hat{\boldsymbol{\theta}}$  are the parameters that maximize the likelihood simultaneously, while  $\tilde{\boldsymbol{\theta}}(\mu)$  are the values of the nuisance parameters that maximize the likelihood for a fixed value of  $\mu$ . Since  $\int_0^1 \chi_1^2(x) dx = 0.683 = 1\sigma$ , to compute the 68% confidence level (CL) on  $\hat{\mu}$ , we can compute the profile likelihood and find the values of  $\mu$  for which holds  $-2\Delta \ln \mathcal{L}(\mu) = 1$ .

To understand how different groups of nuisances affect the profile likelihood and the total uncertainty on  $\mu$ , some of the  $\tilde{\boldsymbol{\theta}}(\mu)$  can be frozen in turn to the values  $\hat{\boldsymbol{\theta}}$ .

The so-called "breakdown procedure" consists of freezing sequentially groups of nuisance parameters, computing the profile likelihood and the relative uncertainties, and subtracting them in quadrature from the total uncertainty.

**Impact plots** Impact plots are a useful tool to understand the impact that each nuisance has on the total uncertainty. An example is shown in Fig. ??.

- In the left panel there are the names of the systematic sources, sorted according to their impact on the total uncertainty.
- In the center panel, there are two elements:

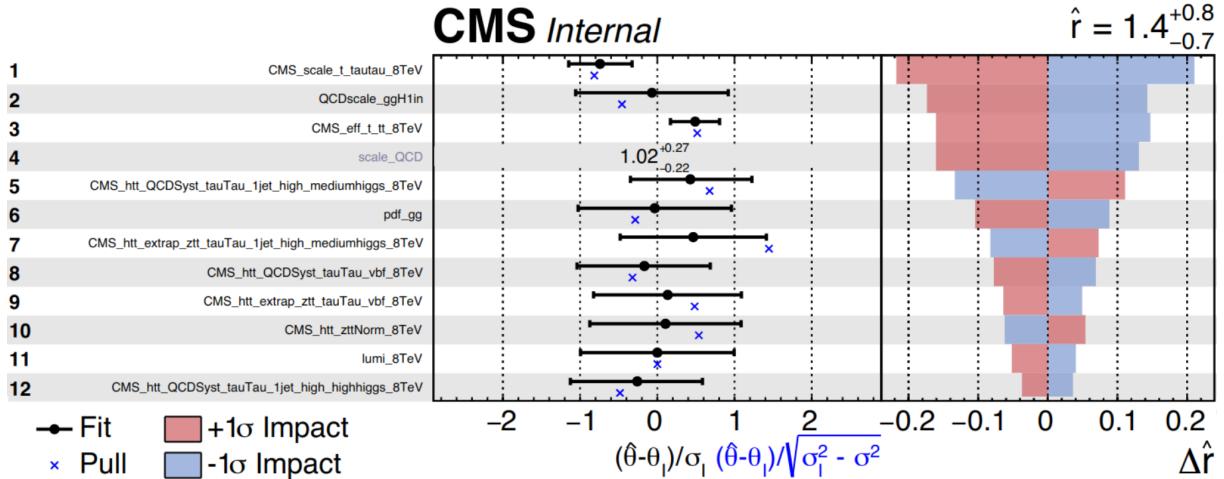
- The black bars ("Fit") represent the profile likelihood of the nuisance parameter. For each nuisance

$$\hat{\theta}_k = \underset{\theta_k}{\operatorname{argmin}} -2 \ln \left( \frac{\mathcal{L}(\theta_k, \tilde{\mu}(\theta_k), \tilde{\boldsymbol{\theta}}(\theta_k))}{\mathcal{L}(\hat{\mu}, \hat{\boldsymbol{\theta}}, \hat{\theta}_k)} \right) \quad (5.30)$$

is computed, along with the respective uncertainty, exploiting Wilks' theorem as explained before.

- The blue crosses ("Pulls") corresponds to  $(\hat{\theta} - \theta_{\text{pre-fit}})/\sqrt{\sigma_{\text{pre-fit}}^2 - \sigma_{\text{post-fit}}^2}$ . They tell us how much the fitted values are in tension with the pre-fit values defined by the priors.

- In the right panel there are the impacts. Once we have the fit for each nuisance, we can perform again the profile likelihood procedure on  $\mu$  but fixing each nuisance in turn to their  $\pm 1\sigma_{\text{post-fit}}$  value. The blue and the red bars represent the values  $\Delta^\pm \hat{\mu} = \hat{\mu}(\hat{\theta}) - \hat{\mu}(\hat{\theta}_{\pm 1\sigma})$ .



**Figure 5.11:** Example of an impact plot. In this plot, the POI  $\mu$  is called  $r$ , and  $\theta_I$ ,  $\sigma_I$  are the pre-fit values.

### 5.2.3 Asimov dataset and sensitivity

The term  $n_{cb}$  in Eq. (??) is the number of observed events in the b-th bin, so, it contains real data.

To perform a simulation of an analysis,  $n_{cb}$  can be replaced by MC events, creating the so-called "Asimov dataset" [?]

$$n_{cb}^{\text{Asimov}} = \sum_p y_{cpb} \quad (5.31)$$

In this way will be always  $\hat{\mu} = 1$  and  $\hat{\theta} = 0$ .

Using the Asimov dataset, the statistical uncertainties on  $\hat{\mu}$  can be estimated through an approximation.

1. Let's consider the Poisson log-likelihood for a single bin

$$\ln(\mathcal{L}_b(\mu)) = n_b \ln(\mu s_b + b_b) - (\mu s_b + b_b) - \ln n_b! \quad (5.32)$$

and let's find the maximum likelihood estimator for  $\mu$

$$\partial_\mu \ln \mathcal{L}_b = 0 \implies \hat{\mu} = \frac{n_b - b_b}{s_b} \quad (5.33)$$

The variance of the estimator  $\hat{\mu}_b$  is

$$\text{Var}(\hat{\mu}_b) = \frac{\text{Var}(n_b)}{s_b^2} = \frac{\mu s_b + b_b}{s_b^2} \quad (5.34)$$

The variance of  $\hat{\mu}_b$  depends on  $\mu$  itself, but, in the context of the Asimov dataset, we can approximate it imposing  $\mu = \hat{\mu}_b = 1$ , so the variance becomes

$$\text{Var}(\hat{\mu}_b) \sim \frac{s_b + b_b}{s_b^2} \quad (5.35)$$

2. The Cramer-Rao theorem [? ] states that, for a probability density function of the exponential family, to whom the Poisson distribution belongs, it holds

$$\text{Var}(\hat{\mu}) = -\frac{(1 - \partial_\mu b)^2}{E[\partial_\mu^2 \ln \mathcal{L}]} \quad (5.36)$$

where  $b$  is the bias of the estimator. The maximum likelihood estimator is asymptotically unbiased, so we can neglect  $b$ .

$$E[\partial_\mu^2 \ln \mathcal{L}_b] \sim -\frac{1}{\text{Var}(\hat{\mu}_b)} \quad (5.37)$$

and for the total likelihood

$$E[\partial_\mu^2 \ln \mathcal{L}] = \sum_b E[\partial_\mu^2 \ln \mathcal{L}_b] = \sum_b -\frac{1}{\text{Var}(\hat{\mu}_b)} \quad (5.38)$$

but since it holds  $\text{Var}(\hat{\mu}) = -1/E[\partial_\mu^2 \ln \mathcal{L}]$ , then

$$\text{Var}(\hat{\mu}) = \left[ \sum_b \frac{1}{\text{Var}(\hat{\mu}_b)} \right]^{-1} = \left[ \sum_b \frac{s_b^2}{s_b + b_b} \right]^{-1} \quad (5.39)$$

So, to estimate the uncertainty on  $\hat{\mu}$  using the Asimov dataset, we can compute the so-called **sensitivity**

$$Q = \sum_c \sum_b \frac{s_{cb}^2}{s_{cb} + b_{cb}} \quad (5.40)$$

and assert  $\sigma(\hat{\mu}) \sim 1/\sqrt{Q}$

Furthermore, the  $(x_b, Q_b)$  plot helps us to choose the binning of the histogram to optimize the signal sensitivity.

# *Chapter 6*

---

## Object and event selection

---

### 6.1 Event signatures

In this chapter, the  $t\bar{t} \rightarrow b\ell\nu bcb$  process will be denoted with "signal" or "cb events", while other processes will be referred to as "background".

The final state of a signal event consists of one muon or electron, three b jets, one c jet, and missing energy due to the neutrino from the leptonic W decay.

The invariant mass of one b jet and the c jet should be consistent with the mass of the W boson, and, adding another b jet, the invariant mass of the three jets should be consistent with the mass of the top quark.

The signal's final state can be mimicked by the following backgrounds.

- Semileptonic  $t\bar{t}$  ( $t\bar{t} \rightarrow b\ell\nu bq\bar{q}$ ): if one or more jets are mistagged or there are additional heavy flavor jets.  
Both the production of additional heavy-flavor jets and the flavor mistagging are suppressed but have to be evaluated.
- Dileptonic  $t\bar{t}$  ( $t\bar{t} \rightarrow b\bar{\ell}\nu \bar{b}\ell\bar{\nu}$ ): the presence of two leptons in the final states increase the trigger probability, while additional jets can be misidentified as c and b jets originating from the hadronic decay of the W boson.
- Dihadronic  $t\bar{t}$  ( $t\bar{t} \rightarrow b\bar{q}q \bar{b}q\bar{q}$ ): the lepton that would trigger the event can originate from the semileptonic decay of a meson in the jet if it passes the isolation selections, and the missing b jet can be obtained by mistagging or by additional heavy flavor jets.
- W+jets ( $W \rightarrow \ell\nu$ ): This process is the one with the largest cross-section but the requirement of three additional b jets and one additional c jet suppresses it significantly.
- WW+jets: the semileptonic decay of WW pairs produced directly by  $pp$  collision have required to be produced in association with two additional b jets to present the same signal signatures. Another difference with the signal is the inconsistency of the invariant mass of three jets with the top mass, which, however, is difficult to exploit due to the combinatorial ambiguities and the low tri-jet mass resolution.

- $tW + \text{jets}$ : it is similar to the  $WW + \text{jets}$  process, with the difference that only one additional b jet, instead of two, is required to match the signal signature.
- $tq$  (t-channel): if the top quark decays into a lepton, two additional b jets are required to match the signal signatures. Furthermore, if the other quark is not a charm quark, an additional c jet is required too. However, to produce a single top in a t channel, one incoming b quark is needed. Since the b quark production at LHC is mainly due to gluon splitting, the  $tq$  (t-channel) is often produced in association with an additional b jet.

The production cross-sections of these processes at  $\sqrt{s} = 13 \text{ TeV}$  with  $pp$  collisions are reported in Tab.??

$pp \rightarrow$	$t\bar{t}$				$W$	$WW$	$tW$	$tq$ (t-channel)
$\sigma(\text{pb})$ (13 TeV)	832				190000	118	79	214
	signal ( $b\ell\nu bcb$ )	semiLept ( $b\ell\nu b\bar{q}q$ )	diLept ( $b\ell\nu \bar{b}\ell\nu$ )	diHad ( $\bar{b}\bar{q}q b\bar{q}q$ )	Lept ( $\ell\nu$ )	semiLept ( $\ell\nu q\bar{q}$ )	semiLept ( $b\ell\nu q\bar{q}$ )	Lept ( $b\ell\nu q$ )
$\mathcal{BR}$	$3.7 \cdot 10^{-4}$	0.439	0.106	0.455	0.326	0.439	0.439	0.326
$\mathcal{BR} \cdot \sigma(pb)$	0.307	365	88	379	61000	51.8	34.7	69.8
$\mathcal{BR} \cdot \mathcal{L}_I \sigma$	$4.2 \cdot 10^4$	$5.0 \cdot 10^7$	$1.2 \cdot 10^7$	$5.2 \cdot 10^7$	$8.5 \cdot 10^9$	$7.1 \cdot 10^6$	$4.8 \cdot 10^6$	$9.6 \cdot 10^6$

**Table 6.1:** Production cross-sections of signal and backgrounds at  $\sqrt{s} = 13 \text{ TeV}$ . In the first row, there are the production cross-sections from  $pp$  collisions at 13 TeV, in the following rows there are the respective branching fractions of the different final states, the cross-section of the respective final states, and the total number of events expected in all the Run2, with a total integrated luminosity of  $\mathcal{L}_I = 138 \text{ fb}^{-1}$

**Analysis strategy** In the following sections, the selection criteria for physics objects and events are described, defining different signal regions for  $\ell = \mu, e$ . Therefore, after reconstructing the kinematics of the events, the signal extraction is performed through a template fit on a NN score, by exploiting multivariate analysis techniques, and also studying the relevant signatures of the events.

This work will be conducted only on MC simulations and the observed data will be replaced by the Asimov dataset.

## 6.2 Monte Carlo samples

The simulation of a hadron-hadron collision relies on different steps and elements [? ]:

1. **PDF set**: The first thing that has to be defined is the parton distribution function of the proton since it defines the initial partons involved in the hard scattering and their fraction of carrying momenta.

In all the simulated samples, the used PDF set is NNPDF3.1 NNLO [? ].

2. **Generator**: The kinematics of the outgoing particles are based on matrix-element calculated in perturbation theory, using Markov chain Monte Carlo integrators.

The generators that were used to build the MC samples are MADGRAPH 5,

aMC@NLO [? ], and POWHEG 2 [? ], setting  $\alpha_s$  to a value of 0.118 at the Z pole mass. All the samples are generated with a different number of additional partons, and with different precision, i.e. considering only the leading order (LO) calculations or going forward to the next-to-leading order (NLO) to include virtual corrections.

3. **Parton Showers** The soft and collinear emission is impossible to be handled at the matrix element level. These effects are reproduced by the parton showers. The shower is constructed recursively, increasing the number of partons by one at each step until an energy cutoff at around the hadronization scale is met.

The software that handles the parton showers is PYTHIA [? ].

4. **Matching** The matrix element calculations provide an accurate description of events with well-separated particles in the final state, while the accurate reproduction of collinear and soft emissions is handled by the parton showers.

Matching methods have the scope of improving the accuracy of the description of multijet final states, combining both approaches.

The Monte Carlo samples used in this work are generated using the MLM matching method [? ] if the matrix element calculations are performed at the LO, or using the FxFx method [? ] if the matrix element calculations are performed at the NLO.

5. **Hadronization** The last step is the hadronization that is simulated by exploiting the Lund string model [? ] and is handled by PYTHIA using the TuneCP5 set of tuning parameters [? ].

The reconstruction and the calibration are performed with the *Ultra Legacy 18* (UL18) setting, which reproduces the state of the detector in 2018.

The used samples are NanoAOD files [? ], a lightweight data format that consists only of flat ROOT NTuple.

A comparison of all the used samples is reported in Table ??.

Dataset	Generator	Additional Partons	$\mathcal{L}_I^{\text{MC}} / \mathcal{L}_I^{\text{RunII}}$	Label
t̄ signal (bℓν bcb)	MADGRAPH (LO) +MADSPIN	3	76.5	$\ell = \mu, e, \tau$ signal(Mu,Ele,Tau)
t̄ semiLept (bℓν bqq)	POWHEG (NLO)	1	2.24	TTsemiLept(Mu,Ele,Tau)
t̄ diLept (bℓν bℓν̄)	POWHEG (NLO)	1	1.71	TTdiLept
t̄ diHad (bq̄ bq̄)	POWHEG (NLO)	1	0.87	TTdiHad
W+jets (W → ℓν)	MADGRAPH (LO) +MADSPIN	4	0.01	WJets
WW+jets ℓν q̄q̄	aMC@NLO (NLO) +MADSPIN	1	0.61	WWJets
tW (bℓν q̄q̄)	POWHEG (NLO)	1	0.58	tW
tq (t-channel) bℓν q̄	POWHEG (NLO)	1	0.42	tq

**Table 6.2:** MC samples used in this work along with the respective generator used, the number of additional partons set in the generator, and the ratio between the effective luminosity of the MC sample and the integrated luminosity of the RunII. In the "Label" column there are the names that are used as labels in the rest of the thesis.

Simulated events are normalized to their expected contributions using event weights

$$w_{pe} = \frac{\mathcal{L}_I \cdot \sigma_p \cdot w_e}{\sum_j w_j} \quad (6.1)$$

where  $w_e$  is the event weight produced by the generator program. The events weights can also multiplied with other corrections needed to improve the agreement between MC and data (e.g. flavor corrections, see sec. ??).

## 6.3 Physics objects selections

As already said, the idea of the analysis is to focus on semileptonic (SL) decay to exploit the lepton as a probe, so we are interested in selecting one prompt lepton, along with four jets. The selection of physics objects is crucial to reject non-prompt leptons that can originate from the semileptonic decays of mesons and jets that can originate from prompt leptons or pile-up.

The objects taken into consideration are Particle Flow candidates (see sec. ??) and the applied selection cuts are working points (WP) recommended by the CMS Physics Object Groups (POG).

Since the signal cross-section is  $\sigma_{\text{signal}} \sim 0.3 pb$ , the requirements will be minimal to preserve the statistics.

### 6.3.1 Muons

**Identification** The selected muons are PF candidates reconstructed as *Loose Muons* [?], i.e. muon candidates that are reconstructed either as a global muon or as an arbitrated tracker muon (aTk Muons)<sup>0</sup>.

Furthermore, the selected muons must be in the coverage of the tracker  $|\eta| < 2.4$ , and its transverse momentum has to be  $p_T > 26 \text{ GeV}$  to pass the HLT single muon trigger.

**Isolation** To select prompt muons from the W decay, the particle flow isolation criteria are applied.

The PF isolation is the scalar sum of the transverse momentum of all the charged particles, neutral hadrons, and photons inside a cone of radius  $\Delta R < 0.4$  around the selected muon, subtracting the contributions that arise from pile-up effects [? ].

$$\text{Iso}_\mu = \left( \sum_{\text{charged}} p_T + \max \left( 0, \sum_{\text{neutral}} E_T + \sum_{\text{photons}} E_T - \sum_{\text{PileUp}} p_T / 2 \right) \right) / p_T(\mu) \quad (6.2)$$

The selected working point is the *Loose PF Isolation*:  $\text{Iso}_\mu < 0.25$ , which leads to a muon selection efficiency of  $\sim 98\%$  evaluated on  $Z \rightarrow \mu\bar{\mu}$  samples.

Muons	
Muon Identification	Loose (Global/aTk Muon)
$p_T$	$> 26 \text{ GeV}$
$ \eta $	$< 2.4$
PF Isolation	Loose ( $\text{Iso}_\mu < 0.25$ )

Table 6.3: Muons selection cuts.

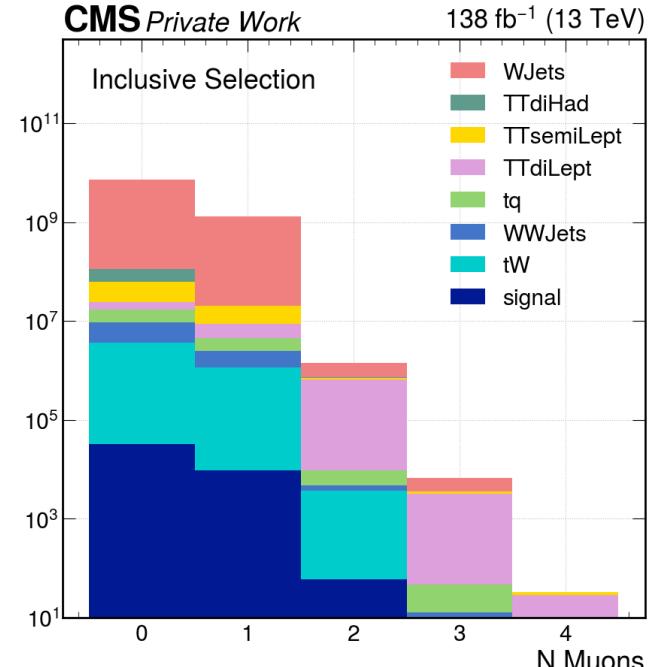


Figure 6.1: Distribution on the number of Muons per event.

<sup>0</sup>Arbitration is the pattern recognition process that assigns each segment uniquely to a single tracker muon.

### 6.3.2 Electrons

The selection of electron candidates follows the same principles as the muon selections. They are required to be into the tracker coverage  $|\eta| < 2.4$  and its momentum has to be  $p_T > 30 \text{ GeV}$ .

The identification and isolation of the electron are based on the *medium* working point of *Fall17IsoV2* [?], a BDT model which assigns a score to the identification and isolation of the electron, taking as input informations about tracks, calo-clusters, shower shape, the amount of collinear radiation to the electron track, and the PF isolation.

The MVA ID is trained on DrellYan+Jets MC samples, with prompt electrons as signal and unmatched plus non-prompt electrons as background, ignoring electrons from tau decays.

The medium working point has an overall signal efficiency of  $\sim 90\%$  and a background rejection efficiency  $> 98.5\%$ .

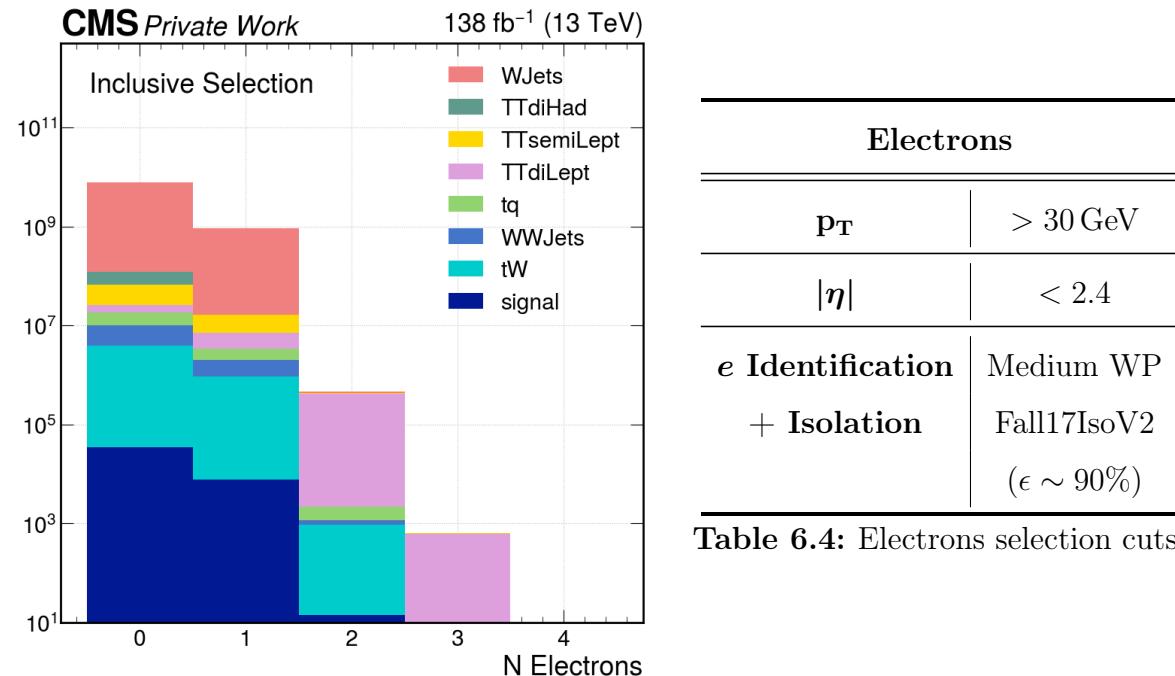


Table 6.4: Electrons selection cuts.

**Figure 6.2:** Distribution on the number of Electrons per event.

### 6.3.3 Jets

The jets taken into consideration are the PF AK4 Jet candidates in the calorimeter coverage  $|\eta| < 4.8$  and with  $p_T > 20 \text{ GeV}$ .

**Identification** The jet identification is based on the *Tight jetID* that imposes different requirements on the composition of the jets in different  $|\eta|$  regions.

The requirements are listed in Tab.??

$ \eta  \in$	[0, 2.6]	[2.6, 2.7]	[2.7, 3.0]	[3.0, 5.0]
<b>Neutral Hadron Fraction</b>	< 0.90	< 0.90	-	> 0.20
<b>Neutral EM Fraction</b>	< 0.90	< 0.99	$\in [0.02, 99]$	< 0.90
<b>Number of Constituents</b>	> 1	-	-	-
<b>Charged Hadron Fraction</b>	> 0	-	-	-
<b>Charged Multiplicity</b>	> 0	> 0	-	-
<b>Number of Neutral Particles</b>	-	-	> 2	> 10

**Table 6.5:** Tight jet identification requirements.

The efficiency of the tight Id working point is  $\epsilon > 99\%$ , whereas the background rejection is  $> 98\%$ .

**Cross-cleaning** Since the leptons are included in the jet clustering process, to avoid considering the leading lepton as an additional jet, a simple  $\Delta R$  matching is performed. If the jet axis and the leading lepton are separated less than  $\Delta R < 0.4$ , the jet is removed.

**Pile-up rejection** The identification of jets from pile-up relies on three types of properties of the jets:

- In the tracker acceptance, the tracks associated with the jets can be used to match the jet with the primary interaction vertex.
- The shape of the calorimetric deposit is different in case of overlap of multiple interactions.
- The objects' multiplicity.

The identification of jets from pile-up is done exploiting the *Jet puid*, a BDT model score [? ]. Its inputs are listed in Tab ??.

This BDT is trained on DrellYan+Jets MC samples with jets of  $p_T < 50$  GeV, where there is the highest composition of PU jets.

Input variable	Definition
$LV \sum p_T$ fraction	Fraction of $p_T$ of charged particles associated with the LV, defined as $\sum_{i \in LV} p_{T,i} / \sum_i p_{T,i}$ where $i$ iterates over all charged PF particles in the jet
$N_{\text{vertices}}$	Number of vertices in the event
$\langle \Delta R^2 \rangle$	Square distance from the jet axis scaled by $p_T^2$ average of jet constituents: $\sum_i \Delta R^2 p_{T,i}^2 / \sum_i p_{T,i}^2$
$f_{\text{ring}X}, X = 1, 2, 3, \text{ and } 4$	Fraction of $p_T$ of the constituents ( $\sum p_{T,i} / p_T^{\text{jet}}$ ) in the region $R_i < \Delta R < R_{i+1}$ around the jet axis, where $R_i = 0, 0.1, 0.2, \text{ and } 0.3$ for $X = 1, 2, 3, \text{ and } 4$
$p_T^{\text{lead}} / p_T^{\text{jet}}$	$p_T$ fraction carried by the leading PF candidate
$p_T^{\text{ch.}} / p_T^{\text{jet}}$	$p_T$ fraction carried by the leading charged PF candidate
$ \vec{m} $	Pull magnitude, defined as $ (\sum_i p_T^i  r_i \vec{r}_i ) / p_T^{\text{jet}} $ where $\vec{r}_i$ is the direction of the particle $i$ from the direction of the jet
$N_{\text{total}}$	Number of PF candidates
$N_{\text{charged}}$	Number of charged PF candidates
$\sigma_1$	Major axis of the jet ellipsoid in the $\eta\phi$ space
$\sigma_2$	Minor axis of the jet ellipsoid in the $\eta\phi$ space
$p_T^D$	Jet fragmentation distribution, defined as $\sqrt{\sum_i p_{T,i}^2 / \sum_i p_{T,i}}$

**Table 6.6:** Puid BDT input variables [? ].

The chosen working point is the *Loose puid*, which corresponds to 99% efficiency for prompt jets with  $|\eta| < 2.5$  and 95% efficiency for prompt jets with  $|\eta| > 2.5$ .

**Jet energy scale corrections** To allow a proper mapping of the energy of the reconstructed jet to the truth particle-level jet energy, a set of jet energy corrections is applied.

In CMS, the jet energy scale corrections are separated into different steps applied sequentially and in a fixed order.

Each correction takes care of a different effect and consists of a rescaling of the jet four-vector [? ].

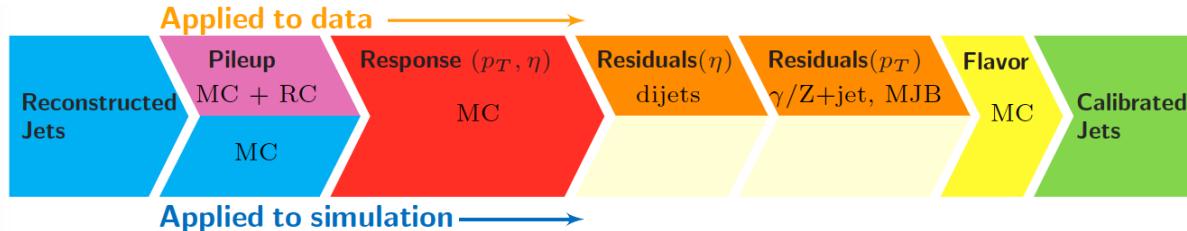


Figure 6.3: Jet energy scale correction chain.

1. **L1 Pile-Up:** remove the energy coming from pileup events, removing any dependence on the luminosity.

The pileup corrections are determined from simulated QCD dijet events with and without pileup overlay.

2. **L2L3 MC-truth corrections:** this correction has the scope to improve the accordance between the reconstructed jet energy and the particle-level energy, and to make the response of the jet energy uniform in  $p_T$  and  $\eta$ . The related scale factors are evaluated on QCD dijet samples.

The pile-up and MC-truth corrections are applied both on MC and data, while additional residual corrections are applied only on data: they are meant to correct differences of the order of 1% within jet response in data and MC.

A summary of the jet energy corrections chain is shown in Fig. ??.

**Jet energy resolution corrections** The jet momentum resolution is different in the data with respect to the simulation. To replicate the resolution observed in the data, the jets' four vectors are rescaled in a procedure called *smearing* [? ]:

if a reconstructed jet and a jet at the generator level (GenJet) are matched through the relation  $\Delta R(\text{Jet}, \text{GenJet}) < 0.2$  and  $|p_T - p_T^{\text{Gen}}| < 3\sigma p_T$ , the jet four-vector is scaled by a factor

$$f = 1 + SF \frac{p_T - p_T^{\text{Gen}}}{p_T} \quad (6.3)$$

where  $\sigma$  is the jet energy resolution measured with data, and SF is the scale factors that depend on the  $\eta$  region in which the jet belongs. Otherwise, if there is not any match between the jet and the genjets, a random Gaussian smearing is applied.

**Flavor tagging corrections** The distribution shapes of the DEEPJET discriminators are another element that shows discrepancies between data and simulated samples.

To address this problem the events are reweighted: a weight is applied to each selected jet and an event weight is defined as the product of all the jet weights inside the event. The event weights obtained from these corrections are then multiplied by the event weights provided by the generator program, as already explained in sec. ??.

In this work, the applied corrections are the ones related to the b tagging and to the c tagging.

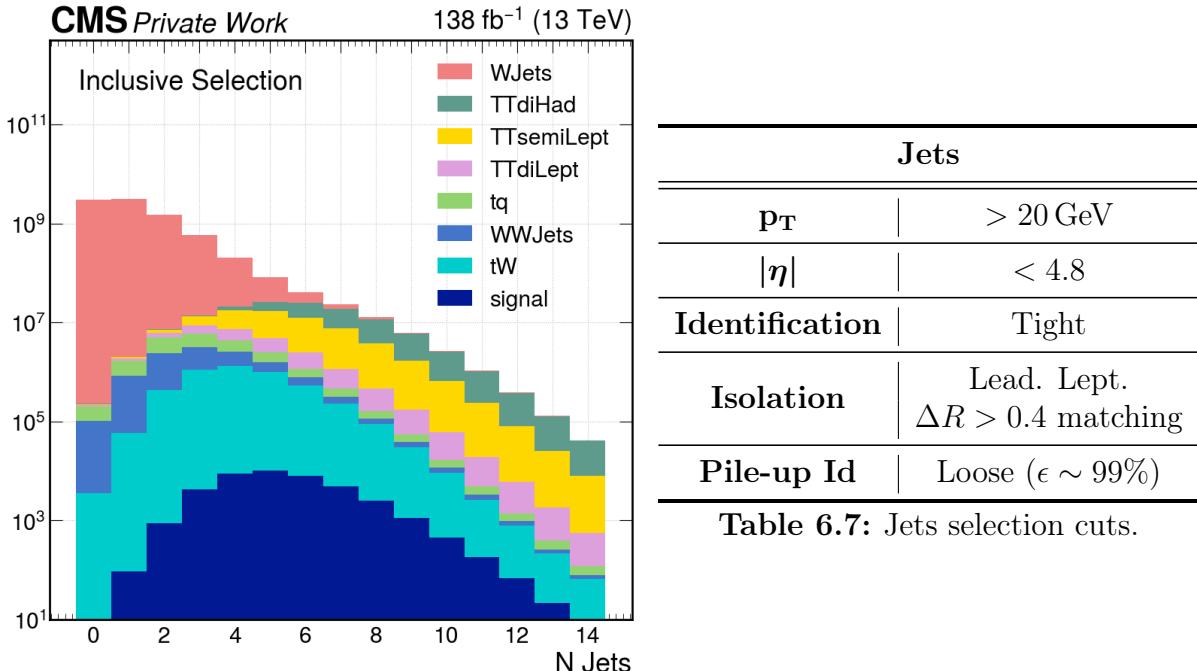
These scale factors are computed through an iterative tag-and-probe procedure [?], in different bins of  $\eta, p_T$ , and DEEPJET score to compute simultaneously SF for both heavy and light flavor jets.

- For heavy-flavor jets, dileptonic  $t\bar{t}$  events are exploited, using one b jet that passes the medium working point as a tag and the second b jet as a probe.
- For light jets,  $Z+jets$  dileptonic samples are used, using a b tag loose working point as a veto to select the tag jet.

The procedure to obtain SFs related to the c-tagging is similar, using  $W+c$  samples, but for them, the binning depends both on the CvB and CvL discriminators defined as

$$\begin{aligned} \text{CvL} &= \frac{P(c)}{P(c) + P(uds)} \\ \text{CvB} &= \frac{P(c)}{P(c) + P(b)} \cdot P(b) \end{aligned} \quad (6.4)$$

where  $P(b)$ ,  $P(c)$ , and  $P(uds)$  are the b,c, and light flavor DEEPJET scores.

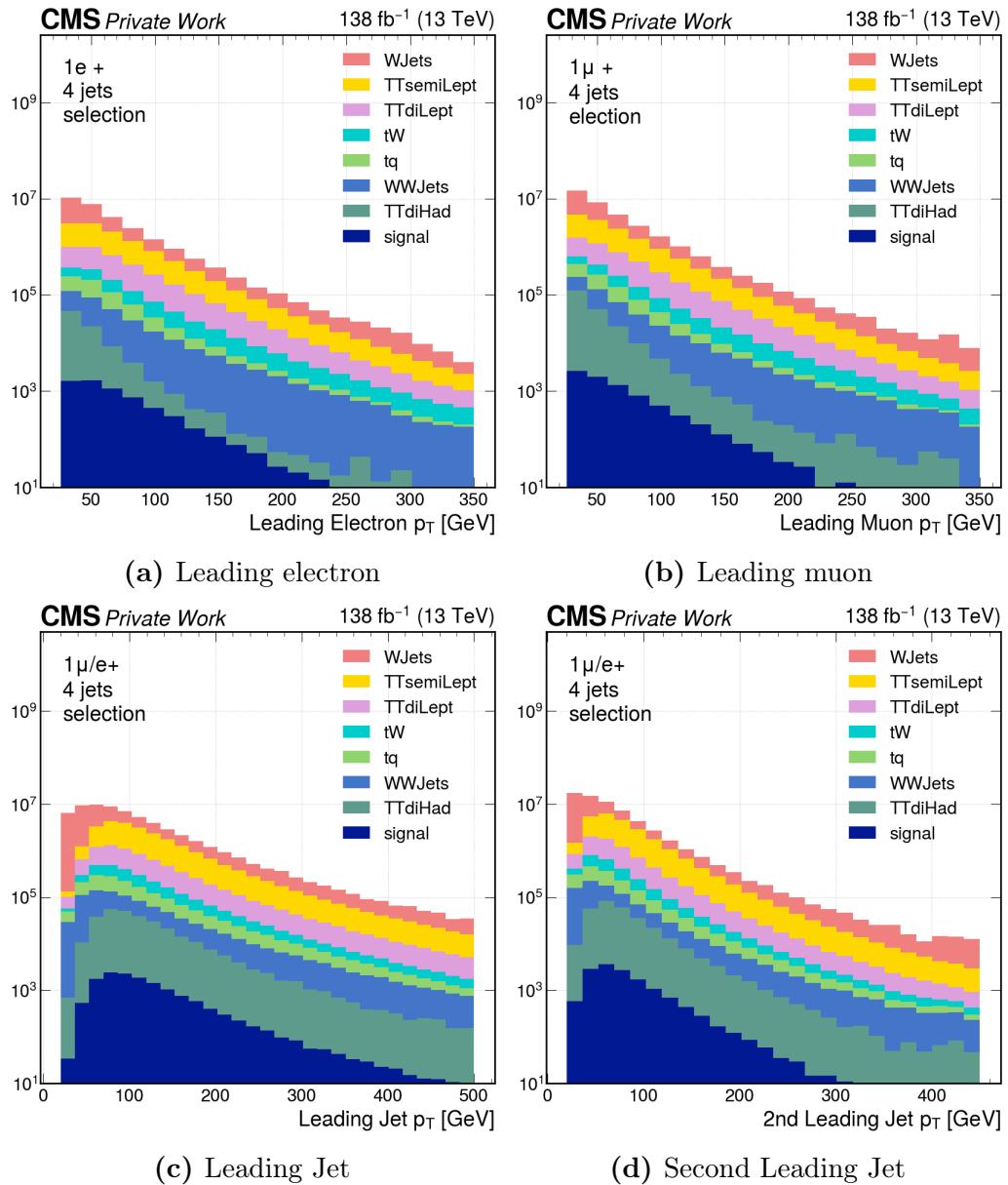


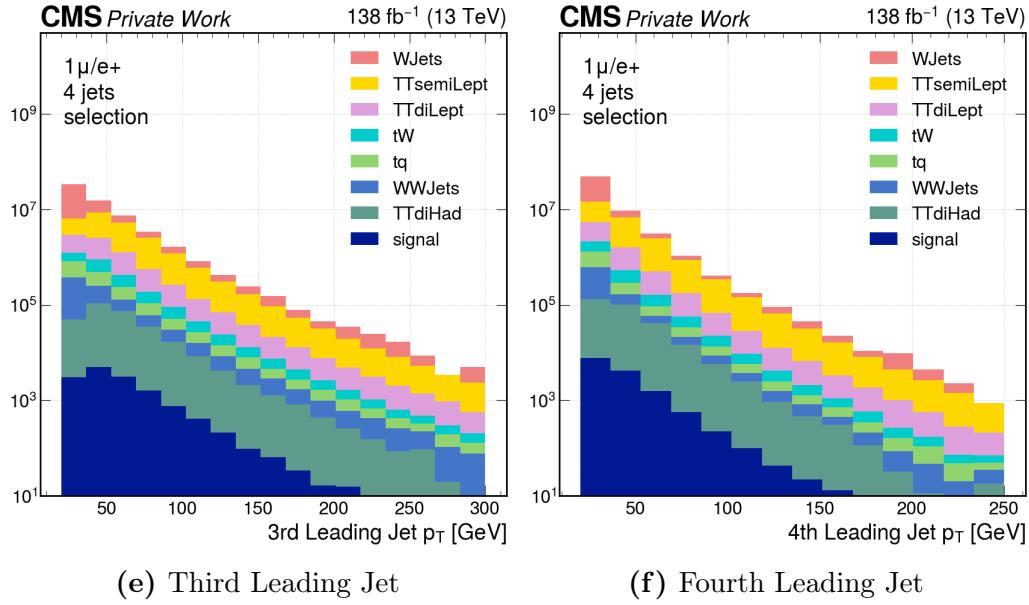
**Figure 6.4:** Distribution on the number of Jets per event.

## 6.4 Event selections

As already discussed, the chosen event selection is minimal to avoid large losses in signal acceptance efficiency.

The signal events are targeted in two different final states, defined by the flavor of the lepton used as a probe, so after the physics objects selection and the requirement of the presence of at least 4 jets, two different channels are defined: a Muon channel that contains events with at least one selected muon, and an Electron channel that contains events with at least one selected electron.





**Figure 6.5:**  $p_T$  distributions after the object selection and the  $1e/\mu + 4$  jets requirements.

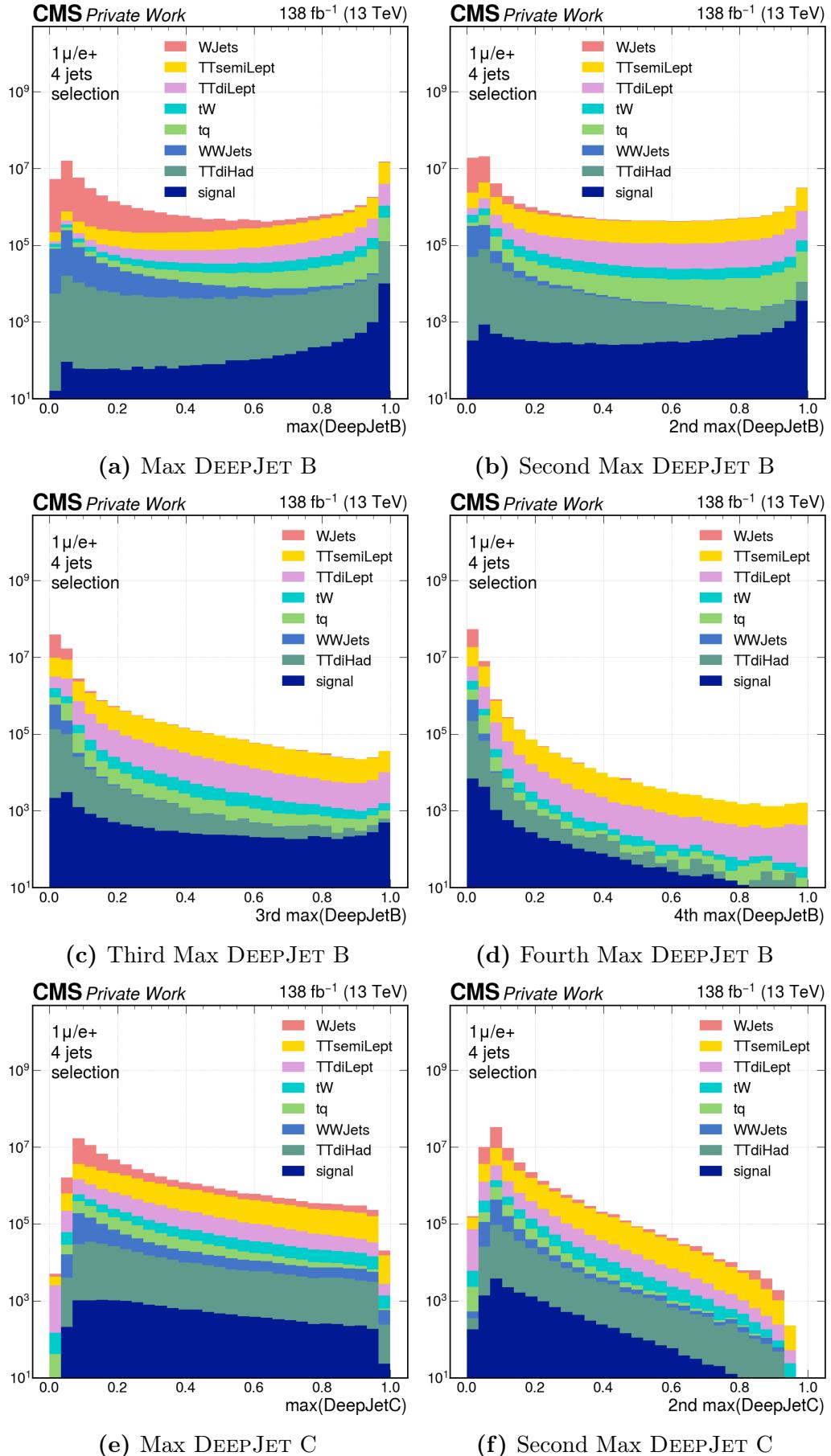
An additional selection on the most b-tagged jet is imposed using the medium working point.

The cumulative efficiencies of each cut are described in Tab. ?? and the final event counts normalized with the RunII luminosity and the respective cross-sections in Fig. ??

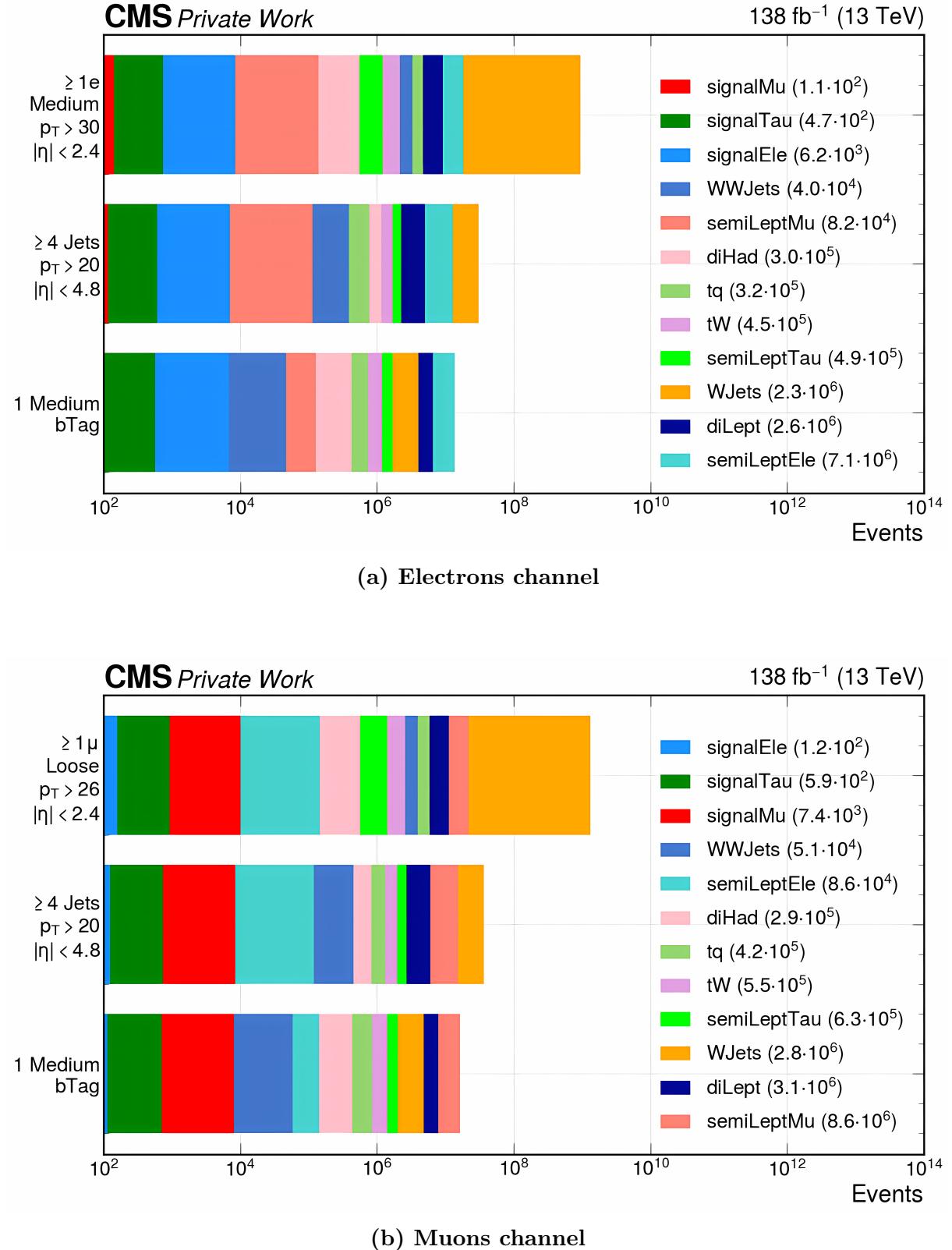
	Muons			Electrons		
Cumulative efficiencies	$\geq 1\mu$	$\geq 4$ Jets	$\text{max(btag)} > 0.27$	$\geq 1e$	$\geq 4$ Jets	$\text{max(btag)} > 0.27$
<b>signalMu</b>	64.1%	53.5%	51.8%	0.22%	0.18%	0.16%
<b>signalEle</b>	0.65%	0.50%	0.46%	52.3%	43.5%	42.1%
<b>signalTau</b>	4.86%	3.95%	3.80%	3.37%	2.79%	2.70%
<b>TTsemiLeptMu</b>	63.6%	54.0%	49.8%	0.15%	0.12%	0.09%
<b>TTsemiLeptEle</b>	0.48%	0.39%	0.30%	51.9%	45.6%	41.7%
<b>TTsemiLeptTau</b>	4.59%	3.84%	3.48%	3.24%	2.82%	2.58%
<b>TTdiLept</b>	40.9%	27.4%	25.3%	33.8%	22.7%	20.9%
<b>TTdiHad</b>	0.45%	0.40%	0.31%	0.16%	0.15%	0.12%
<b>WJets Lept</b>	15.0%	0.24%	0.03%	10.6%	0.20%	0.03%
<b>WWJets SL</b>	19.5%	5.53%	0.79%	15.3%	4.42%	0.64%
<b>tW SL</b>	23.6%	13.6%	11.3%	19.0%	10.9%	9.11%
<b>tq Lept</b>	21.6%	6.23%	5.42%	15.1%	4.40%	3.81%

**Table 6.8:** Cumulative efficiencies of the selections in the Muons and Electrons channel after the physics objects selection

The main background contribution in each channel is from the  $t\bar{t}$  semileptonic decay, which is 3 orders of magnitude greater than the signal. Another significant contribution is from the  $t\bar{t}$  dileptonic (DL) process. The contribution from the WW+Jets process is two orders of magnitude smaller than  $t\bar{t}$  SL and  $t\bar{t}$  DL processes, while the remaining processes' contribution is one order of magnitude smaller than the leading backgrounds.



**Figure 6.6:** B-tag and c-tag score distributions after the object selection and the  $1 \mu/e + 4$  jets requirements.



**Figure 6.7:** Event number for each process for each cut. In the legend the final number of events for each process after the preselection is shown. The processes are ordered in their final event number.

## 6.5 Neutrino reconstruction

To fully reconstruct the kinematics of the event and to allow the reconstruction of the leptoninc decaying top quark, the longitudinal component of the neutrino was reconstructed exploiting the transverse momentum of the missing energy, its pseudorapidity, the lepton's four vector, and fixing the mass of the W boson to  $m_W = 80.385$  GeV.

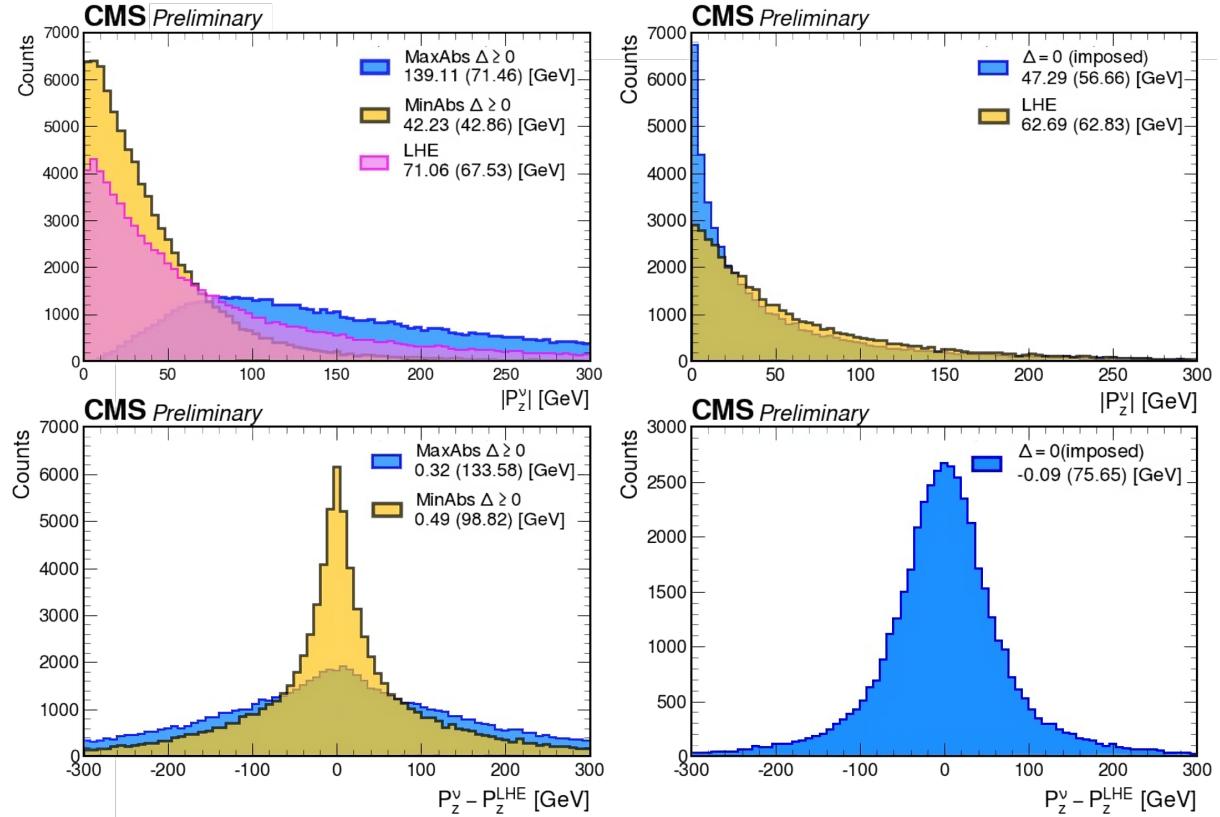
$$(p_\nu^\mu + p_\ell^\mu)^2 = m_W^2 \quad (6.5)$$

The equation to solve for  $p_z^\nu$  is a second order equation

$$p_z^\nu = \frac{p_z^\ell(m_W^2 + 2\vec{p}_T^\ell \cdot \vec{p}_T^\nu) \pm \sqrt{(p_z^\ell)^2(m_W^2 + 2\vec{p}_T^\ell \cdot \vec{p}_T^\nu)^2 - |p_T^\ell|^2[4(E_\ell p_T^\nu)^2 - (m_W^2 + 2\vec{p}_T^\ell \cdot \vec{p}_T^\nu)^2]}}{2|p_T^\nu|^2}$$

The equation has two solutions, that can also be complex. If the equation has a complex solution the real part is taken but, if the equation allows two real solutions, we must choose one of them.

In Fig. ?? it is shown a comparison between the solution with the smaller magnitude, the solution with the greater magnitude, and the truth value of  $p_z^\nu$  at the LHE level. This study was performed using signal samples.

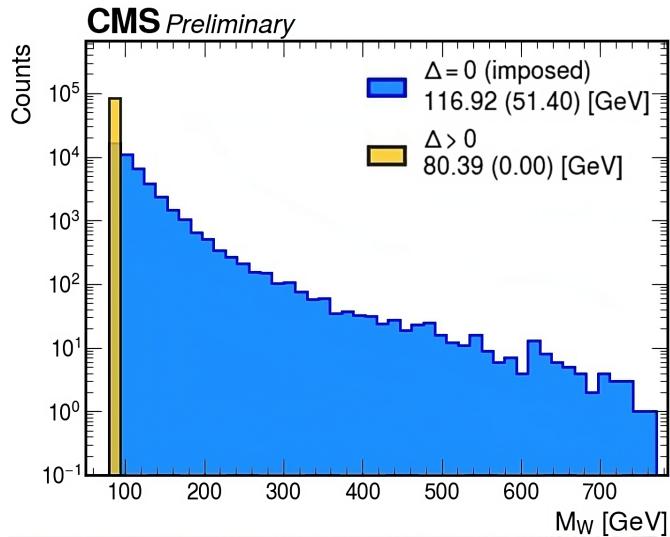


**Figure 6.8:** In the left panels there are the events that have two real solutions, while in the right panels the ones with the complex solutions of which only the real part is shown. In the top panel the distribution of the absolute value of the solutions and the LHE truth value are shown, while in the bottom panels, there are the distributions of the differences between the solutions and the LHE truth values. The  $\Delta$  in the legend is the discriminant of the second-order equation. The distributions are in arbitrary units.

The distribution of the solutions with the smaller absolute value is in better accordance with the truth value, so the strategy will be:

- The solutions are complex: take the real part.
- There are two real solutions: take the smallest in the absolute value.

Of course, in the complex solution case, if we consider only the real part, the chosen value will violate the W mass constraint, so, the invariant mass of the sum of the lepton and the reconstructed neutrino will not be  $m_W$ . In Fig. ?? the distribution of the leptonic W mass is shown.



**Figure 6.9:** Distribution of the reconstructed W mass in the real solution case and the complex solution case, taking only the real part.

## *Chapter 7*

---

# Signal reconstruction and interpretation

---

This chapter is dedicated to the kinematical reconstruction and interpretation of the signal events, tackling the Jet-parton assignment problem.

Initially, we tackle the jet-parton assignment in a simplified form by focusing on the selection of the jet associated with the b quark parton that arises in the decay of the top quark that decays into the leptonic W boson. This simplified benchmark is used to perform a feature ranking exploiting the  $N + 1$ ,  $N - 1$  method and to select the best model that will be employed in the complete jet-parton assignment task and that will be adapted to discriminate the signal from the background, as presented in the next chapter.

All the investigations presented in this chapter are carried out within the Muon signal region, as defined in Table ??, after the kinematic reconstruction of the neutrino described in sec. ??.

### 7.1 Jet-parton assignment

The Jet-Parton assignment (JPA) task consists of associating a jet to each of the four quark partons that belong to the signal final state to reconstruct the event's topology and kinematics.

This task is challenging due to the combinatorial nature: if there are  $N$  jets in an event, there are a total of  $N!/(N - 4)!$  possible combinations for assigning a jet to the 4 partons, so, for example, in the case of an event with 10 jets, there are over 5000 possible combinations.

Typically, the assignment is accomplished through a kinematic fit that consists of finding the combination of jets for each event that minimizes the chi-square of all the

invariant masses.

$$\chi^2 = \frac{(m_{j_1 j_2}^{\text{inv.}} - m_W)^2}{\sigma_W^2} + \frac{(m_{j_1 j_2 j_3}^{\text{inv.}} - m_t)^2}{\sigma_{t_H}^2} + \frac{(m_{j_4 \ell \nu}^{\text{inv.}} - m_t)^2}{\sigma_{t_L}^2} \quad (7.1)$$

where  $\sigma_W$ ,  $\sigma_{t_L}$ , and  $\sigma_{t_H}$  are the experimental dijet and trijet invariant mass resolution widths of the hadronic W and leptonic/hadronic top quark decays, respectively.

This is an unsupervised approach but in this work, another approach is employed, leveraging supervised multivariate models like neural networks that allow us to exploit not only the kinematic variables of jets but also their flavor score and the correlations between all the observables.

Signal events	
Muon region	
$5.7 \cdot 10^7$	
$\Delta R(\text{part.-jet}) < 0.4$	
$\forall \text{parton } (43\%)$	
$2.5 \cdot 10^7$	
Training (80%)	Validation (20%)
$2.0 \cdot 10^7$	$5 \cdot 10^6$

**Table 7.1:** Number of events of the training and validation datasets employed to perform the JPA.

The dataset employed to conduct this study is composed of  $5.7 \cdot 10^7$  signal events that have passed the muon channel selections.

To avoid ambiguities and to assign to each jet a label, a further requirement is imposed: all the four quark partons at the generator level in the event have to match a distinct reconstructed jet through the condition  $\Delta R < 0.4$ .

The fraction of signal events in the muon channel that satisfy this requirement is 43%. Of these 25 million of events, 80% are used to train the models, while the remaining 20% are reserved for the model validation.

### 7.1.1 Leptonic bJet benchmark

In principle, the b quark parton in the signal final state originating by the leptonic top decay is the easiest to assign to a reconstructed jet for two reasons:

- The top quarks are produced back-to-back in the transverse plane and the decay products of each top quark are in the direction of the momentum of the respective top quark.  
Given that, the b parton originating by the leptonic top decay is closer on average to the lepton in the  $\phi - \eta$  plane than the other quark partons.
- Since it is a heavy flavor parton, we can exploit the b-tagging capabilities of the experiment. In an ideal situation, the b-tagging would simplify the problem by reducing the number of jet candidates to just three.

Several multivariate models were evaluated to select the most accurate that will be used to perform the jet-parton assignment on all four quark partons.

In the upcoming sections, we will refer to the b parton originating by the leptonic top decay as "the leptonic b parton" and the corresponding jet with "the leptonic b jet". As well, the problem of assigning the leptonic b parton to a jet for each event will be denoted as "simplified JPA problem".

## Jet-wise models

The initial class of models tested uses as input only observables associated with the jets. In other words, the event's structure is disregarded and the data provided to the models is a matrix that has different jets on the rows and different features of the jets on the columns.

The inputs fed to the models are shown in Tab. ??.

The simplified jet-wise JPA problem is a binary classification task where the model's goal is to determine whether a particular jet is associated with the leptonic b parton.

The event's structure is reconstructed afterward, and the jet with the highest score within each event will be classified as the leptonic b jet. The accuracy is then defined as the ratio between the number of events in which the leptonic b parton is associated with the right jet and the total number of events.

Input feature	Description
Jet_pt	Jet $p_T$
Jet_eta	Jet $\eta$
Jet_mass	Jet invariant mass
Jet_btag	DEEPJET b-tag score
Jet_CvL.	DEEPJET CvL score
Jet_CvB	DEEPJET CvB score
max_dEta_Jets	$\max(\Delta\eta)$ computed with respect to other jets in the event.
min_dEta_Jets	$\min(\Delta\eta)$ computed with respect to other jets in the event.
max_dPhi_Jets	$\max(\Delta\phi)$ computed with respect to other jets in the event.
min_dPhi_Jets	$\min(\Delta\phi)$ computed with respect to other jets in the event.
dPhi_Jet_nu	$\Delta\phi$ Jet- $\nu$
dEta_Jet_nu	$\Delta\eta$ Jet- $\nu$
dEta_Jet_mu	$\Delta\eta$ Jet- $\mu$
dPhi_Jet_mu	$\Delta\phi$ Jet- $\mu$
T_mass	Jet+ $\mu + \nu$ invariant mass

**Table 7.2:** Input features of the jet-wise models tested for the simplified JPA task.

The jet-wise models tested are four:

- **Fisher discriminant** The Fisher's linear discriminant [?] is a model whose goal is to find a hyperplane that maximizes the separation between the classes. The Fisher discriminant identifies two regions, separating the predicted leptonic b jets from the other jets. The absolute value of the score assigned to each jet is the distance between the jet represented in the jet's feature space and the separating hyperplane, while the score sign will be positive if the jet belongs to the predicted leptonic b jet region, and negative otherwise. This is a linear model and its complexity is not sufficient to tackle the JPA problem. Indeed, its accuracy on both training and validation datasets is just 45%.
- **k-NN** k-Nearest Neighbors (k-NN) [?] is a model that leverages the Euclidean distance within the normalized jet feature space to classify the data.

After choosing the parameter  $k$ , the model classifies a jet as a leptonic b jet if the majority of the  $k$  nearest jets in the jet feature space are labeled as leptonic b jet and vice versa.

The score assigned to each jet is determined by the fraction of jets labeled as leptonic b jets within the  $k$ -neighborhood of the jet itself.

The model was assessed for various values of  $k$ , and the corresponding accuracies are shown in Tab. ??

$k$	Accuracy
5	48%
10	51%
20	54%
40	55%
80	55%
500	55%

**Table 7.3:** k-NN accuracy for different values of  $k$ .

- **BDT** A boosted decision tree (BDT) is an ensemble of multiple decision trees. [? ]. that assigns a score to each jet through a weighted combination of the individual tree predictions, with more weight given to the trees that perform better.
- The BDT employed for the simplified JPA task is an ensemble of 200 decision trees with a maximum depth of five, and its accuracy was found to be 65% on both the training and the validation datasets.
- **MLP** The highest-accuracy jet-wise model tested is a feed-forward neural network. The network consists of three hidden layers, each containing 40 neurons and using the ReLU as the activation function. The minimized loss function is the cross-entropy reweighted to address the class imbalance, and the output layer is equipped with the softmax function in such a way as to assign a score in the  $[0,1]$  interval to each jet.
- This model maintains an accuracy of 67% on both the training and validation datasets and exhibits stability in performance even with increased network size.

Model	Accuracy
Fisher	45%
k-NN	55%
BDT	65%
MPL	67%

**Table 7.4:** Jet-wise models accuracy summary.

## Feature ranking

The most straightforward approach for identifying discriminative features is to examine the distributions of various observables and select those with the most distinct shapes. The "shape difference" between two histograms can be quantified by the following metric

$$d = \frac{1}{2} \sum_b \left| y_b^{(1)} - y_b^{(2)} \right| \quad (7.2)$$

where  $y_b^{(i)}$  is the bin height of the  $b$ -th bin of the  $i$ -th histogram. The histograms are normalized such that their integral is 1.

However, this method is not satisfactory since it does not account for correlations between the observables, therefore, more sophisticated multivariate methods can be employed.

As demonstrated in the previous section, the MLP model outperforms other jet-wise models in terms of accuracy, so we can use it to rank the jet's features.

To accomplish this, we employ a method known as the " $N + 1, N - 1$ " approach, which involves two distinct steps:

1.  **$N - 1$  step:** In this step, the model is initially trained and evaluated using all input features. Then, it is retrained  $N$  times, each time removing a different feature. After these  $N$  training iterations, the feature that, when removed, has the least impact on accuracy is eliminated from the feature set. This process is repeated until only one feature remains.

After the  $N$  training, the input feature that, if taken out, changes less the accuracy is removed from the feature set, and the method is iterated until all features but one are removed.

2.  **$N + 1$  step:** beginning with the feature with the highest  $d$  score, we gradually introduce the feature that most significantly enhances accuracy into the feature set iteratively.

The  $N + 1, N - 1$  method is based on the idea that important features have a significant impact on the model's performance, and by systematically adding or removing features, we can assess their importance. Since the complexity of the network is fixed, the model is more prone to overfitting when trained with a smaller set of input features, so, to prevent it, an early stopping strategy was employed.

The  $N+1$  and  $N-1$  rankings are shown in figure Fig. ?? and Tab. ?? lists the eight features that significantly impact the model's accuracy. Both the  $N+1$  and  $N-1$  steps converge to the same eight features, albeit in a slightly different order.

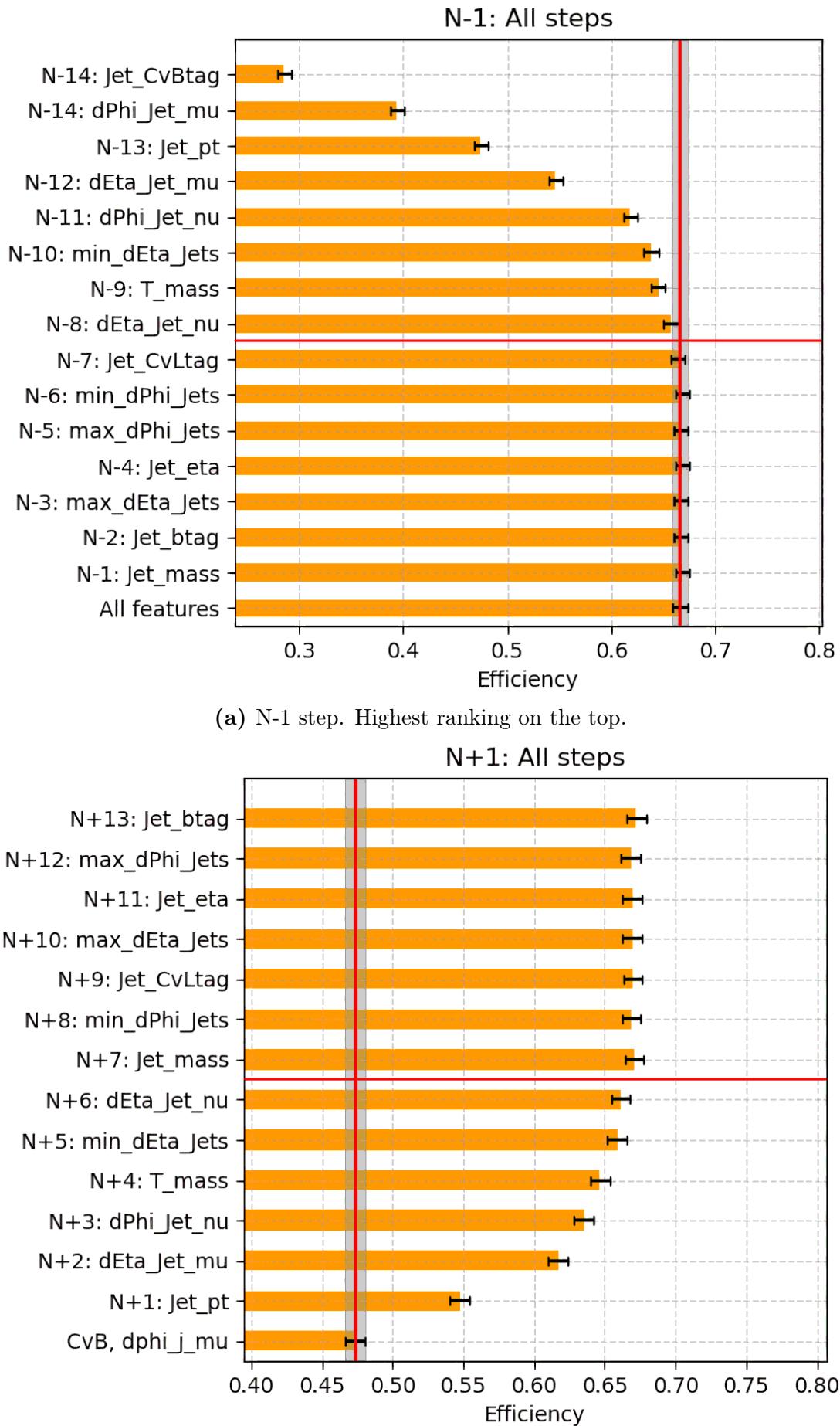
The most highly ranked feature is the CvB score, as we are seeking to detect a b-jet. However, upon repeating the procedure, the btag and the CvB score may interchange positions in the ranking.

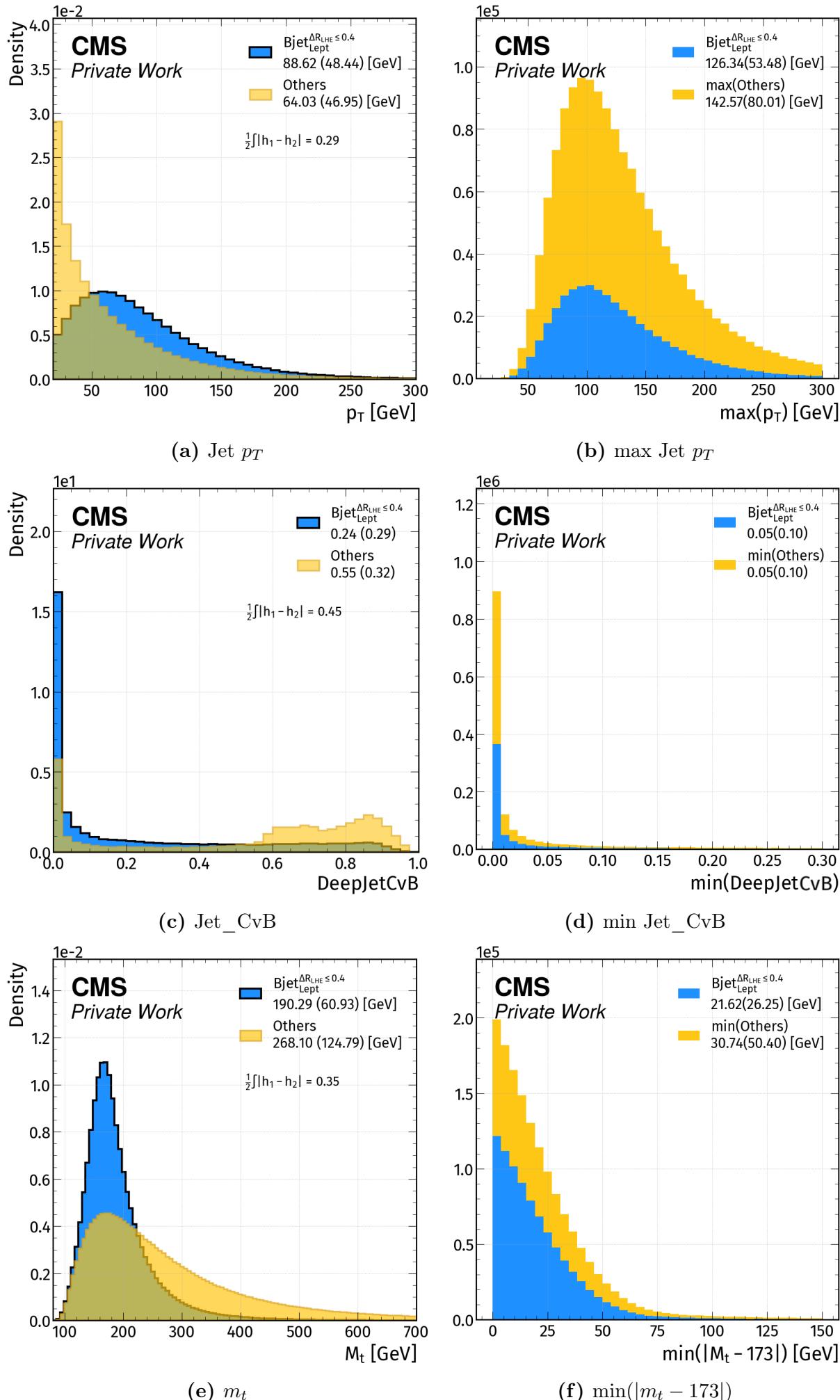
This highlights a correlation and redundancy between these two features.

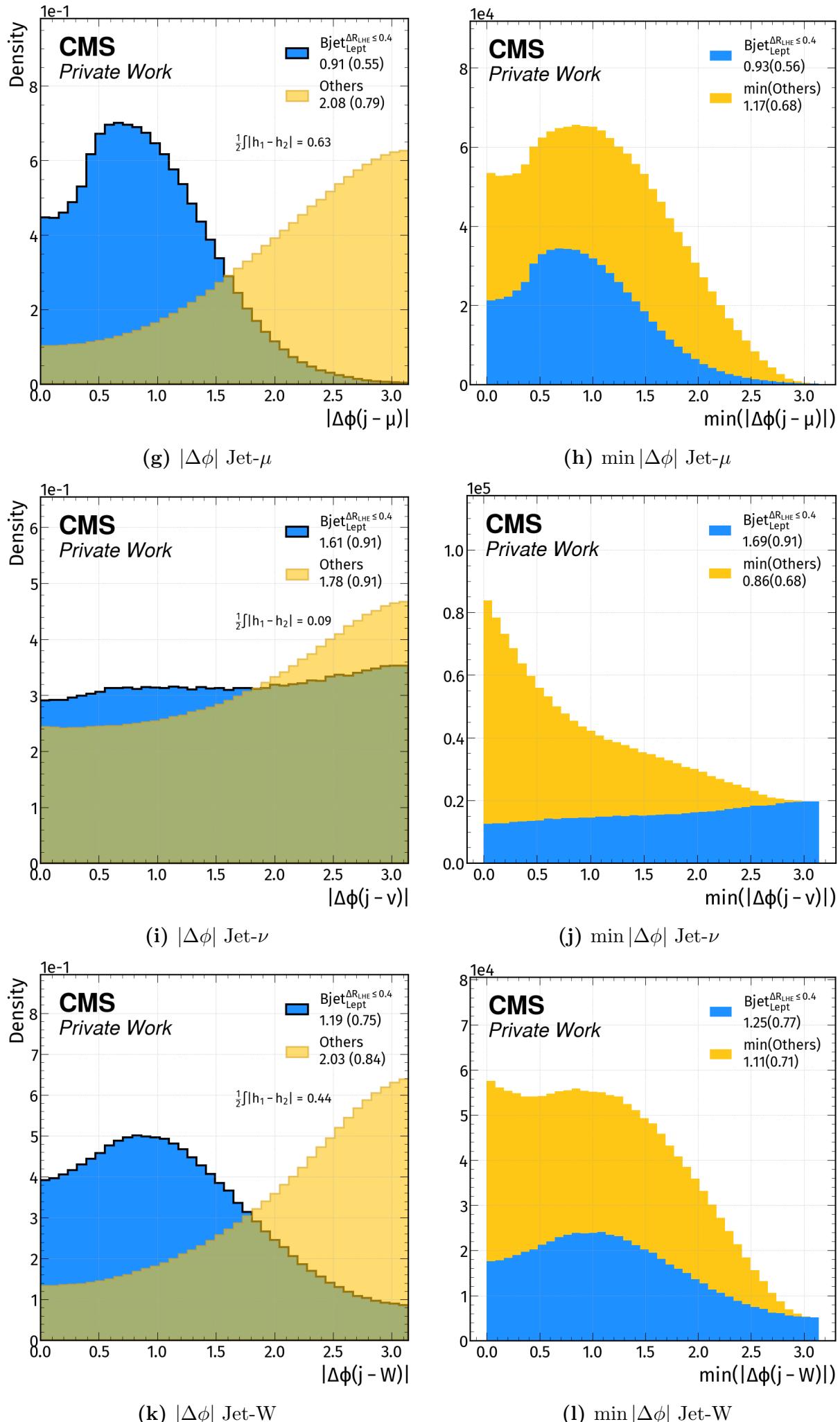
Based on the earlier discussion regarding the proximity of the leptonic b-jet to the lepton, these eight features encompass all the aspects related to the separation between the jet, lepton, and neutrino. Additional relevant features include the invariant mass of the jet, lepton, and neutrino system, which must be compatible with the top quark mass, as well as the jet's transverse momentum since the  $p_T$  of the jets originating from the top quark decay is slightly higher than that of the jets originating from the hadronic W decay.

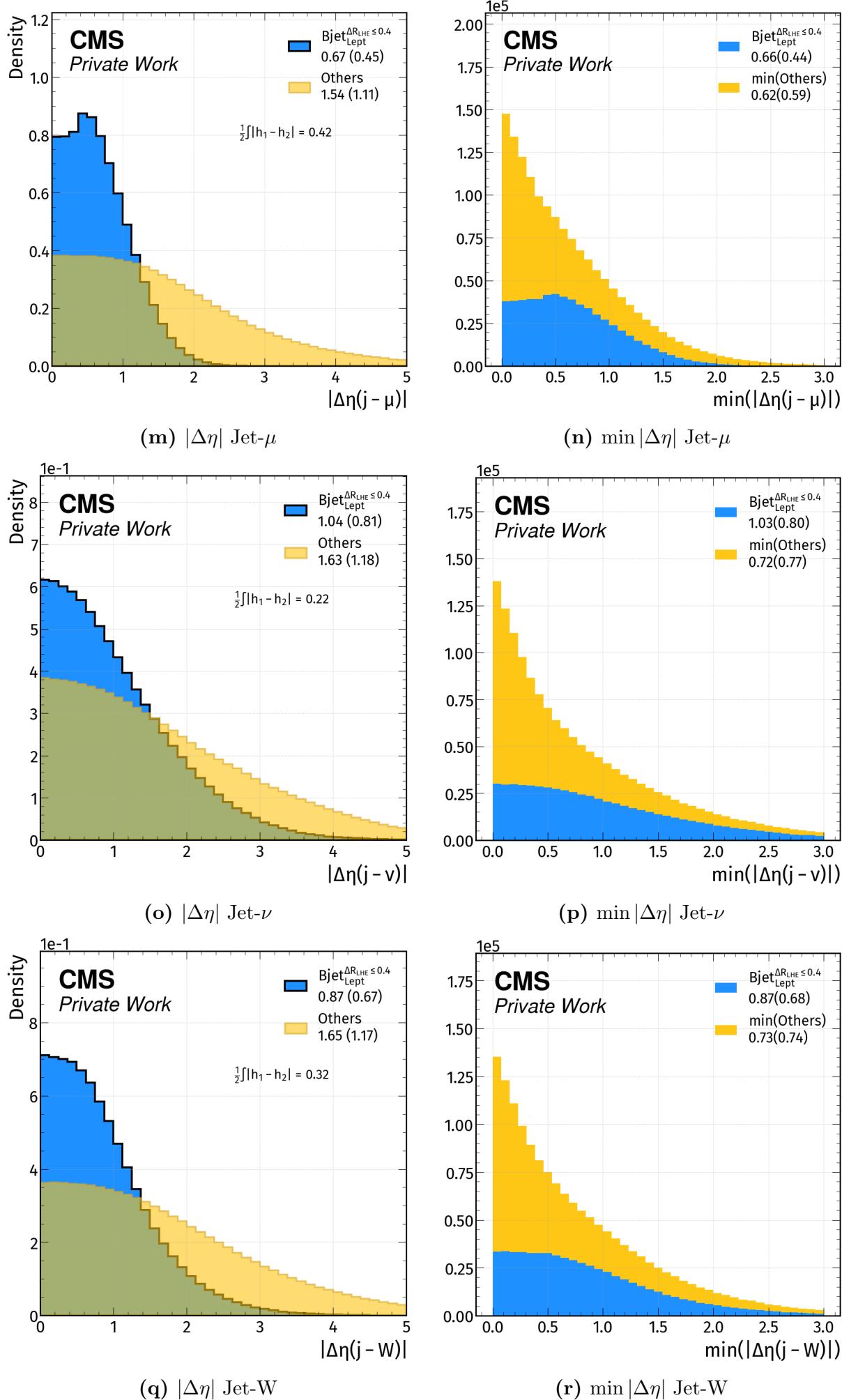
Jet feature	N-1 rank	N+1 rank	$d$
Jet_CvB	1	2	0.45
dPhi_Jet_mu	2	1	0.63
Jet_pt	3	3	0.29
dEta_Jet_mu	4	4	0.42
dPhi_Jet_nu	5	5	0.09
min_dEta_jets	6	7	0.24
T_mass	7	6	0.35
dEta_Jet_nu	8	8	0.22

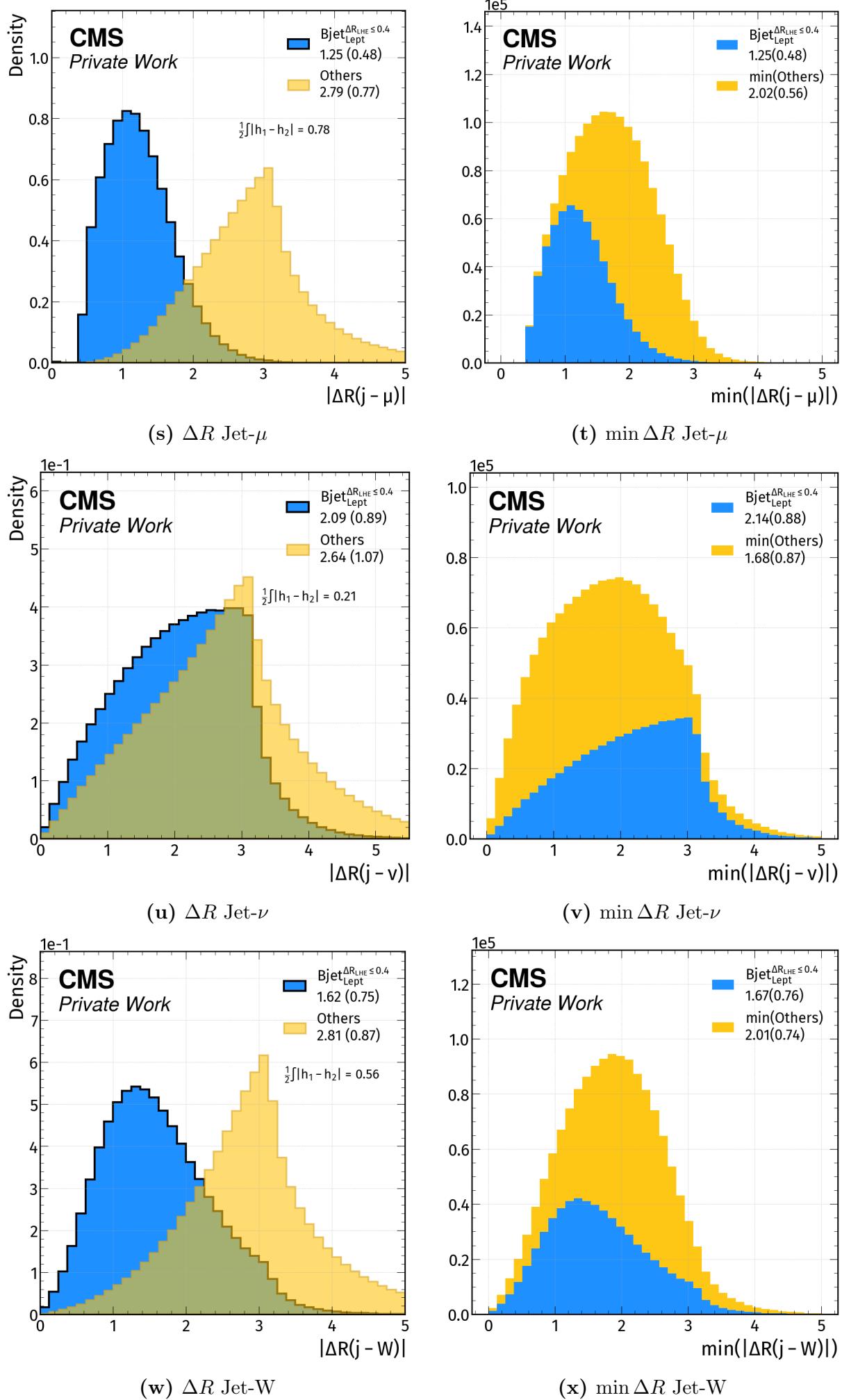
**Table 7.5:**  $N+1, N-1$  ranking of the eight features that have an impact on the accuracy.

**Figure 7.1:** Jet-wise N+1,N-1 feature ranking.









**Figure 7.2:** Distributions of the most relevant features. On the left side, there are the distributions of the observables of the leptonic b Jet and of the other jets, both normalized to 1. On the right side, there are the stacked histograms of the distributions of the minimum or the maximum of each observable for each event. The legend includes information on the mean and standard deviation.

### Event-wise models

Another category of models under examination falls into the event-wise models class. These models receive a matrix as input, where rows represent various events, and columns encompass diverse event and object characteristics. Event-wise models retain the event structure, providing a more direct approach to addressing the JPA problem and eliminating the need to reconstruct the event after the network evaluation.

However, constructing such a model is not trivial: An MLP, for instance, would have a distinct set of input units for each jet or lepton, each assigned to different features, and  $N$  outputs, where  $N$  is the maximum number of jets in the events. The prediction would involve identifying the index of the predicted leptonic b Jet for each event.

This approach has multiple criticalities:

- Since the labels correspond to the index of the jet within the event, the model lacks permutation invariance, which necessitates introducing an order among the jets in the event.
  - Although the network's output may appear as a straightforward index, it constitutes a classification problem with  $N$  classes that leads to two complexities:
    - The high dimensionality of the output space demands a significant amount of data for effective population, a well-known challenge referred to as the "curse of dimensionality."
    - To address the imbalance between the  $N$  classes, the cross-entropy loss must be reweighted, but, since the network is not permutation invariant, neither is the reweighting.
  - $N$  is the maximum number of jets considered in the event, but, when an event has less than  $M$  jets, dealing with the units related to the remaining  $N-M$  jets becomes a concern.
- We could discuss different methods to assign a value for the features of the absent jets, but no choice would be optimal.

To tackle these challenges, one potential solution involves leveraging an architecture better suited to the problem, such as attention networks. In the following section, we introduce a neural network based on the attention mechanism to address the JPA problem, which we will henceforth refer to as "JPANet" (Jet-Parton Attention Net) that afterward will be employed to tackle the full JPA problem.

### 7.1.2 Full Jet-parton assignment

**JPA<sub>Net</sub>** JPA<sub>Net</sub> is an attention network adapted to the topology of the signal events. The network takes in two different inputs: one representing the leptonic W boson, and the other containing the features of all the jets. These inputs are channeled into separate subnetworks referred to as "heads" and are subsequently merged through a cross-attention mechanism.

- **W head:** The input tensor for the W head has a dimension of  $(N_{\text{events}}, 1, 7)$ , with 7 representing the number of features related to the leading lepton and the reconstructed neutrino originating in the leptonic W decay listed in Tab. ??.

The inputs are normalized with a Batch norm layer before being passed through a feed-forward network consisting of two layers, each containing 80 units. The resulting output from the W head constitutes the "W embedding."

- **Jet head:** The input tensor for the Jet head has dimensions of  $(N_{\text{events}}, 7, 5)$ , where the second dimension corresponds to the various jets and the third dimension pertains to distinct features. These input features encompass the kinematic properties of each jet, alongside its CvB score and the invariant mass of the combined jet-lepton-neutrino system.

Because in approximately 97% of events, all the jets matching the four quark partons can be found among the  $p_T$  leading seven jets with the highest transverse momentum, we opted to limit our selection to just seven jets (Fig. ??).

These selected inputs are first passed through a batch normalization layer <sup>1</sup> and then through a feed-forward neural network, which consists of two layers, each containing 80 units. The output from the first MLP serves as the input for a residual network, comprising a self-attention layer and a layer normalization layer. Finally, the output from the self-attention jet ResNet is directed into another MLP, which consists of three layers, each having 80 units. The output of the jet head constitutes the "Jet embedding."

After creating the W and the Jet embedding, we exploit a cross-attention layer to combine all the information of the events to assign an attention score to each jet in the jet sequence. This process involves concatenating the W embedding and jet embedding sequences and employing them as the Key and Value components in the scaled dot product, with the Query derived from the jet embedding sequence.

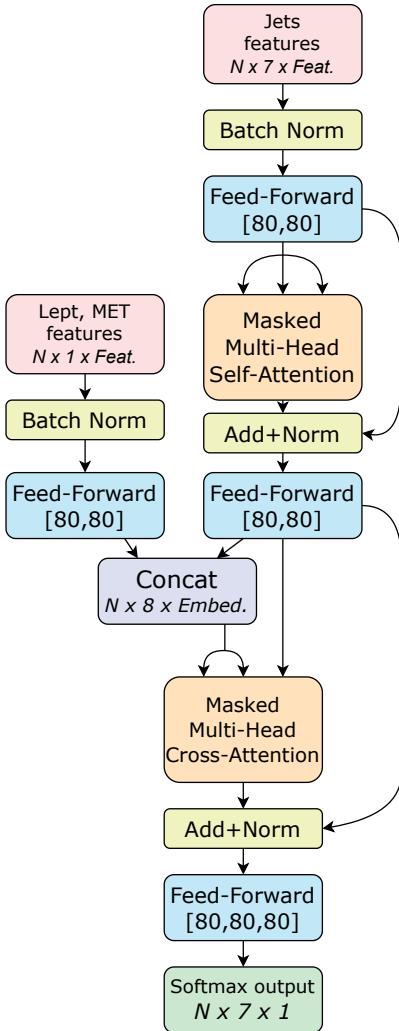
Feature	Description
$p_T^\ell$	Lepton $p_T$
$\phi^\ell$	Lepton $\phi$
$\eta^\ell$	Lepton $\eta$
$p_T^\nu$	Neutrino $p_T$
$\phi^\nu$	Neutrino $\phi$
$\eta^\nu$	Neutrino $\eta$
$m_W$	Leptonic W mass

Table 7.6: W head features.

Feature	Description
$p_T^j$	Jet $p_T$
$\phi^j$	Jet $\phi$
$\eta^j$	Jet $\eta$
$m_{t_L}$	Lept. top mass
Jet_CvB	DEEPJET CvB

Table 7.7: Jet head features.

<sup>1</sup>Since  $m_{t_L}$  can have significant outliers, the variable is first transformed with the function  $\log(1 + m_{t_L})$



**Figure 7.3:** JPA Net.

The output of the cross-attention layer is added to the jet embedding residuals, normalized, and subsequently passed into the final Multi-Layer Perceptron (MLP), which consists of three layers, each containing 80 units. The output layer is equipped with a softmax function, ensuring that the outputs represent scores for each element in the jet sequence that are normalized to sum up to 1.

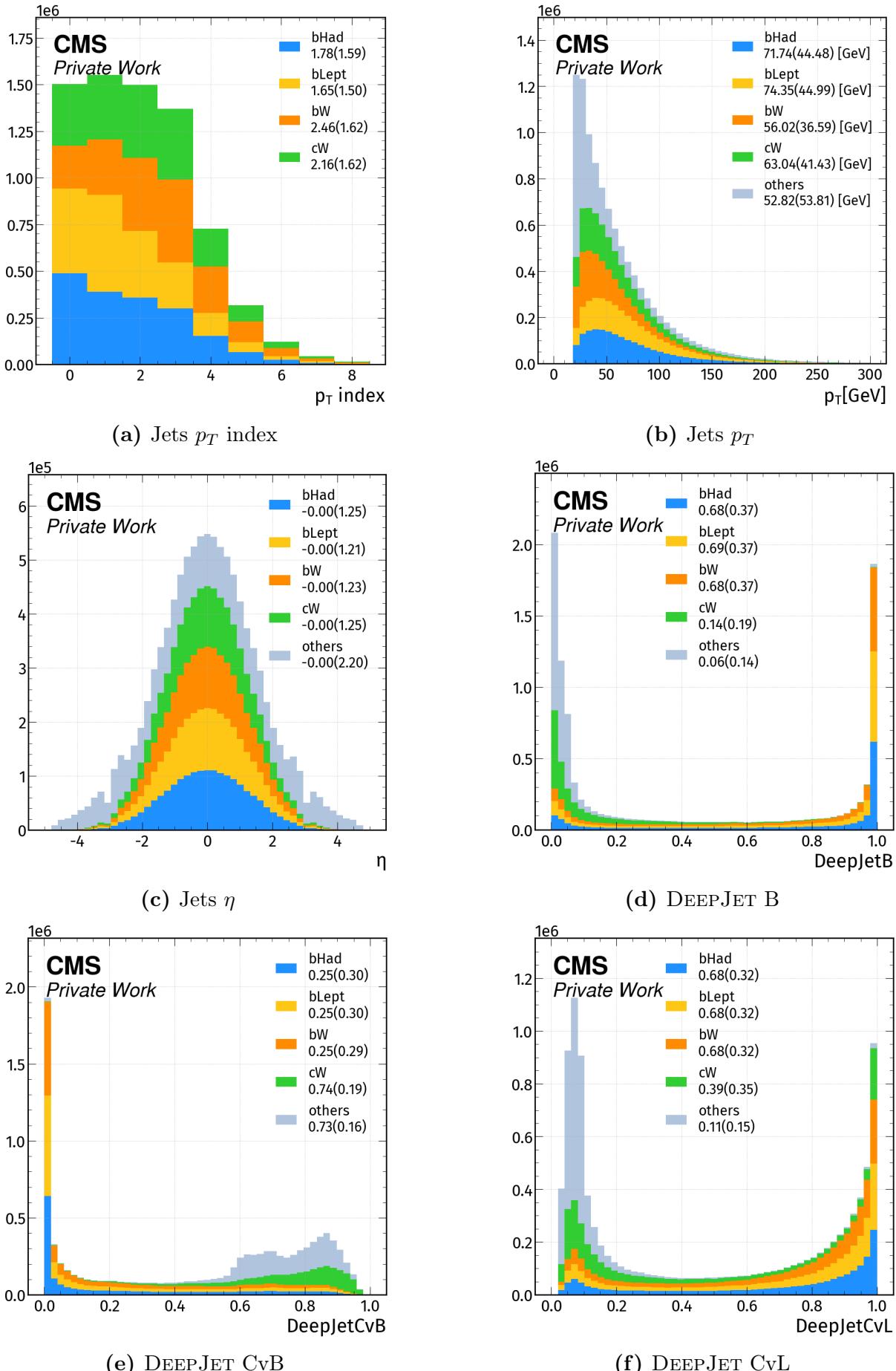
The network employs the Swish activation function and includes a 1.2% dropout probability to mitigate overfitting. Additionally, it utilizes a batch size of 20,000 and the RAdam optimizer, which minimizes the cross-entropy loss with a learning rate of  $10^{-3}$ .

Since the networks took 3 hours to train on an Nvidia A100 (40GB) GPU shared with the entire CERN community, a comprehensive hyperparameter optimization strategy could not be implemented. As a result, the choices made are based on a small number of trial runs. Thanks to its architecture, JPA effectively addresses all the issues that an event-wise MLP would encounter. The network inherently maintains permutation invariance for physics objects and the absent jets are handled just by forcing their respective attention scores to zero.

**Leptonic bJet Performance** On both the training and validation datasets, JPA Net successfully matched the correct jet to the leptonic b quark parton in 82% of the events. This high accuracy represents a substantial enhancement compared to all the assessed jet-wise models, making JPA Net the preferred choice for carrying out the complete JPA task.

**Full Jet-parton assignment** To carry out the full JPA task using JPA Net, a modification is needed primarily in the output layer. The chosen approach involves generating an output, denoted as  $A_{jp}$ , for each event, with a shape of  $(7, 4)$ , such that  $\sum_j A_{jp} = 1 \forall p$ . In other words, to each quark parton, are assigned one score per jet that sum up to one. Additionally, since now we are looking for jets of different flavors, the DEEPJET CvL score is introduced in the jet features. The optimal combination of jets is determined by maximizing the product of scores across all possible combinations of assigning the jets to the four quark partons. Despite a significant number of potential combinations ( $7!/(7 - 4)! = 840$  in total), JPA Net demonstrated the ability to accurately identify the correct jet-parton assignments in 54% of the events in both the training and validation datasets.

JPA Net accuracy	
Lept. b JPA	82%
Full JPA	54%



**Figure 7.4:** Stacked plots of observables of jets that match the four quark partons.

---

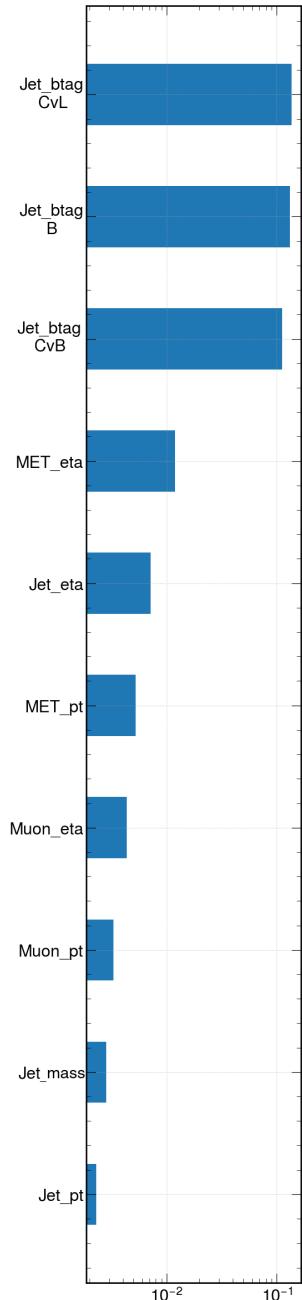
*Chapter 8*

---

## Event classification and signal extraction

---

This chapter is focused on the signal-background discrimination. It will begin by demonstrating a basic one-dimensional feature ranking. Subsequently, JPANet will be adapted to tackle the signal-background discrimination challenge, and its score will be employed to execute a template fitting procedure on both the Muon and Electron channels, utilizing the Asimov dataset. Once the statistical uncertainties associated with the  $W \rightarrow cb$  rate are determined, we will also introduce and assess the systematic uncertainties and their impact on the signal rate extraction.



## 8.1 Unidimensional feature ranking

To begin, we can once again employ the metric  $d$  to evaluate and rank the observables that discriminate the signal from the background.

In this context,  $d$  is computed between the signal and the  $t\bar{t}$  semileptonic process, which is the primary source of background.

All observables are treated as flattened, meaning that all objects within the events are combined in the same histograms, disregarding the event's inherent structure.

The ranking process is carried out within the Muon channel following the initial preselection.

The results of this one-dimensional ranking are depicted in Fig. ??.

Notably, the observables with the highest ranking are those associated with b-tagging, exhibiting  $d$  values of approximately 0.1.

Other observables produce results that are about an order of magnitude smaller, and the latests are due mostly to statistical fluctuations.

**Figure 8.1:** Unidimensional feature ranking using the metric  $d$  between flattened signal observables and  $t\bar{t}$  semileptonic observables

## 8.2 Event classification

### SBA $\text{Net}$

Since we have already demonstrated the ability of JPANet to reconstruct and interpret the signal events, we can exploit the same model to tackle the signal-background discrimination task. To achieve this, we simply need to make some adjustments to JPANet, creating another network that will be denoted as "Signal-Background Attention Network" (SBA $\text{Net}$ ). The training and evaluation of this network will be conducted independently for the Muon and Electron channels.

The observables used as inputs are the kinematic variables ( $p_T, \eta, \phi$ ) of jets, leptons, and

the reconstructed neutrino, along with the bTag, CvB, and CvL scores of all jets. In addition, to provide an enriched representation of the events, all the invariant masses of physics object pairs were added as input to the network. The output is a score assigned to each event that is treated as the "probability" of the event to be a signal event.

The adjustments applied to JPANet to build SBANet are the following:

**Hyper-parameters:** After experimenting with various configurations, we found that the network required increased complexity. This entailed raising the number of units in each feed-forward layer to 128 and adjusting the dropout probability to 2%.

**Self-attention pooling:** The JPANet's output initially produces a 7x4 matrix, as it assigns scores to pairs of jet-partons. However, we need to derive a single score for the entire event. To accomplish this, we condensed the physics objects embedding sequence into a single dimension using a self-attention pooling layer. This layer assigns weights to each element of the sequence (i.e., the physics objects) based on attention scores and computes a weighted average across the sequence. Subsequently, we introduced three additional feed-forward layers.

**Event self-attention:** We aimed to have the self-attention pooling compute attention scores over all physics objects, not just the jets. Consequently, we transformed JPANet's cross-attention into a self-attention mechanism covering the concatenated sequence of jets and the leptonic W.

**Output layer** The output of SBANet is a 3-dimensional layer equipped with the softmax function. Three classes are defined: the signal, the  $t\bar{t}$  semileptonic process, and the  $t\bar{t}$  dileptonic process. This enables the network to better distinguish the signal from the two primary sources of background, even though, ultimately, we only consider the signal score.

**Secondary leptons head** To enhance SBANet's ability to identify di-leptonic events, we introduced an additional head. In the muon(electron) channel, this head comprises a sequence of three elements: the second leading muon(electron), the leading and the second leading electron(muon). The lepton features are limited to kinematic variables ( $p_T, \eta, \phi$ ), and the head is linked to a 2-layer feed-forward neural network. The sequence of secondary leptons' embeddings is then combined with the complete event sequence before the self-attention pooling layer.

### Invariant mass heads and pair-wise attention biases

The two self-attention layers in the network were modified to include pair-wise physics object features, such as invariant masses. Let's recall what is the scaled dot product:

$$\text{Softmax} \left( \frac{QK^T}{\sqrt{d_k}} \right) V \quad (8.1)$$

The product  $QK^T$  is a  $N \times N$  matrix where N is the length of the sequence.

Given that, we can include the invariant masses of couples of physics objects by just adding a term in the softmax function

$$\text{Softmax} \left( \frac{QK^T + \hat{M}}{\sqrt{d_k}} \right) V \quad (8.2)$$

where  $\hat{M}$  is a linear embedding of the matrix  $M$  that contains all the invariant masses

$$\begin{aligned} M_{ij} &= \sqrt{(p_i + p_j)^\mu (p_i + p_j)_\mu} && \text{if } i \neq j \\ M_{ij} &= m_i && \text{if } i = j \end{aligned} \quad (8.3)$$

The  $M$  matrix is symmetric and has  $N(N + 1)/2$  independent components.

The strategy to implement this in SBANet is adding another input head containing for each event a vector of length 36<sup>1</sup> with the square invariant masses of the couple of jets and the leptonic  $W$ .

$$\text{Mass head inputs} = \left[ \begin{array}{c} m_W, m_{Wj_1}, \dots, m_{Wj_7}, [m_{j_1}, m_{j_1j_2}, \dots, m_{j_6}, m_{j_6j_7}, m_{j_7}] \end{array} \right]$$

This vector contains all the independent components of the  $M$  matrix.

Given the presence of significant outliers in the invariant masses, they are transformed with the function  $\log(1 + m)$  and then normalized with a batch norm layer.

There are two attention layers in SBANet: the first is a self-attention layer that works only on the jet sequence, while the second works on the jet + leptonic  $W$  sequence, so we have to build two different mass embeddings.

For the first jet self-attention, only the mass inputs that are not related to the leptonic  $W$  boson are considered. They are 28 elements that are fed to a 4-layer feed-forward neural network of size [28,128,128,28]. This network that has the same number of neurons in the first and the last layer is called Autoencoder.

The 28 output elements of the autoencoder are used to build the full  $7 \times 7$   $M$  matrix that will be used as attention bias in the jet self-attention.

The same thing is done with the second event self-attention but considering all the mass head inputs and, in this case, the autoencoder size will be [36,128,128,36].

**Training and evaluation** The network is trained and evaluated separately in the Muon and Electron channels. The dataset is divided into 85% training and 15% validation and is illustrated in Tab. ??.<sup>2</sup>

To take into account the imbalance between the three classes, the cross-entropy loss function of SBANet is reweighted.

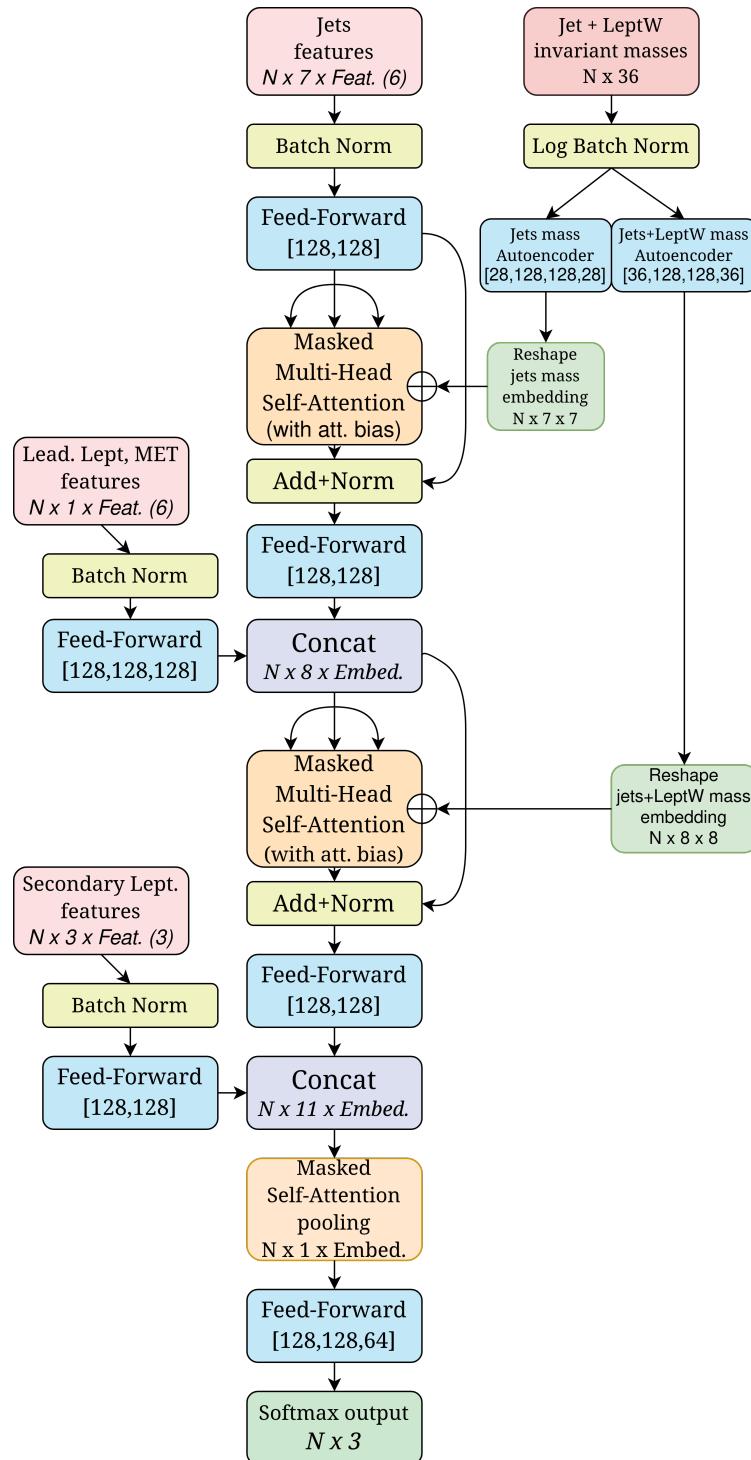
	Training	Validation
signal	$9.1 \cdot 10^6$	$1.6 \cdot 10^6$
TTSL	$1.0 \cdot 10^8$	$1.8 \cdot 10^7$
TTDL	$2.0 \cdot 10^7$	$3.5 \cdot 10^6$

**Table 8.1:** SBANet train and validation datasets before the Muon/Electron channel preselection.

<sup>1</sup>Because the final sequence is composed of 7 jets + the leptonic  $W$ , so  $8 * (8 + 1)/2 = 36$

<sup>2</sup>These events are not included in the effective MC luminosity shown in Tab. ???. These are additional events used only to train and validate the network

The loss functions are shown in Fig. ?? and don't show any sign of overfitting.



Head	Input features
Jets	$[p_T, \eta, \phi, \text{bTag}, \text{CvB}, \text{CvL}]$
LeptW	$[p_T^\ell, \eta^\ell, \phi^\ell, p_T^\nu, \eta^\nu, \phi^\nu]$
Additional Lept.	$[p_T, \eta, \phi]$
Masses	36 Jets+LeptW inv. masses

Act. function	Swish
Loss	Cross-Entropy
Optimizer	RAdam
Learn. rate	0.001
Batch size	20000
Dropout	0.02

Figure 8.2: SBANet schematics, inputs, and hyper-parameters.

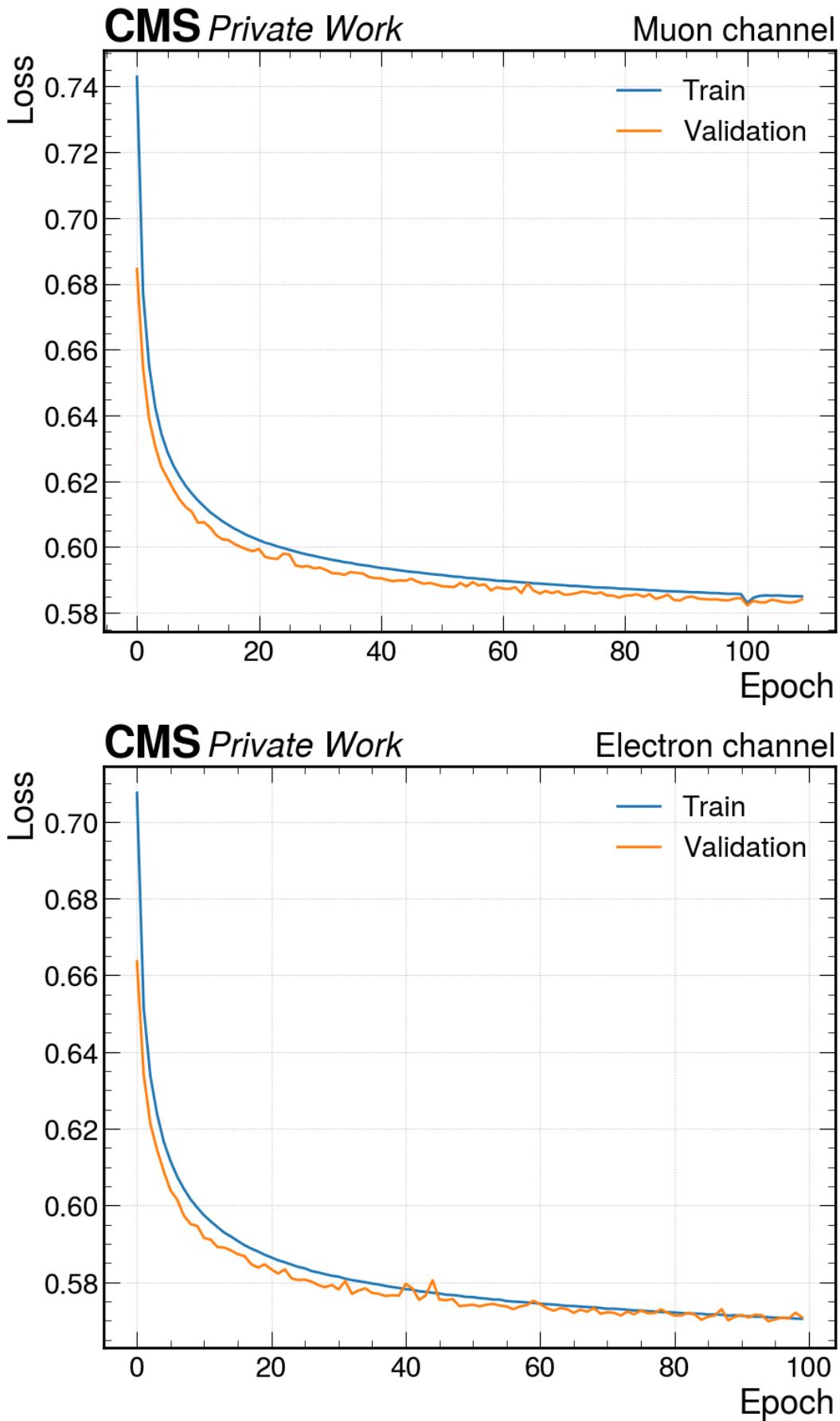
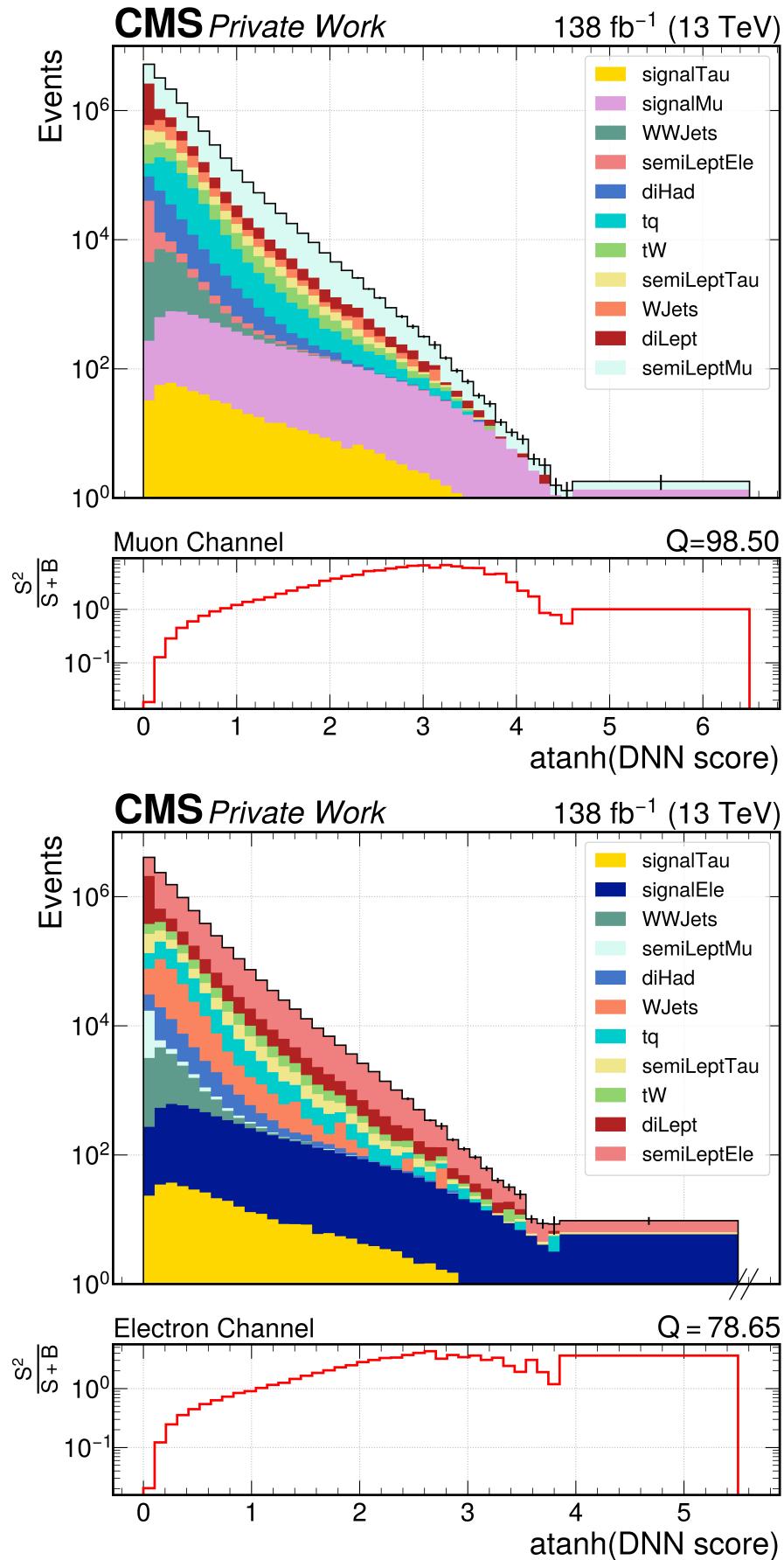


Figure 8.3: SBA-Net training and validation loss for the muon and electron channel.



**Figure 8.4:** Stacked distribution of the SBA-Net scores for each process and signal sensitivity ( $Q$ ) plot in the muon (top) and electron (bottom) channel.

## 8.3 $|V_{cb}|$ extraction

To extract  $|V_{cb}|^2$ , the SBA-Net score is employed to perform a template fit.

Channel	$Q$	$\frac{\sigma[ V_{cb} ^2]}{ V_{cb} ^2}$
Muon	98.50	10.1%
Electron	68.65	11.3%
Combined	177.15	7.5%

**Table 8.2:** Signal sensitivity and estimate of the statistical uncertainties on  $|V_{cb}|^2$  for each channel.

To expand the region with the highest signal purity, the scores are transformed with the inverse hyperbolic tangent function and then binned in such a way as to maximize the signal sensitivity, maintaining at least one event in each bin. Even before executing the fit, the statistical uncertainties on  $|V_{cb}|^2$  can be estimated through the signal sensitivity.

**Template fit** The statistical uncertainties on  $|V_{cb}|^2$  are derived using the profile likelihood method using the Asimov dataset, as described in Sec. ??.

The signal events in the histogram are normalized with the RunII luminosity

$$N_s = \mathcal{L}_{\text{RunII}}^I \times \epsilon_{\text{channel}} \times \sigma(t\bar{t}) \times \mathcal{BR}(W \rightarrow \ell\nu) \times (1 - \mathcal{BR}(W \rightarrow \ell\nu)) \times \frac{|V_{cb}|^2}{2}$$

where  $|V_{cb}|^2$  is the parameter of interest to fit, so the expectation value  $\lambda$  of the Poisson distribution in the likelihood relative to the b-th bin will be

$$\lambda^{(b)}(|V_{cb}|^2) = |V_{cb}|^2 \cdot \frac{N_s^{(b)}}{\widetilde{|V_{cb}|^2}} + \sum_p^{\text{bkgs}} N_p^{(b)}$$

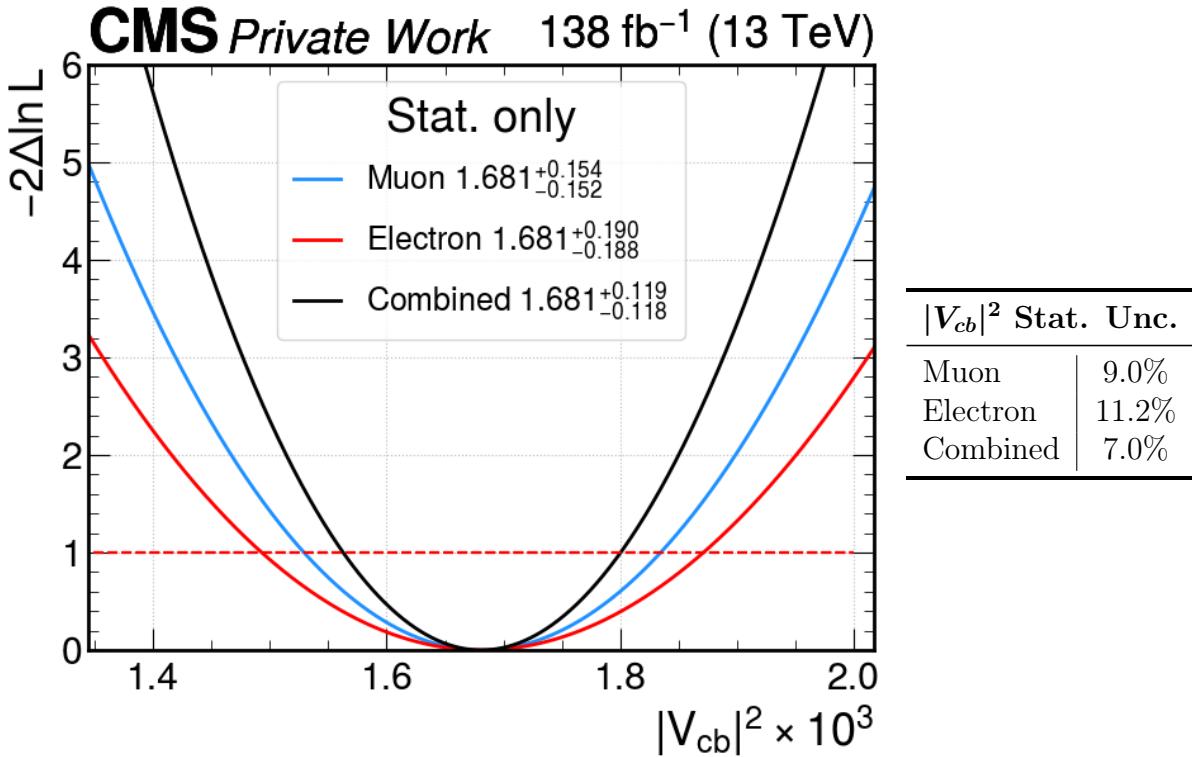
where  $\widetilde{|V_{cb}|^2}$  is the value used to normalize the histograms and  $N_p^{(b)}$  are the number of events for the process  $p$  in the b-th bin. The result of the fit is shown in Fig. ??

### 8.3.1 Systematic uncertainties

In this analysis, the most relevant systematic uncertainties are the following:

- **b/c-tagging uncertainties** The derivation of the b/c-tagging scale factors, described in Sec. ??, is subject to different systematical uncertainties [? ].
- **JES/JER** Uncertainties related to the jet energy scale and resolution corrections described in Sec. ??.
- **Luminosity** Uncertainties on the RunII luminosity: 1.6%
- **Channel acceptance** Uncertainties on the acceptance of each signal region: 1%(2%) Muon(Electron) channel.
- **t̄ cross-section** Uncertainties on the t̄ cross-section: 5%

The impact of the systematic uncertainties is evaluated using the profile likelihood method with normal priors for each nuisance. The impacts for each channel and the combined impacts are shown in Fig. ??, Fig. ??, Fig. ?? and are summarized in Tab. ??.



**Figure 8.5:** Profile likelihood without nuisances on the SBA-Net score templates in the muon channel, electron channel, and combined.

Uncertainties	Muon	Electron	Combined
Statistical	9.0%	11.2%	7.0%
b-tagging	9.3%	8.7%	8.8%
c-tagging	5.7%	4.1%	7.6%
JES/R	8.6%	11.5%	6.6%
Lumi+XSec+Acceptances	6.0%	7.5%	4.1%
Total	18%	21%	15%

**Table 8.3:** Impact of the systematic uncertainties and total uncertainties on the  $|V_{cb}|^2$  determination for each channel.

The uncertainties on  $|V_{cb}|$  are half of those on  $|V_{cb}|^2$ , so the expected relative uncertainty on  $|V_{cb}|$  is 7.8% (3.5% Stat. + 7% Syst.)

$$|V_{cb}| = 0.0410 \pm 0.0032$$

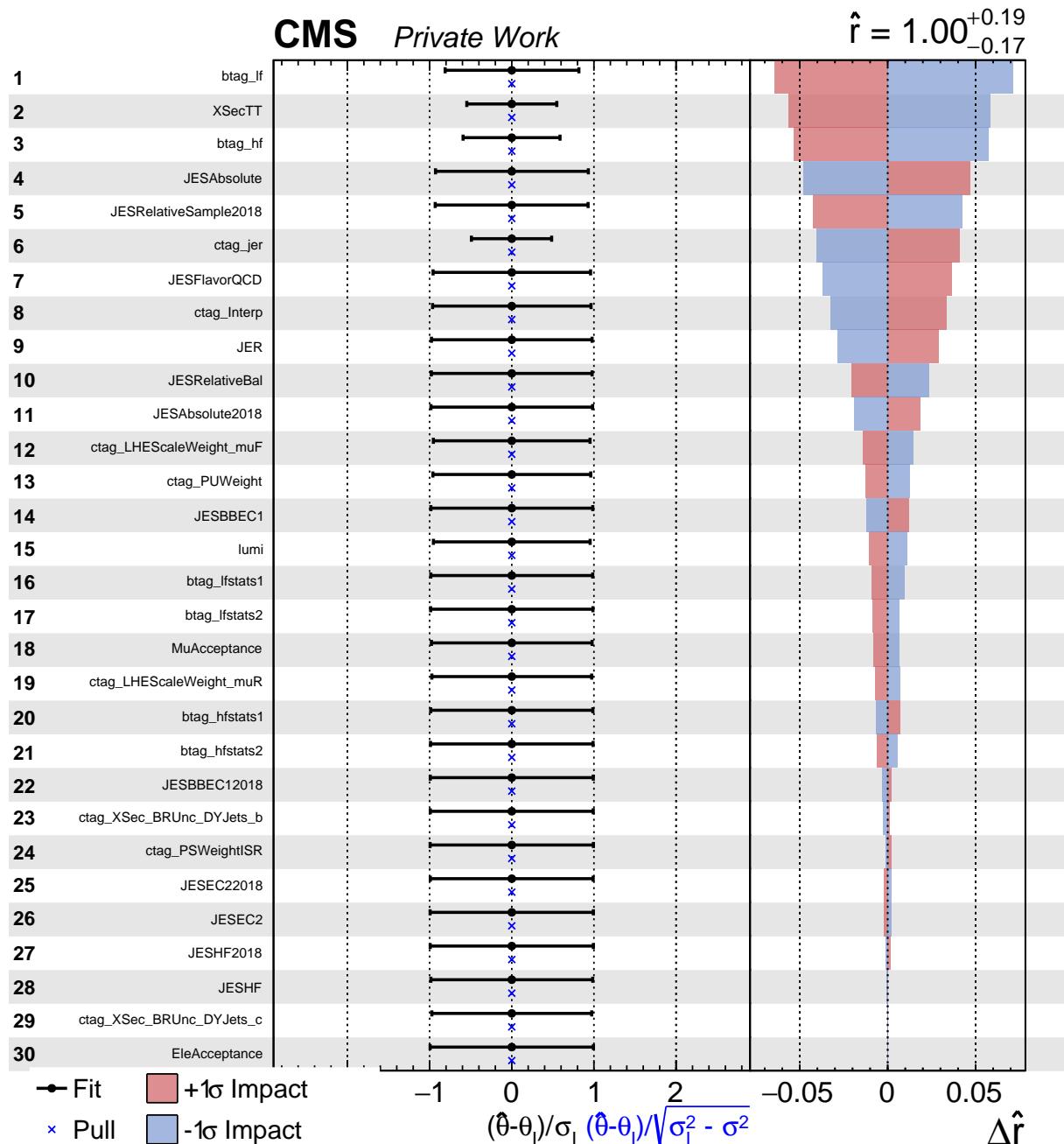
This relative uncertainty is comparable with the 7% difference shown by the  $|V_{cb}|$  puzzle. The systematical uncertainties can be further reduced by deriving  $|V_{cb}|$  from the branching fractions ratio instead of from the signal rate.

$$|V_{cb}|^2 = 2 \frac{\mathcal{BR}(W \rightarrow cb)}{\mathcal{BR}(W \rightarrow qq)}$$

Since different systematic variations are similar for all the semileptonic  $t\bar{t}$  processes,

regardless the flavor of the hadronic W decay, the ratio should allow some systematic uncertainties cancellations, improving the presented determination and providing a great opportunity to address the  $|V_{cb}|$  puzzle.

## Muons



**Figure 8.6:**  $|V_{cb}|^2$  systematic uncertainties impacts on the Muon channel.

# Electrons

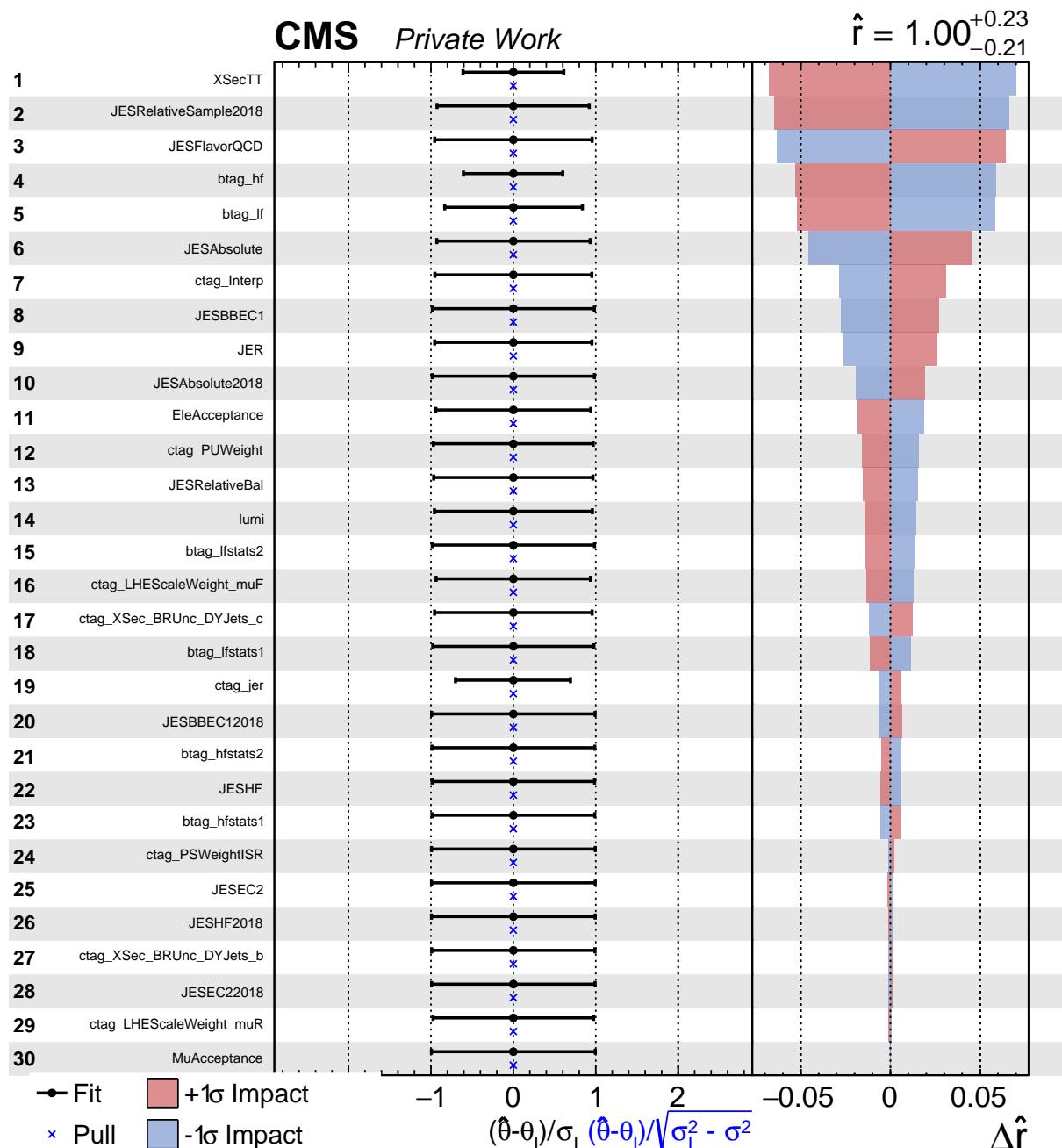
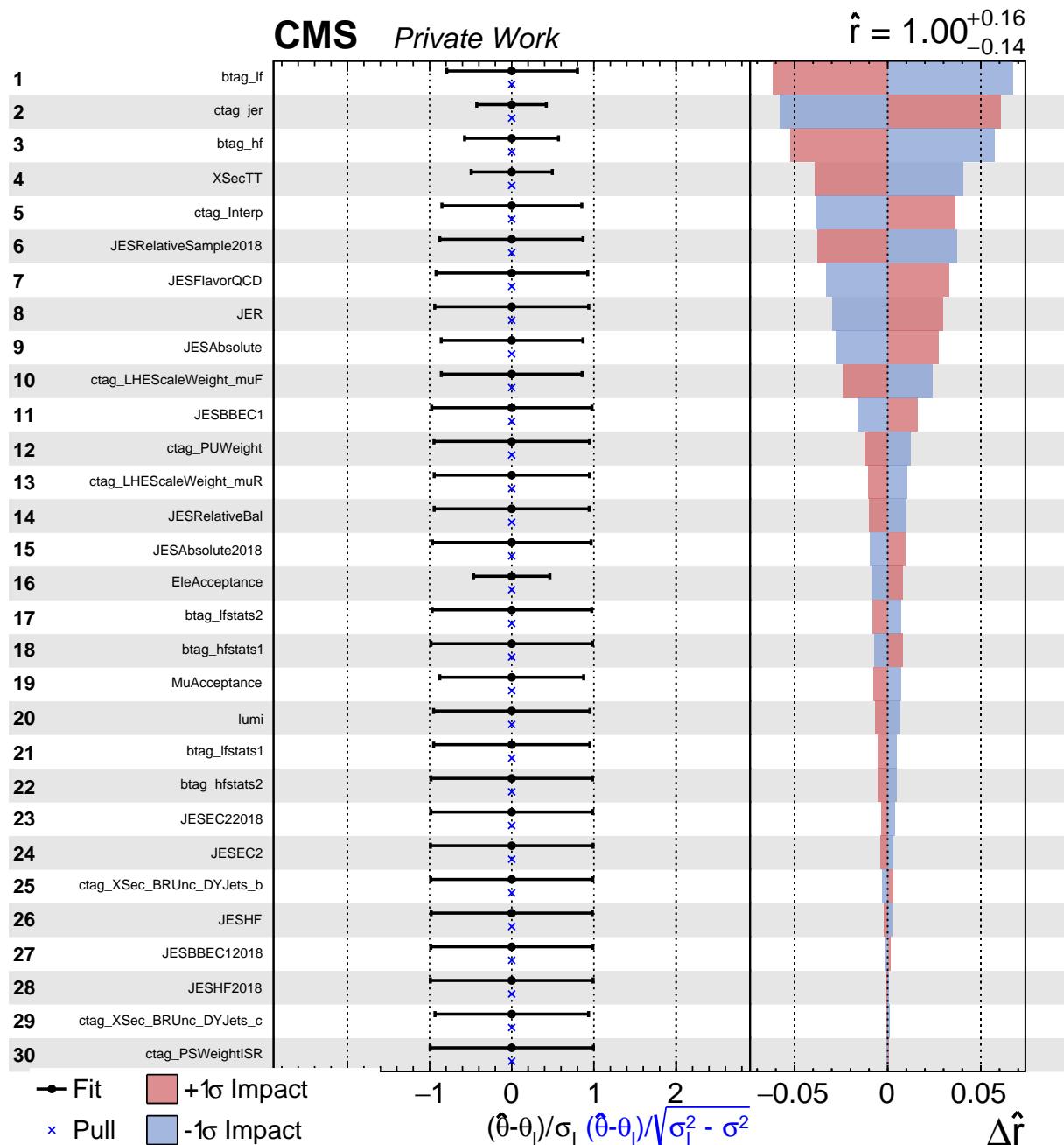


Figure 8.7:  $|V_{cb}|^2$  systematic uncertainties impacts on the Electron channel.

# Combined



**Figure 8.8:**  $|V_{cb}|^2$  systematic uncertainties impacts on both Electron and Muon channel combined.

## *Chapter 9*

---

# Conclusion

---

Currently,  $|V_{cb}|$  is the CKM element measured with the largest relative uncertainties beside  $|V_{ub}|$ . Furthermore, the determination performed exploiting the semileptonic decays of B mesons exhibits a 7% difference between the exclusive and the inclusive measurements, which leads to a  $3\sigma$  tension.

This thesis introduces a method to perform a direct measurement of  $|V_{cb}|$  at LHC through the hadronic decays of W bosons, exploiting the semileptonic decay of W pairs to avoid the QCD background that would make an inclusive determination an impossible task.

The presented analysis focused on events where the W pairs are produced by di-top decays and where a W boson decays into a muon or electron and the other into a cb quark pair.

To discriminate the signal from the background, cutting-edge machine-learning techniques were employed, in particular attention networks. A first attention network was built to address the signal interpretation optimization problem and then adapted to the signal-background discrimination task.

The network was trained separately on muon and electron signal regions and the output scores were employed to extract  $|V_{cb}|$  through a template fit.

The fit was performed using the profile likelihood method, adding additional nuisance parameters to evaluate the systematic uncertainties on the b/c tagging and the jet energy scale/resolution. Using the Asimov dataset, a combined relative statistical precision on  $|V_{cb}|$  of 7.8% (3.5% Stat. + 7% Syst.) was obtained, which is comparable with the 7% difference of the  $|V_{cb}|$  puzzle.

In the end, the presented method is a promising prospect to tackle the  $|V_{cb}|$  puzzle with a novel strategy, providing a new determination at the natural energy scale of the W boson which is also independent of non-perturbative QCD theoretical uncertainties. On top of that, while in this work  $|V_{cb}|$  was fitted directly as the parameter of interest, the determination could be further improved by deriving  $|V_{cb}|$  from the ratio of branching fractions  $|V_{cb}|^2 = 2\mathcal{BR}(W \rightarrow cb)/\mathcal{BR}(W \rightarrow qq)$ , since it would allow

cancellation in systematical uncertainties.

---

# References

---

- [1] E. Noether, “Invariante Variationsprobleme”, *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse* (1918) 235–257.
- [2] Burgard Carsten and Galbraith David, “Standard model of physics | TikZ example”. <https://texexample.net/tikz/examples/model-physics/>.
- [3] R. Feynman, “QED: The Strange Theory of Light and Matter”,.
- [4] A. Deur, S. J. Brodsky, and G. F. de Teramond, “The QCD Running Coupling”, *Progress in Particle and Nuclear Physics* **90** (4, 2016) 1–74, [doi:10.1016/j.ppnp.2016.04.003](https://doi.org/10.1016/j.ppnp.2016.04.003).
- [5] E. Fermi, “Tentativo di una Teoria Dei Raggi  $\beta$ ”, *Il Nuovo Cimento* **11** (1, 1934) 1–19, [doi:10.1007/BF02959820/METRICS](https://doi.org/10.1007/BF02959820).
- [6] M. Kobayashi and T. Maskawa, “CP-Violation in the Renormalizable Theory of Weak Interaction”, *Progress of Theoretical Physics* **49** (2, 1973) 652–657, [doi:10.1143/PTP.49.652](https://doi.org/10.1143/PTP.49.652).
- [7] S. Weinberg, “A Model of Leptons”, *Physical Review Letters* **19** (11, 1967) 1264, [doi:10.1103/PhysRevLett.19.1264](https://doi.org/10.1103/PhysRevLett.19.1264).
- [8] A. Salam and J. C. Ward, “Electromagnetic and weak interactions”, *Physics Letters* **13** (11, 1964) 168–171, [doi:10.1016/0031-9163\(64\)90711-5](https://doi.org/10.1016/0031-9163(64)90711-5).
- [9] S. L. Glashow, “Partial Symmetries of Weak Interactions”, *Nucl.Phys.* **22** (1961), no. 4, 579–588, [doi:10.1016/0029-5582\(61\)90469-2](https://doi.org/10.1016/0029-5582(61)90469-2).
- [10] P. W. Higgs, “Broken symmetries and the masses of gauge bosons”, *Physical Review Letters* **13** (10, 1964) 508–509, [doi:10.1103/PHYSREVLETT.13.508/FIGURE/1/THUMB](https://doi.org/10.1103/PHYSREVLETT.13.508/FIGURE/1/THUMB).
- [11] F. Englert and R. Brout, “Broken symmetry and the mass of gauge vector mesons”, *Physical Review Letters* **13** (8, 1964) 321–323, [doi:10.1103/PHYSREVLETT.13.321/FIGURE/1/THUMB](https://doi.org/10.1103/PHYSREVLETT.13.321/FIGURE/1/THUMB).
- [12] “Higgs Potential – TikZ.net”. <https://tikz.net/higgs-potential/>.
- [13] N. Cabibbo, “Unitary Symmetry and Leptonic Decays”, *Phys. Rev. Lett.* **10**

- (1963) 531, [doi:10.1103/PhysRevLett.10.531](https://doi.org/10.1103/PhysRevLett.10.531).
- [14] Particle Data Group, “Review of Particle Physics”, *PTEP* **2022** (2022) 083C01, [doi:10.1093/ptep/ptac097](https://doi.org/10.1093/ptep/ptac097).
- [15] J. Charles et al., “CP violation and the CKM matrix: assessing the impact of the asymmetric B factories”, *The European Physical Journal C - Particles and Fields* **2005** *41:1* **41** (2005), no. 1, 1–131, [doi:10.1140/EPJC/S2005-02169-1](https://doi.org/10.1140/EPJC/S2005-02169-1).
- [16] A. Höcker, H. Lacker, S. Laplace, and F. L. Diberder, “A new approach to a global fit of the CKM matrix”, *European Physical Journal C* **21** (6, 2001) 225–259, [doi:10.1007/S100520100729/METRICS](https://doi.org/10.1007/S100520100729/METRICS).
- [17] “CKMfitter”. [http://ckmfitter.in2p3.fr/www/results/plots\\_spring21/ckm\\_res\\_spring21.html](http://ckmfitter.in2p3.fr/www/results/plots_spring21/ckm_res_spring21.html).
- [18] F. Abe et al., “Observation of Top Quark Production in  $p\bar{p}$  Collisions with the Collider Detector at Fermilab”, *Physical Review Letters* **74** (4, 1995) 2626, [doi:10.1103/PhysRevLett.74.2626](https://doi.org/10.1103/PhysRevLett.74.2626).
- [19] I. Bigi, “The QCD Perspective on Lifetimes of Heavy-Flavour Hadrons”, *arXiv* (8, 1995) hep-ph/9508408, [doi:10.48550/ARXIV.HEP-PH/9508408](https://doi.org/10.48550/ARXIV.HEP-PH/9508408), [arXiv:9508408](https://arxiv.org/abs/9508408).
- [20] R. Feynman, “The behavior of hadron collisions at extreme energies”, *Conf.Proc.C* **690905** (1969) 237–258.
- [21] G. Altarelli and G. Parisi, “Asymptotic Freedom in Parton Language”, *Nucl.Phys.B* **126** (8, 1977) 298–318, [doi:10.1016/0550-3213\(77\)90384-4](https://doi.org/10.1016/0550-3213(77)90384-4).
- [22] V. N. Gribov and L. N. Lipatov, “Deep inelastic ep scattering in perturbation theory”, *Sov. J. Nucl. Phys.* **15** (1972) 438–450.
- [23] Y. L. Dokshitzer, “Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics.”, *Sov.Phys.JETP* **46** (1977) 641–653.
- [24] R. D. Ball et al., “The path to proton structure at 1% accuracy”, *The European Physical Journal C* **2022** *82:5* **82** (5, 2022) 1–119, [doi:10.1140/EPJC/S10052-022-10328-7](https://doi.org/10.1140/EPJC/S10052-022-10328-7).
- [25] M. Czakon and A. Mitov, “Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders”, *Comput.Phys.Commun.* **185** (2014), no. 11, 2930, [doi:10.1016/J.CPC.2014.06.021](https://doi.org/10.1016/J.CPC.2014.06.021).
- [26] “TtbarNNLO < LHCPhysics < TWiki”.  
<https://twiki.cern.ch/twiki/bin/view/LHCPhysics/TtbarNNLO>.
- [27] G. Ricciardi and M. Rotondo, “Determination of the Cabibbo-Kobayashi-Maskawa matrix element  $|V_{cb}|$ ”, *Journal of Physics G: Nuclear and Particle Physics* **47** (12, 2019) [doi:10.1088/1361-6471/ab9f01](https://doi.org/10.1088/1361-6471/ab9f01).
- [28] M. Bona et al., “New UTfit analysis of the unitarity triangle in the

- Cabibbo–Kobayashi–Maskawa scheme”, *Rendiconti Lincei. Scienze Fisiche e Naturali* 2023 34:1 **34** (2, 2023) 37–57, doi:10.1007/S12210-023-01137-5.
- [29] A. B. Smith, “Determination of  $|V_{cb}|$  Using Moments of Inclusive B Decay Spectra”, *Nuclear Physics B - Proceedings Supplements* **142** (5, 2005) 318–325, doi:10.1016/J.NUCLPHYSBPS.2005.01.055.
- [30] A. Alberti, P. Gambino, K. J. Healey, and S. Nandi, “The Inclusive Determination of  $|V_{cb}|$ ”, *Nuclear and Particle Physics Proceedings* **273-275** (4, 2016) 1325–1329, doi:10.1016/J.NUCLPHYSBPS.2015.09.212.
- [31] M. Bordone, B. Capdevila, and P. Gambino, “Three loop calculations and inclusive  $V_{cb}$ ”, *Physics Letters B* **822** (11, 2021) 136679, doi:10.1016/J.PHYSLETB.2021.136679.
- [32] S. Lu et al., “W boson mass and properties at LEP”, *Journal of Physics: Conference Series* **110** (5, 2008) 042019, doi:10.1088/1742-6596/110/4/042019.
- [33] R. Barate et al., “A direct measurement of  $|V_{cs}|$  in hadronic W decays using a charm tag”, *Physics Letters B* **465** (10, 1999) 349–362, doi:10.1016/S0370-2693(99)01088-6.
- [34] G. Abbiendi et al., “A measurement of the rate of charm production in W decays”, *Physics Letters B* **490** (9, 2000) 71–86, doi:10.1016/S0370-2693(00)00971-0.
- [35] P. Abreu et al., “Measurement of  $|V_{cs}|$  using W decays at LEP2”, *Physics Letters B* **439** (10, 1998) 209–224, doi:10.1016/S0370-2693(98)01061-2.
- [36] O. S. Brüning et al., “LHC Design Report”, *CERN, Geneva* **1** (2004) 548, doi:10.5170/CERN-2004-003-V-1.
- [37] S. Chatrchyan et al., “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”, *Physics Letters B* **716** (9, 2012) 30–61, doi:10.1016/J.PHYSLETB.2012.08.021.
- [38] G. Aad et al., “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”, *Physics Letters B* **716** (9, 2012) 1–29, doi:10.1016/J.PHYSLETB.2012.08.020.
- [39] “Panoramas”. <https://panoramas-outreach.cern.ch/index.html>.
- [40] CMS Collaboration, “Precision luminosity measurement in proton-proton collisions at  $\sqrt{s} = 13$  TeV in 2015 and 2016 at CMS”, *Eur. Phys. J. C* **81** (4, 2021) 800, doi:10.1140/epjc/s10052-021-09538-2.
- [41] “LumiPublicResults < CMSPublic < TWiki”.  
<https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults>.
- [42] O. Aberle et al., “submitter : High-Luminosity Large Hadron Collider (HL-LHC): Technical design report”, doi:10.23731/CYRM-2020-0010.
- [43] S. Chatrchyan et al., “The CMS experiment at the CERN LHC”, *Journal of*

- Instrumentation* **3** (8, 2008) S08004, doi:10.1088/1748-0221/3/08/S08004.
- [44] M. Spiropulu and S. Stapnes, “LHC’s ATLAS AND CMS DETECTORS”, <https://doi.org/10.1142/S0217751X08042341> **23** (1, 2012) 4081–4105, doi:10.1142/S0217751X08042341.
- [45] L. A. Coroller et al., “The CMS magnet project : Technical Design Report”, doi:10.17181/CERN.6ZU0.V4T9.
- [46] S. Chatrchyan et al., “Precise mapping of the magnetic field in the CMS barrel yoke using cosmic rays”, *Journal of Instrumentation* **5** (10, 2010) T03021, doi:10.1088/1748-0221/5/03/T03021.
- [47] P. Castaldi et al., “The CMS tracker system project : Technical Design Report”.
- [48] A. Dominguez et al., “CMS Technical Design Report for the Pixel Detector Upgrade”, technical report, 2012.
- [49] “DPGResultsTRK < CMSPublic < TWiki”. <https://twiki.cern.ch/twiki/bin/view/CMSPublic/DPGResultsTRK>.
- [50] W. Adam et al., “The CMS Phase-1 pixel detector upgrade”, *Journal of Instrumentation* **16** (2, 2021) P02027, doi:10.1088/1748-0221/16/02/P02027.
- [51] M. Friedl, “The CMS silicon strip tracker and its electronic readout”, <https://cds.cern.ch/record/783937>, 2001.
- [52] “TrackerMaterialBudgetplots < CMSPublic < TWiki”. <https://twiki.cern.ch/twiki/bin/view/CMSPublic/TrackerMaterialBudgetplots>.
- [53] H. Hofer ETH Zürich HansHofer et al., “The CMS electromagnetic calorimeter project : Technical Design Report”.
- [54] C. Biino, “The CMS Electromagnetic Calorimeter: overview, lessons learned during Run 1 and future projections”, *Journal of Physics: Conference Series* **587** (2, 2015) 012001, doi:10.1088/1742-6596/587/1/012001.
- [55] A. Benaglia, “The CMS ECAL performance with examples”, *Journal of Instrumentation* **9** (2, 2014) C02008, doi:10.1088/1748-0221/9/02/C02008.
- [56] P. Adzic et al., “Energy resolution of the barrel of the CMS electromagnetic calorimeter”, *JINST* **2** (4, 2007) P04004–P04004, doi:10.1088/1748-0221/2/04/P04004.
- [57] S. Chatrchyan et al., “Energy calibration and resolution of the CMS electromagnetic calorimeter in pp collisions at  $\sqrt{s} = 7$  TeV”, *Journal of Instrumentation* **8** (9, 2013) P09009, doi:10.1088/1748-0221/8/09/P09009.
- [58] “EcalDPGResultsCMSPDS2019005 < CMSPublic < TWiki”. <https://twiki.cern.ch/twiki/bin/view/CMSPublic/EcalDPGResultsCMSPDS2019005>.
- [59] D. Green, “The CMS hadron calorimeter project : Technical Design Report”, *Slac.Stanford.Edu* (1997) 31–34.

- [60] J. Anderson et al., “CMS Technical Design Report for the Phase 1 Upgrade of the Hadron Calorimeter”, *J. Mans* (9, 2012).
- [61] S. Abdullin et al., “The CMS Barrel Calorimeter Response to Particle Beams from 2 to 350 GeV/c”, <https://cds.cern.ch/record/1166316>, 6, 2008.
- [62] J. Layter et al., “The CMS muon project : Technical Design Report”,
- [63] D. Piccolo, “CMS muon detector and trigger performance”, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **628** (2, 2011) 142–145, [doi:10.1016/J.NIMA.2010.06.303](https://doi.org/10.1016/J.NIMA.2010.06.303).
- [64] “Muon Drift Tubes | CMS Experiment”.  
<https://cms.cern/detector/detecting-muons/muon-drift-tubes>.
- [65] “EndcapReshufflingRpc < Main < TWiki”.  
<https://twiki.cern.ch/twiki/bin/view/Main/EndcapReshufflingRpc>.
- [66] D. Acosta et al., “Efficiency of finding muon track trigger primitives in CMS cathode strip chambers”, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **592** (7, 2008) 26–37, [doi:10.1016/J.NIMA.2008.03.118](https://doi.org/10.1016/J.NIMA.2008.03.118).
- [67] G. Franzoni, “Dataset definition for CMS operations and physics analyses”, *Nuclear and Particle Physics Proceedings* **273-275** (4, 2016) 929–933, [doi:10.1016/J.NUCLPHYSBPS.2015.09.144](https://doi.org/10.1016/J.NUCLPHYSBPS.2015.09.144).
- [68] R. Musenich et al., “CMS TriDAS project : Technical Design Report, Volume 1: The Trigger Systems”, *Design* (2000), no. CMS-TDR-006-1, 599.
- [69] J. Brooke et al., “Hardware and Firmware for the CMS Global Calorimeter Trigger”, [doi:10.5170/CERN-2003-006.226](https://doi.org/10.5170/CERN-2003-006.226).
- [70] A. Rácz and P. Sphicas, “CMS The TriDAS Project : Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger”,
- [71] A. M. Sirunyan et al., “Particle-flow reconstruction and global event description with the CMS detector”, *Journal of Instrumentation* **12** (10, 2017) P10003, [doi:10.1088/1748-0221/12/10/P10003](https://doi.org/10.1088/1748-0221/12/10/P10003).
- [72] S. Chatrchyan et al., “Description and performance of track and primary-vertex reconstruction with the CMS tracker”, *Journal of Instrumentation* **9** (5, 2014) P10009, [doi:10.1088/1748-0221/9/10/P10009](https://doi.org/10.1088/1748-0221/9/10/P10009).
- [73] R. Frühwirth, “Application of Kalman filtering to track and vertex fitting”, *Nucl.Instrum.Meth.A* **262** (12, 1987) 444–450, [doi:10.1016/0168-9002\(87\)90887-4](https://doi.org/10.1016/0168-9002(87)90887-4).
- [74] X. Ai et al., “A GPU-Based Kalman Filter for Track Fitting”, *Computing and Software for Big Science* **5** (12, 2021) 1–16,

[doi:10.1007/S41781-021-00065-Z/FIGURES/17](https://doi.org/10.1007/S41781-021-00065-Z/FIGURES/17).

- [75] W. Adam, R. Frühwirth, A. Strandlie, and T. Todorov, “Reconstruction of electrons with the Gaussian sum filter in the CMS tracker at LHC”, *eConf C0303241* (9, 2003) [doi:10.1088/0954-3899/31/9/N01](https://doi.org/10.1088/0954-3899/31/9/N01).
- [76] H. Bethe and W. Heitler, “On the stopping of fast particles and on the creation of positive electrons”, *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* **146** (8, 1934) 83–112, [doi:10.1098/RSPA.1934.0140](https://doi.org/10.1098/RSPA.1934.0140).
- [77] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, “Optimization by Simulated Annealing”, *Science* **220** (5, 1983) 671–680, [doi:10.1126/SCIENCE.220.4598.671](https://doi.org/10.1126/SCIENCE.220.4598.671).
- [78] B. V. Dasarathy, “Nearest neighbor (NN) norms : nn pattern classification techniques”,.
- [79] A. M. Sirunyan et al., “Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC”, *Journal of Instrumentation* **16** (5, 2021) P05014, [doi:10.1088/1748-0221/16/05/P05014](https://doi.org/10.1088/1748-0221/16/05/P05014).
- [80] M. Cacciari, G. P. Salam, and G. Soyez, “The anti-kt jet clustering algorithm”, *Journal of High Energy Physics* **2008** (4, 2008) 063, [doi:10.1088/1126-6708/2008/04/063](https://doi.org/10.1088/1126-6708/2008/04/063).
- [81] V. Khachatryan et al., “Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV”, *Journal of Instrumentation* **12** (2, 2017) P02014, [doi:10.1088/1748-0221/12/02/P02014](https://doi.org/10.1088/1748-0221/12/02/P02014).
- [82] E. Bols et al., “Jet flavour classification using DeepJet”, *Journal of Instrumentation* **15** (12, 2020) P12012, [doi:10.1088/1748-0221/15/12/P12012](https://doi.org/10.1088/1748-0221/15/12/P12012).
- [83] K. O’Shea and R. Nash, “An Introduction to Convolutional Neural Networks”, *International Journal for Research in Applied Science and Engineering Technology* **10** (11, 2015) 943–947, [doi:10.22214/ijraset.2022.47789](https://doi.org/10.22214/ijraset.2022.47789).
- [84] A. Sherstinsky, “Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network”, *Physica D: Nonlinear Phenomena* **404** (8, 2018) [doi:10.1016/j.physd.2019.132306](https://doi.org/10.1016/j.physd.2019.132306).
- [85] J. Alwall et al., “MadGraph 5 : Going Beyond”, *JHEP* **06** (2011), no. 6, 128, [doi:10.1007/JHEP06\(2011\)128](https://doi.org/10.1007/JHEP06(2011)128).
- [86] S. Alioli, P. Nason, C. Oleari, and E. Re, “A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX”, *JHEP* **06** (2010), no. 6, 043, [doi:10.1007/JHEP06\(2010\)043](https://doi.org/10.1007/JHEP06(2010)043).
- [87] T. Sjöstrand, S. Mrenna, and P. Skands, “PYTHIA 6.4 Physics and Manual”, *JHEP* **05** (2006), no. 5, 026, [doi:10.1088/1126-6708/2006/05/026](https://doi.org/10.1088/1126-6708/2006/05/026).

- [88] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjöstrand, “Parton fragmentation and string dynamics”, *Physics Reports* **97** (7, 1983) 31–145, [doi:10.1016/0370-1573\(83\)90080-7](https://doi.org/10.1016/0370-1573(83)90080-7).
- [89] S. Agostinelli et al., “GEANT4—a simulation toolkit”, *Nucl.Instrum.Meth.A* **506** (7, 2003) 250–303, [doi:10.1016/S0168-9002\(03\)01368-8](https://doi.org/10.1016/S0168-9002(03)01368-8).
- [90] S. Sekmen, “Recent Developments in CMS Fast Simulation”, *PoS ICHEP2016* (2016) 181, [doi:10.22323/1.282.0181](https://doi.org/10.22323/1.282.0181).
- [91] F. Vaselli, A. Rizzi, F. Cattafesta, and G. Cicconofri, “FlashSim prototype: an end-to-end fast simulation using Normalizing Flow”, <https://cds.cern.ch/record/2858890>, 4, 2023.
- [92] F. Rosenblatt, “The perceptron: A probabilistic model for information storage and organization in the brain”, *Psychological Review* **65** (11, 1958) 386–408, [doi:10.1037/H0042519](https://doi.org/10.1037/H0042519).
- [93] K. Hornik, M. Stinchcombe, and H. White, “Multilayer feedforward networks are universal approximators”, *Neural Networks* **2** (1, 1989) 359–366, [doi:10.1016/0893-6080\(89\)90020-8](https://doi.org/10.1016/0893-6080(89)90020-8).
- [94] D. P. Kingma and J. L. Ba, “Adam: A Method for Stochastic Optimization”, *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings* (12, 2014).
- [95] L. Liu et al., “On the Variance of the Adaptive Learning Rate and Beyond”, *8th International Conference on Learning Representations, ICLR 2020* (8, 2019).
- [96] T. Ho-Phuoc, “CIFAR10 to Compare Visual Recognition Performance between Deep Neural Networks and Humans”.
- [97] N. Srivastava et al., “Dropout: A Simple Way to Prevent Neural Networks from Overfitting”, *Journal of Machine Learning Research* **15** (2014), no. 56, 1929–1958.
- [98] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition”, *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2016-December* (12, 2015) 770–778, [doi:10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90).
- [99] A. M. Fred Agarap, “Deep Learning using Rectified Linear Units (ReLU)”,.
- [100] P. Ramachandran, B. Zoph, and Q. V. Le Google Brain, “Searching for Activation Functions”, *6th International Conference on Learning Representations, ICLR 2018 - Workshop Track Proceedings* (10, 2017).
- [101] “Normalizing your data (specifically, input and batch normalization).”. <https://www.jeremyjordan.me/batch-normalization/>.
- [102] S. Ioffe and C. Szegedy, “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift”, *32nd International Conference on*

- Machine Learning, ICML 2015* **1** (2, 2015) 448–456.
- [103] J. L. Ba, J. R. Kiros, and G. E. Hinton, “Layer Normalization”,
- [104] X. Glorot and Y. Bengio, “Understanding the difficulty of training deep feedforward neural networks”,  
<https://proceedings.mlr.press/v9/glorot10a.html>, 3, 2010.
- [105] K. He, X. Zhang, S. Ren, and J. Sun, “Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification”, 2015.
- [106] A. Vaswani et al., “Attention Is All You Need”, *Advances in Neural Information Processing Systems 2017-December* (6, 2017) 5999–6009.
- [107] P. Safari, M. India, and J. Hernando, “Self-attention encoding and pooling for speaker recognition”, *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH 2020-October* (8, 2020) 941–945, doi:10.21437/Interspeech.2020-1446.
- [108] V. Khachatryan et al., “Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV”, *The European Physical Journal C 2015 75:5* **75** (5, 2015) 1–50, doi:10.1140/EPJC/S10052-015-3351-7.
- [109] F. James, “MINUIT: Function Minimization and Error Analysis Reference Manual”, <http://cds.cern.ch/record/2296388>, 1998.
- [110] J. Conway, “Incorporating Nuisance Parameters in Likelihoods for Multisource Spectra”, doi:10.5170/CERN-2011-006.115.
- [111] E. T. Jaynes, “Probability Theory: The Logic of Science”, *Probability Theory* (4, 2003) doi:10.1017/CBO9780511790423.
- [112] R. Barlow and C. Beeston, “Fitting using finite Monte Carlo samples”, *Computer Physics Communications* **77** (10, 1993) 219–228,  
doi:10.1016/0010-4655(93)90005-W.
- [113] W. S. Cleveland, “Robust locally weighted regression and smoothing scatterplots”, *Journal of the American Statistical Association* **74** (1979), no. 368, 829–836,  
doi:10.1080/01621459.1979.10481038.
- [114] F. James, “Statistical methods in experimental physics, 2nd edition”, *Statistical Methods in Experimental Physics, 2nd Edition* (1, 2006) 1–346,  
doi:10.1142/6096.
- [115] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, “Asymptotic formulae for likelihood-based tests of new physics”, *The European Physical Journal C 2011 71:2* **71** (2, 2011) 1–19, doi:10.1140/EPJC/S10052-011-1554-0.
- [116] C. Bierlich et al., “A comprehensive guide to the physics and usage of PYTHIA 8.3”, *SciPost Physics Codebases* **8** (11, 2022) 008,

[doi:10.21468/SCIPOSTPHYSODEB.8/PDF](https://doi.org/10.21468/SCIPOSTPHYSODEB.8/PDF).

- [117] The NNPDF Collaboration et al., “Parton distributions from high-precision collider data”, *European Physical Journal C* **77** (6, 2017) [doi:10.1140/epjc/s10052-017-5199-5](https://doi.org/10.1140/epjc/s10052-017-5199-5).
- [118] M. L. Mangano, M. Moretti, F. Piccinini, and M. Treccani, “Matching matrix elements and shower evolution for top-quark production in hadronic collisions”, *Journal of High Energy Physics* **2007** (11, 2006) [doi:10.1088/1126-6708/2007/01/013](https://doi.org/10.1088/1126-6708/2007/01/013).
- [119] R. Frederix and S. Frixione, “Merging meets matching in MC@NLO”, *Journal of High Energy Physics 2012 2012:12* **2012** (12, 2012) 1–41, [doi:10.1007/JHEP12\(2012\)061](https://doi.org/10.1007/JHEP12(2012)061).
- [120] A. M. Sirunyan et al., “Extraction and validation of a new set of CMS pythia8 tunes from underlying-event measurements”, *The European Physical Journal C 2019 80:1* **80** (1, 2020) 1–47, [doi:10.1140/EPJC/S10052-019-7499-4](https://doi.org/10.1140/EPJC/S10052-019-7499-4).
- [121] Y. Liu et al., “The NanoAOD event data format in CMS”, *Journal of Physics: Conference Series* **1525** (4, 2020) 012038, [doi:10.1088/1742-6596/1525/1/012038](https://doi.org/10.1088/1742-6596/1525/1/012038).
- [122] “Muon identification and isolation efficiencies with 2017 and 2018 data”, <https://cds.cern.ch/record/2629364>, 7, 2018.
- [123] “Electron and Photon performance in CMS with the full 2017 data sample and additional 2016 highlights for the CALOR 2018 Conference”, <https://cds.cern.ch/record/2320638>, 5, 2018.
- [124] CMS Collaboration, “Pileup mitigation at CMS in 13 TeV data”, *Journal of Instrumentation* **15** (3, 2020) [doi:10.1088/1748-0221/15/09/P09018](https://doi.org/10.1088/1748-0221/15/09/P09018).
- [125] “Jet energy scale and resolution measurement with Run2 Legacy Data Collected by CMS at 13TeV”, <https://cds.cern.ch/record/2792322>, 9, 2021.
- [126] “B-tagging performance of the CMS Legacy dataset 2018.”, <https://cds.cern.ch/record/2759970>, 3, 2021.
- [127] R. A. Fisher, “The use of multiple measurements in taxonomic problems”, *Annals of Eugenics* **7** (9, 1936) 179–188, [doi:10.1111/J.1469-1809.1936.TB02137.X](https://doi.org/10.1111/J.1469-1809.1936.TB02137.X).
- [128] P. Cunningham and S. J. Delany, “k-Nearest Neighbour Classifiers: 2nd Edition (with Python examples)”, *ACM Computing Surveys* **54** (4, 2020) [doi:10.1145/3459665](https://doi.org/10.1145/3459665).
- [129] Y. Coadou, “Boosted decision trees”, *Artificial Intelligence for High Energy Physics* (6, 2022) 9–58, [doi:10.1142/9789811234033\\_0002](https://doi.org/10.1142/9789811234033_0002).
- [130] A. Sirunyan et al., “Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV”, *Journal of Instrumentation* **13** (may, 2018) P05011,

doi:10.1088/1748-0221/13/05/P05011.

