SAPIENZA
UNIVERSITÀ DI ROMA

Search for lepton flavor violation in $\tau \to 3\mu$ decay with the Run 3 data of the CMS experiment at LHC

Scuola di Dottorato in Scienze Astronomiche, Chimiche, Fisiche e Matematiche 'Vito Volterra'
Physics (XXXVIII cycle)

Chiara Basile
ID number 1807325

Advisors
Dott.ssa Chiara Ilaria Rovelli
Dott.ssa Livia Soffi

Academic Year 2024/2025

Thesis not yet defended

Search for lepton flavor violation in $\tau \to 3\mu$ decay with the Run 3 data of the CMS experiment at LHC

Sapienza University of Rome

This thesis has been typeset by LaTeX and the Sapthesis class.

Author's email: chiara.basile@uniroma1.it

Abstract

Lepton flavor conservation is one of the accidental symmetries of the Standard Model (SM) that has been proven to be non-fundamental following the observation of neutrino oscillations. Among the possible lepton flavor–violating (LFV) processes, the decay of the τ lepton into three muons ($\tau \to 3\mu$) provides a particularly sensitive probe of new physics scenarios involving for instance heavy mediators or an extended Higgs sector.
This thesis presents a search for the LFV decay $\tau \to 3\mu$ using proton–proton collision data, collected with the CMS detector at a center-of-mass energy of 13.6 TeV during 2022 and 2023, and corresponding to an integrated luminosity of 62.2 fb^{-1}. The analysis targets τ leptons produced in vector boson decays. Candidate events are reconstructed from three muons originating from a common vertex in the detector and are selected using multivariate techniques. The search performs a bump hunt in the invariant mass distribution of the three muons, which would exhibit a narrow resonance around the τ lepton mass (1.777 GeV) in the presence of a LFV signal.
At the time of writing, the analysis is undergoing final internal review within the CMS Collaboration. Therefore, the results remain blind: data in the signal region are not used, and only the expected sensitivity, derived from simulated signal samples, is reported. The expected upper limit on the branching ratio is $\mathcal{B}r(\tau \to 3\mu) < 7.2 \times 10^{-8}$ at 90% confidence level, obtained from the combination of the 2022 and 2023 datasets.

The search for the $\tau \to 3\mu$ process is complemented by performance studies of the upgraded readout electronics for the CMS electromagnetic calorimeter barrel in preparation for the High-Luminosity LHC phase, which are documented in Appendix A. Beam test measurements demonstrate that the new front-end electronics achieves an energy resolution better than 0.6% and a time resolution below 30 ps for 50 GeV electrons, fully meeting the design specifications.

Contents

Introduction

The SM of particle physics is the most successful and precise theory currently available to describe the properties and interactions of elementary particles. Its formulation is based on quantum field theory, and it is constructed as a gauge theory invariant under the symmetry group $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$. Symmetry principles play a central role in the SM, as they determine which interactions among fundamental particles are allowed and govern their dynamics. The remarkable predictive power and internal consistency of the SM has been confirmed through decades of experimental tests with increasing precision. Nonetheless, there is experimental evidence that do not find an explanation within the theory. These observations provide direct evidence that the SM is an effective low-energy approximation of a more fundamental theory yet to be discovered. Among the open questions of the standard model is the inclusion of the neutrino oscillations, which have been observed experimentally. Furthermore, neutrino oscillation experiments have conclusively demonstrated that neutrinos have non-zero masses and that lepton flavor is not conserved in the neutral lepton sector, both phenomena that cannot be accommodated within the minimal SM.

In the charged lepton sector, lepton flavor violation (LFV) processes have not yet been observed. In the Standard Model extended to include neutrino masses, predicted branching fractions are many orders of magnitude below the current experimental sensitivity. The observation of LFV in charged lepton decays would therefore constitute unambiguous evidence of physics beyond the Standard Model. Among the possible LFV decay modes, the process $\tau \to 3\mu$ is particularly compelling experimentally. It features a clean experimental signature with three muons in the final state, whose invariant mass would peak at the τ lepton mass in the presence of a signal. A wide class of new physics models—such as supersymmetric extensions, heavy neutrino mediators, or multi-Higgs scenarios—predicts branching fractions for $\tau \to 3\mu$ that are within the reach of current experiments.
The $\tau \to 3\mu$ decay has been searched for by several experiments to date at both electron–positron and proton–proton colliders. The most stringent limit on the branching fraction, $\mathcal{B}r(\tau \to 3\mu) < 2.1{\times}10^{-8}$ at 90% confidence level, has been set by the Belle II experiment.
In this work, a search for the $\tau \to 3\mu$ decay is presented using proton–proton collision data collected by the Compact Muon Solenoid (CMS) experiment at the Large Hadron Collider (LHC). The dataset corresponds to an integrated luminosity of 62.2 fb^{-1}, recorded at a center-of-mass energy of 13.6 TeV during the 2022 and 2023 data-taking periods of Run 3. In proton–proton collisions, τ leptons are primarily

produced in decays of heavy-flavored mesons or electroweak bosons. The analysis presented here specifically targets τ leptons originating from W boson decays.
The analysis is designed as a search for a localized excess over a smooth background in the invariant mass distribution of three muons. Possible $\tau \to 3\mu$ candidates are reconstructed from events selected by the CMS trigger system. Offline, three muons originating from a common vertex are combined to form $\tau \to 3\mu$ candidates. The event missing transverse energy is combined with the reconstructed τ lepton kinematics to infer the kinematics of the associated W boson in the transverse plane. A multivariate selection is employed to maximize the separation between potential signal and background processes, exploiting a set of kinematic and topological observables that characterize the $W \to \tau(3\mu)\nu$ candidates. The expected LFV signal is modeled using simulated samples, with yield normalized to the integrated luminosity of the data. The background contribution is determined primarily through a data-driven approach. The analysis is performed blind, meaning that the data in the signal region are not examined until all algorithms and selection procedures have been finalized, in order to avoid potential bias in the search. Although the analysis is optimized to target τ leptons originating from W boson decays, the selection strategy retains significant acceptance for $\tau \to 3\mu$ decays where the τ is produced in Z boson decays. Therefore, the final result accounts for the additional contribution from $Z \to \tau(3\mu)\tau$ processes.
The exclusion limits on $\mathcal{B}r(\tau \to 3\mu)$ are derived from a profile likelihood fit performed simultaneously across all analysis categories. Currently, the analysis is under review in the CMS Collaboration and data in the signal region remain blind; only the expected exclusion limits are therefore reported.

This work is organized as follows. Chapter 1 introduces the theoretical framework of the Standard Model and the experimental context of the $\tau \to 3\mu$ search. In Chapter 2 the LHC and the CMS experiment are detailed providing the experimental context for this specific work. Chapter 3 describes the reconstruction of $\tau \to 3\mu$ candidates and gives the definition of the relevant observables. The sources of background and the strategies designed for their suppression are presented in Chapter 4, while Chapter 5 discusses the acceptance of $Z \to \tau(3\mu)\tau$ events and the modeling of both signal and background components. Finally, Chapter 6 details the statistical treatment based on a profiled likelihood fit and presents the resulting exclusion limits.

While the core of the thesis describes the search for the $\tau \to 3\mu$ process, Appendix A, describes the work performed during my PhD in the context of the CMS detector upgrade for the High-Luminosity LHC (HL-LHC) phase. It focuses on the characterization of the upgraded readout electronics for the CMS electromagnetic calorimeter. Dedicated test-beam campaigns were conducted to evaluate the energy and timing performance of the new front-end electronics. The results demonstrate an energy resolution better than 0.6% and a time resolution below 30 ps for 50 GeV electrons, meeting the design specifications required for operation at the HL-LHC.

Chapter 1

Lepton flavor theory in the Standard Model and beyond

This chapter briefly describes the theoretical framework of the SM of particle physics and its extensions relevant to lepton flavor physics. The theory describes the building blocks of our Universe, the elementary particles, and the dynamics of their interaction within the coherent and self-consistent framework of the quantum field theory. The SM currently provides the most precise and solid description of particle physics, nevertheless it leaves several fundamental questions unanswered. These open issues motivate the exploration of physics beyond the Standard Model, including phenomena like lepton flavor violation, which provide a powerful window into possible new interactions and particles. Section 1.1 introduces the fundamental building blocks of matter, their interactions, and the Higgs mechanisms responsible for the generation of particle masses. The discussion then focuses on the flavor sector, highlighting how fermion mixing arises and how flavor-changing phenomena are incorporated in the SM. In Section 1.2 the role of accidental symmetries and their breaking is discussed, providing the theoretical motivation for searches of charged lepton flavor violation, such as the $\tau \to 3\mu$ decay. Finally, in Section 1.3 the current experimental scenario for the searches of the lepton flavor violating $\tau \to 3\mu$ process is presented, including the latest achievement by the CMS experiment.
The following discussion set the stage for the experimental work that is the main topic of this thesis.

1.1 The Standard Model of Particle Physics

The SM of particle physics is a gauge theory describing the elementary particles and their interactions under three of the four fundamental forces of Nature: the strong, electromagnetic and weak interactions. The first step towards the mathematical formulation of the SM theory was Sheldon Glashow's unification of the electromagnetic and weak interactions [1]. Several years later Steven Weinberg [2] and Abdus Salam [3] incorporated the Brout-Englert-Higgs [4; 5] mechanism into Glashow's electroweak interaction, giving to the theory its modern form. Throughout the

following years and up to now the SM has established itself as a thoroughly tested and astoundingly precise theory over countless experiments [6].

This theory describes the interactions between elementary particles within the framework of the quantum field theory making use of a limited number of fundamental half-integer spin fields, the *fermions*, and spin-1 fields, the *bosons*, whose quantization implies the existence of point-like particles. The properties of fermions and bosons are summarized in Figure 1.1. Fermion fields describe elementary particles while their interactions arise requiring the gauge invariance of the Lagrangian of the theory under the symmetry group given by

$$SU(3)_C \otimes SU(2)_L \otimes U(1)_Y. \tag{1.1}$$

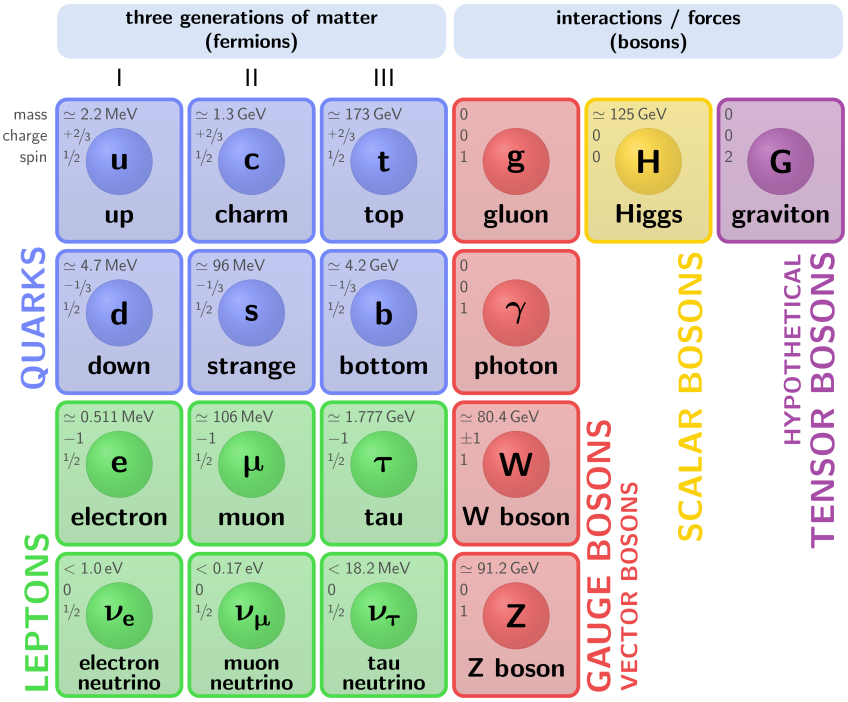


Figure 1.1 The Standard Model of elementary particles, with the three generations of fermions, the gauge bosons in the fourth column, and the Higgs boson in the fifth.

1.1.1 Fermions: the matter building blocks

Fermions are the constituent of matter and they are divided in two categories: leptons and quarks, based on their properties observed in Nature. Three generations of fermions are known for both leptons and quarks. Corresponding particles in different generations share the same quantum numbers and differ only for the mass so that fermions of the second (third) generation have larger mass than the corresponding fermions in the first (second) generation.

Leptons are electroweak interacting fermions of unitary charge. Each lepton generation, also referred to as *flavor*, includes a charged particle associated to a massless

neutrino (ν). The three charged leptons are the electron (e), historically linked to the electric properties of matter; the muon (μ), first observed in 1936 by C. D. Anderson and S. Neddermeyer in cosmic radiations and confirmed one year after in cloud chamber experiments by J. C. Street and E. C. Stevenson [7]; and the (τ) lepton, the heaviest of the three, which was the last to be discovered. The τ lepton was first detected in 1977 by the SLAC–LBL group at the SLAC electron–positron collider [8].
Quarks are strong and electroweak interacting fermions with fractional electric charge, known to be the elementary constituents of nucleons. There are six quark flavors divided into three generation : up (u) and down (d), charm (c) and strange (s), top (t) and bottom (b). Each flavor can assume three different color states, red, blue or green with color being the coupling for the strong force. Quarks do not appear free, but form bounded and colorless states called hadrons. Commonly observed hadrons are mesons, a bound state of a quark and an anti-quark, and baryons, a bound state of three quarks. In the mathematical formulation trough quantum field theory, each one of the twelve fermions of the Standard Model is associated to an anti-particle having same quantum numbers with opposite sign.

Each fermionic field is represented by a four-component Dirac spinor Ψ that, in the massless case, satisfies the equation of motion derived from the free propagation Lagrangian:

$$\mathcal{L}_0 = i\bar{\Psi}\gamma^\mu\partial_\mu\Psi, \tag{1.2}$$

where $\bar{\Psi}$ indicates the Dirac conjugate of $\Psi = \Psi^\dagger\gamma^0$ and the γ-matrices are the Dirac matrices satisfying the Clifford algebra:

$$\{\gamma^\mu, \gamma^\nu\} = 2\eta^{\mu\nu}, \tag{1.3}$$

with η the Minkowski metric. The γ matrices are written using the Pauli matrices (σ_i) of the 1/2 spinorial representation :

$$\gamma^0 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \gamma^i = \begin{pmatrix} 0 & \sigma^i \\ -\sigma^i & 0 \end{pmatrix}, \gamma^5 = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \text{ for } i = 1, 2, 3. \tag{1.4}$$

Projecting the Ψ on the left-handed and on the right-handed chiral components, the Dirac field is written in terms of the Weyl spinors (ψ_L and ψ_R):

$$\Psi = \psi_L + \psi_R, \tag{1.5}$$

$$\psi_L = P_L\Psi = \frac{1-\gamma^5}{2}\Psi, \text{ and} \tag{1.6}$$

$$\psi_R = P_R\Psi = \frac{1+\gamma^5}{2}\Psi. \tag{1.7}$$

For massless particles, chirality, which is a mathematical property of fields, coincides with helicity defined as the spin projection along the direction of motion of a particle. If the particle spin projection is antiparallel to the direction of motion, the particle is left-handed otherwise it is right-ended. The distinction between left- and right-handed fermions plays a crucial role in the SM theory formalization.

1.1.2 Interaction between fermions

The interactions between fermions are introduced in the Lagrangian of the theory imposing its local invariance under the symmetry group given by

$$SU(3)_C \otimes SU(2)_L \otimes U(1)_Y, \tag{1.8}$$

that immediately invokes Emmy Noether's theorem which states that every continuous symmetry of the Lagrangian of a physical system causes the conservation of a physical quantity [9]. The $SU(3)_C$ rules the strong interaction and, therefore give raise to the color quantum number while the electroweak symmetry group $SU(2)_L \otimes U(1)_Y$ is linked to the weak isospin (I_3) and to the hypercharge Y.
In 1956, Madame Wu experiment revealed that weak interactions intrinsically violates parity [10], coupling only left-handed fermions. For this reason fermions are divided into left-handed $SU(2)$ isospin doublet and right handed $SU(2)$ isospin singlets. Fermionic field for leptons and quarks in the SM are expressed, respectively as

$$\ell_{L,f} = \begin{pmatrix} \nu_f \\ \ell_f \end{pmatrix}_L \quad \ell_{R,f} \tag{1.9}$$

$$q_{L,f} = \begin{pmatrix} u_f^\alpha \\ d_f^\alpha \end{pmatrix}_L \quad u_{R,f}^\alpha, d_{R,f}^\alpha \tag{1.10}$$

where f index the flavor and α the color quantum number. The representation above encodes that left-handed fermions transforms as doublets with isospin I = 1/2, while right-handed particles transform as singlet with I=0 under $SU(2)_L$ electroweak symmetry. No right-handed neutrinos are included in the theory as no process involving right-handed neutral leptons has been ever observed up to this moment.
Quarks interact both strongly and electro-weakly, therefore they carry also the color charge α, and transform as triplet under $SU(3)_C$. The presence of colored and non-colored particles naturally splits the SM Lagrangian in the sum of two independent terms: one describing the strong interaction, also referred to as the quantum chromodynamics (QCD), and one describing the electroweak (EW) interactions:

$$\mathcal{L}_{SM} = \mathcal{L}_{QCD} + \mathcal{L}_{EW}. \tag{1.11}$$

To understand the mechanism allowing to describe the particle interaction theory in its Lagrangian formulation it is natural to start from the free Lagrangian of Eq. 1.21. That is invariant under a *global* $U(1)$ transformation, however it is no longer invariant under a time-space dependent transformation. A local gauge transformation acts on a Weyl spinors as

$$\psi \to e^{ig\theta(x)_a\lambda^a}\psi, \text{ and } \bar{\psi} \to \bar{\psi}e^{-ig\theta(x)_a\lambda^a}, \tag{1.12}$$

where g is a scalar coupling and $\theta(x)_a\lambda^a$ is the space-time dependent linear combination of the underlying symmetry group generators, thus making the transformation local.
The requirement of local invariance under a gauge transformation requires the

replacement of the partial derivative ∂ in the free Lagrangian with a covariant derivative operator $D_\mu = \partial_\mu - igG^a_\mu\lambda_a$ that introduces the vector field G^a_μ representing the mediator of the interaction and is described by bosonic fields. This operator is built applying the local transformation to the Lagrangian, and absorbing all the terms that would break the invariance of the theory [11].

The SM electroweak interaction is described imposing the Lagrangian invariance under the $SU(2)_I \otimes U(1)_Y$ symmetry group. Four vector fields are introduced: one boson B for $U(1)_Y$ and three bosons $\{W^{1,2,3}\}$ for $SU(2)_I$. The group transformation and the covariant derivative are expressed as

$$U = e^{i\beta(x)\frac{Y}{2}+i\alpha_i(x)\frac{\sigma^i}{2}}, \tag{1.13}$$

$$D_\mu = \partial_\mu - ig'YB_\mu - ig_W\frac{\sigma^i}{2}W^i_\mu, \tag{1.14}$$

where the scalar quantities g and g' are the generic interaction couplings, while Y and $I_i = \sigma^i/2$ indicate the weak hypercharge and the weak isospin matrices, respectively. The hypercharge and the isospin projection along the quantization axis (I_3) are related to the electric charge by the Gell-Mann-Nishijima formula:

$$\frac{Y}{2} = Q - I_3. \tag{1.15}$$

The theory was built in agreement with the experimental evidences that:

- electroweak interaction fully violates parity in the weak sector,
- charged current interactions (CCI), i.e. interactions with exchange of charge, involves fermions doublets whose charges differ by unity,
- CCI allows the interaction between quark of different families, i.e. allow flavor-mixing,
- no flavor-mixing is observed in neutral current interactions (NCI),
- electromagnetic interaction is independent from the fermions' chirality,

The symmetry invariant Lagrangian embeds the parity violation by coupling only the left-handed fermions to $SU(2)$ operator. Considering for simplicity a single family of fermions the gauge invariant Lagrangian is written as

$$\begin{aligned}
\mathcal{L}_{\text{EW}} = &-\frac{1}{4}B_{\mu\nu}B^{\mu\nu} - \frac{1}{4}W^j_{\nu\mu}W^{\nu\mu}_j \\
&+ \bar{\ell}_L\gamma^\mu\partial_\mu\ell_L + \bar{\ell}_R\gamma^\mu\partial_\mu\ell_R + \bar{q}_L\gamma^\mu\partial_\mu q_L + \bar{u}_R\gamma^\mu\partial_\mu u_R + \bar{d}_R\gamma^\mu\partial_\mu d_R \\
&+ \bar{\ell}_L\gamma^\mu(\partial_\mu - ig'YB_\mu - ig_W\frac{\sigma_i}{2}W^i_\mu)\ell_L + i\bar{\ell}_R\gamma^\mu(-ig'YB_\mu)\ell_R \\
&+ \bar{q}_L\gamma^\mu(\partial_\mu - ig'YB_\mu - ig_W\frac{\sigma_i}{2}W^i_\mu)q_L + i\bar{u}_R\gamma^\mu(-ig'YB_\mu)u_R + i\bar{d}_R\gamma^\mu(-ig'YB_\mu)d_R,
\end{aligned} \tag{1.16}$$

where the same notation in Eq. 1.9 is used for fermions. The first two lines of Eq. 1.16 represent the kinetic therms for the bosonic and fermionic field respectively.

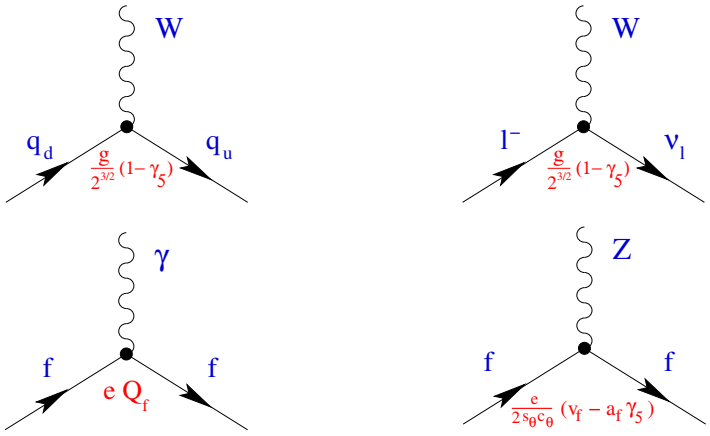


Figure 1.2 Charged (top) and neutral (bottom) current interactions in the SM [11].

The last two rows contains the interaction terms for the fermions and quark and no right-handed neutrino enters the theory. However, physical fields cannot be directly identified with B_μ and W^i_μ fields, but they arise from a linear combination of the gauge bosons. Expanding the matrix products in the interaction terms of Eq. 1.9 it is immediately evident that the charged current interaction is mediated by the linear combinations of W^1_μ and W^2_μ:

$$W^\pm_\mu = \frac{W^1_\mu \mp iW^2_\mu}{\sqrt{2}}, \tag{1.17}$$

where $W^\pm_\mu$ are the physically observed charged bosons interacting as described in Fig. 1.2. Similarly, the gauge fields W^3_μ and B_μ are not the observed Z^0 and γ bosons. Among other reasons, the photon has the same interaction with both fermion chiralities experimentally but since they have different hypercharges the photon field cannot be directly identified with the gauge boson B_μ. The gauge bosons W^3_μ and B_μ are mapped into the physical Z^0 and γ bosons via the following rotation, parametrized by the Weinberg angle θ_W:

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} B_\mu \\ W^3_\mu \end{pmatrix}, \tag{1.18}$$

After this redefinition the neutral current interaction dynamics is described by the diagrams in Fig. 1.2. The $W^\pm$ and Z^0 weak bosons are physical self-interacting particles, detected in experiments, while the photon field A_μ is the well known massless electromagnetic mediator.

Provided this mathematical approach, the $SU(3)_C$ covariant derivative for the strong interaction gets the following formulation:

$$(D_\mu)_{\alpha\beta} = \partial_\mu \delta_{\alpha\beta} - ig_s G^i_\mu \left(\frac{\lambda_i}{2}\right)_{\alpha\beta}, \tag{1.19}$$

where λ_i are the $SU(3)_C$ generators and g_s is the strong coupling constant. The quark Lagrangian modifies to include the associate covariant derivative and the

gluon-gluon interaction term $G^{\alpha}_{\mu\nu}G^{\mu\nu\ \alpha}$ which is also gauge invariant, hence allowed from the theory. The QCD Lagrangian can therefore be written as expressed below, giving raise to the interaction diagrams in Fig. 1.3,

$$\mathcal{L}_{\text{QCD}} = -\frac{1}{4}G^{\alpha}_{\mu\nu}G^{\mu\nu}_{\alpha} + i\sum_{f}\bar{q}^{\alpha}_{f}\gamma^{\mu}(D_{\mu})_{\alpha\beta}q^{\beta}_{f}. \tag{1.20}$$

Although the SM dynamics description presented above successfully describes the electroweak and strong interactions, it does not allow for the inclusion of mass terms for either the fermionic and bosonic fields as they would break the gauge invariance. Consequently a new mechanism should be introduced into the theory to allow for the introduction of the most evident property of matter: its mass. The required mechanism is the well known Higgs mechanism.

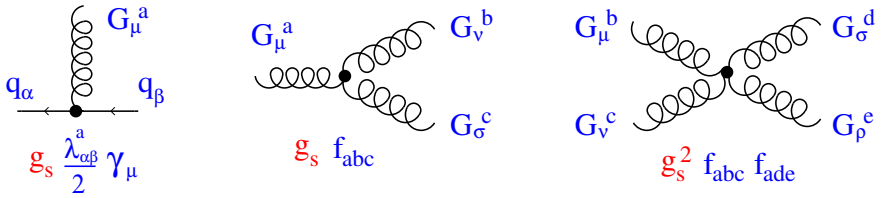


Figure 1.3 Interaction vertices of the QCD Lagrangian [11].

1.1.3 The origin of masses through the Higgs mechanism

The addition of the mass term modifies the Dirac fermion free Lagrangian of Eq. 1.21 as

$$\mathcal{L} = i\bar{\Psi}\gamma^{\mu}\partial_{\mu}\Psi - m\bar{\Psi}\Psi, \tag{1.21}$$

where the mass term mixes the right-handed and the left-handed representation and it not invariant under the SM gauge symmetry group.
The spontaneous symmetry breaking (SSB) mechanism allows for mass generation without breaking the gauge invariance of the SM Lagrangian. The model was published in 1964 by Brout and Englert [4], Higgs [5] and Guralnik, Hagen and Kibble [12], independently. It arbitrarily introduces a complex scalar field $\phi(x)$ with hypercharge 1 and with the structure of a $SU(2)_L$ doublet, namely the Higgs field:

$$\phi(x) = \begin{pmatrix}\phi^{+}(x)\\ \phi^{0}(x)\end{pmatrix}, \tag{1.22}$$

where ϕ^{+} and ϕ^{0} are the positive and neutral isospin component respectiveley. Imposing the invariance under the electroweak symmetry group the Lagrangian of the Higgs field is

$$\mathcal{L}_{H} = (D_{\mu}\phi)^{\dagger}(D^{\mu}\phi) - V(\phi) = (D_{\mu}\phi)^{\dagger}(D^{\mu}\phi) - \mu^{2}\phi^{\dagger}\phi + h(\phi^{\dagger}\phi)^{2}, \tag{1.23}$$

where $V(\phi)$ represents the Higgs potential and D_{μ} are the convariant derivative of Eq. 1.14. The physically interesting configuration of the potential has a degenerate

set of minima. This is realized if the potential is bounded from below ($h > 0$) and if μ^2 is negative. The degenerate set of minima is described by

$$|\phi_0|^2 = -\frac{\mu^2}{2h} := \frac{v^2}{2}, \quad V(\phi_0) = -\frac{h}{4}v^4, \tag{1.24}$$

with v being the vacuum expectation value for the Higgs field. The spontaneous symmetry breaking arises from the choice of a particular ground state, i.e. the vacuum state:

$$\phi_0 = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\v\end{pmatrix}, \tag{1.25}$$

and the Higgs field can be expressed as a perturbation around its minimum:

$$\phi(x) = \frac{1}{\sqrt{2}}e^{i\frac{\sigma_i}{2}\theta^i(x)}\begin{pmatrix}0\\v+H(x)\end{pmatrix}. \tag{1.26}$$

The non-physical fields θ can be removed exploiting the $SU(2)$ invariance of the Lagrangian, leaving only the scalar H field. The remaining field is called Higgs field:

$$\phi(x) = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\v+H(x)\end{pmatrix}. \tag{1.27}$$

The Higgs field is neutral therefore it does not couple with the photon. Consequently a residual symmetry under the $U(1)_Q$ rotation of the electric charge is still present. The spontaneous symmetry breaking pattern is, therefore $SU(2)_I \otimes U(1)Y \to U(1)_Q$ and, thanks to the Goldstone theorem, it follows that three out of four bosons acquire mass while the photon boson, associated to the residual symmetry, remains massless. Substituting the Higgs field representation of Eq. 1.27, expanding the covariant derivative and using the Weinberg rotation of Eq. 1.18, the Higgs Lagrangian is

$$\begin{aligned}\mathcal{L}_H = &\frac{1}{2}\partial_\mu H \partial^\mu H - v^2 h H^2 \\ &+ \frac{g_W^2 v^2}{8}(W_\mu^- W^{-\mu} + W_\mu^+ W^{+\mu}) \\ &+ \frac{v^2(g^2+g_W^2)}{8} Z_\mu Z^\mu.\end{aligned} \tag{1.28}$$

The first line of Eq. 1.28 represents the Higgs boson kinematics term and its mass ($m_H = v\sqrt{2h}$), while the second and third line represent the mass terms for the W and Z bosons ($m_Z = 1/2v\sqrt{g^2+g'^2}$ and $m_W = 1/2vg$). The interaction terms between gauge bosons and the Higgs boson, that enter the Lagrangian when considering higher orders of the Higgs potential $H(x)$, are omitted.

This procedure allows the electroweak theory to account for massive boson fields, while leaving the spinorial fields still massless. The massive nature of fermions is described by the inclusion of the Yukawa interaction term in the Lagrangian, i.e. the coupling between the scalar Higgs field and the fermions preserving the gauge

invariance of the theory. Considering only one fermion family the Yukawa terms get the following form:

$$\mathcal{L}_Y = c_1 \cdot \bar{q}_L \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} d_R + c_2 \cdot \bar{q}_L \begin{pmatrix} \phi^{0*} \\ \phi^- \end{pmatrix} u_R + c_3 \cdot \bar{\ell}_L \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} e_R + \text{h.c.} =$$

$$= -(1 + \frac{H}{v})(m_d \bar{d}d + m_u \bar{u}u + m_e \bar{e}e), \tag{1.29}$$

therefore the Yukawa coupling and the SSB mechanism generate mass terms for leptons and their couplings to the Higgs field.

1.1.4 Flavor mixing

It is known from experiments that there are six different quark flavor and three different lepton flavors. Let us include all the families in the Yukawa Lagrangian of Eq. 1.29 and apply the SSB:

$$\mathcal{L}_Y = \sum_{j,k} \bar{q}_{L,j} \left[c_{jk}^{(d)} \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} d_{R,k} + c_{jk}^{(u)} \cdot \begin{pmatrix} \phi^{0*} \\ \phi^- \end{pmatrix} u_{R,k} \right] + c_{jk}^{(\ell)} \cdot \bar{\ell}_{L,j} \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \ell_{R,k} + \text{h.c.}$$

$$= -(1 + \frac{H}{v})(\bar{d}_L M_d d_R + \bar{u}_L M_u u_R + \bar{\ell} M_\ell \ell_R) + \text{h.c., with} \tag{1.30}$$

$$(M_x)_{jk} = -c_{jk}^{(x)} \frac{v}{\sqrt{2}}, \tag{1.31}$$

where $c_{jk}^{(d)}$, $c_{jk}^{(u)}$ and $c_{jk}^{(\ell)}$ are arbitrary coupling absorbed in the mass matrices which are not diagonal. Following the principle that any terms that preserve the symmetry of the theory should enter the Lagrangian, we introduced mixing coefficient between the flavor eigenstates d, u and ℓ. Physical fields that are mass eigenstates are obtained diagonalizing the mass matrices, thus introducing the rotations:

$$\begin{aligned} u_L &\to V_L^u u_L, \quad & u_R &\to V_R^u u_R, \\ d_L &\to V_L^d d_L, \quad & d_R &\to V_R^d d_R, \\ \ell_L &\to V_L^\ell \ell_L, \quad & \ell_R &\to V_R^\ell \ell_R. \end{aligned} \tag{1.32}$$

While leaving the neutral currents unchanged, this rotation introduces a mixing in the quark section of the electroweak charged currents, allowing transitions between quark families. This phenomenon, known as the Cabibbo-Kobayashi-Maskawa (CKM) mechanism, is described by the CKM unitary matrix and it's partially responsible of the CP violation observed in nature [13]. In the quark charged-weak current sector the interaction lagrangian, extended to all known fermion families is

$$\mathcal{L}_{\text{CC}} = \frac{g}{2\sqrt{2}} \left[W_\mu \sum_{ij} \bar{u}_i \gamma^\mu (1 - \gamma_5) \mathbf{V}_{ij} d_j + W_\mu \sum_l \bar{\nu} \gamma^\mu (1 - \gamma_5) \ell \right] + h.c., \tag{1.33}$$

where $\mathbf{V}$ represents the CKM matrix and in the three flavors scheme it is described by three angles and one phase. The complex phase in the CKM is the only one in the

Figure 1.4 Flavor-changing transitions through the charged-current couplings of the $W^{\pm}$ bosons [11].

SM Lagrangian, therefore the only source of CP-violation phenomena. Experimental measurements show a hierarchical pattern as the elements on the diagonal are very close to unity indicating that the flavor mixing is possible still disfavored. Therefore is usually more informative to represent the CKM in the Wolfenstein parametrization [14]:

$$\mathbf{V} = \begin{pmatrix} 1 - \frac{\lambda^2}{2} & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda^2 & 1 - \frac{\lambda^2}{2} & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix}, \tag{1.34}$$

where $\lambda \approx |\mathbf{V}_{us}| = 0.22431 \pm 0.00085$ represents the coupling between the first and second generation [13]. The mixing between the second and third generations is of the order λ^2, while the mixing between the first and third generation is much weaker, of the order of λ^3.

The neutrinos are massless fields in the minimal SM description, this allows to redefine the neutrino flavors in such a way to eliminate the flavor mixing in the lepton sector, thus lepton flavor conservation in the minimal SM arise.

1.1.5 Open questions of the Standard Model

The SM of particle physics has proven extremely successful in providing precise predictions of particle dynamics. The observation, in 2012 by the ATLAS and CMS experiments at the LHC, of a particle with characteristics compatible with the Higgs boson crowned the success of the SM in its current formulation.

Despite its remarkable achievements, the SM is far from being a complete and satisfactory description of nature, as several theoretical and experimental inconsistencies remain [15]:

- Although the SM is fully compatible with special relativity, it does not incorporate general relativity and the **gravitational interaction**. Thus, only three of the four fundamental forces are described within the theory.
- The SM does not provide an explanation for **dark matter**, whose existence is strongly supported by astrophysical and cosmological observations [16].
- The SM cannot account for the observed **baryon asymmetry** of the universe. While the electroweak and strong interactions are nearly symmetric between

matter and antimatter, our universe is composed almost entirely of matter. The CP violation predicted by the complex phase in the quark mixing CKM matrix (Eq. 1.34) is too small to explain the observed asymmetry [17; 18].

- In the SM, **neutrinos are massless**. However, the discovery of neutrino oscillations has demonstrated that neutrinos have small but non-zero mass differences. Only upper limits on the absolute neutrino mass scale have been established so far [19; 20].

1.2 Accidental symmetries

The minimal formulation of the Standard Model theory imposes explicitly the invariance under the gauge symmetry group $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ which get spontaneously broken into $SU(3)_C \otimes U(1)_Q$ to describe the massive nature of particles. Besides, the SM possesses a number of accidental continuous global symmetries, namely the baryon number symmetry, $U(1)_B$, and three individual lepton number symmetries: $U(1)_e$, $U(1)_\mu$ and $U(1)_\tau$. Those impose the conservation of three quantum numbers separately in particles interactions: the electron ($L_e = +1$ for e^- and ν_e), muon ($L_\mu = +1$ for μ^- and ν_μ) and tau ($L_\tau = +1$ for τ^- and ν_τ) lepton number.
These are also referred to as accidental symmetries, because they do not arise from the explicit requirement of invariance under a symmetry group, i.e. they are not *gauged* by any symmetry group. Albeit, from experiments, they appear to be extremely good symmetries of Nature their theoretical origin beg the question on why and whether they are such good symmetries.

Defining the total lepton number $L \equiv L_e + L_\mu + L_\tau$ the global symmetry group can also be written as

$$U(1)_{B+L} \otimes U(1)_{B-L} \otimes U(1)_{L_\mu - L_\tau} \otimes U(1)_{L_\mu + L_\tau - 2L_e}. \tag{1.35}$$

For the purposes of this work we will concentrate on the lepton flavor symmetry.

1.2.1 Charged lepton flavor violation

As described in Sec. 1.1.4 the flavor sector in the SM arises from the Yukawa interaction between the fermion fields and the Higgs field (see Eq. 1.33). Flavor violation in the quark sector arises from the fact that the diagonalization of the mass matrices requires different operators on up-like and down-like quarks leading to the flavor mixing parametrized in the CKM matrix.
In the lepton sector, where neutrinos are massless, one can apply the same transformation to charged and neutral leptons, because no other term in the Lagrangian involves the lepton doublets. Thus no mixing and no flavor violation arises.
The experimental observation of neutrino oscillations proved that they are massive and, clearly, the flavor conservation does not hold any longer in the minimal Standard Model extension with the generation of neutrinos mass terms. In other words the neutrino oscillation proves that the $U(1)_{L_\mu - L_\tau} \otimes U(1)_{L_\mu + L_\tau - 2L_e}$, in the larger flavor symmetry group of Eq. 1.35, is not conserved in nature. Nonetheless, charged

lepton flavor violation (CLFV) processes are expected to be suppressed by a factor m_{ν_i}/M_W, with m_{ν_i} the mass of the i-th neutrino mass eigenstate and M_W the W boson mass.

Let us consider the SM extended with neutrino masses. The flavor eigenstates of the left-handed neutrino fields $\nu_{L,\ell}$, entering the weak interaction, are linear combination of the mass eigenstates. The rotation from the mass to the flavor eigenstates basis is described by the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix [21; 22], defined as

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = U \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix}, \tag{1.36}$$

where the PMNS matrix can be parametrized in terms of three rotation angles (θ_{12}, θ_{13} and θ_{23}) and one complex phase (δ_{CP}) as follows:

$$U = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \cdot \begin{pmatrix} c_{13} & 0 & s_{13}e^{-i\delta_{\rm CP}} \\ 0 & 1 & 0 \\ -s_{13}e^{i\delta_{\rm CP}} & 0 & c_{13} \end{pmatrix} \cdot \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{1.37}$$

where s_{ij} and c_{ij} indicates $\sin\theta_{ij}$ and $\cos\theta_{ij}$, respectively. The flavor mixing parametrized by the PMNS matrix is analogous to the one parametrized by the CKM matrix.

The possible LFV decay of an heavy lepton ($L = \mu,\ \tau$) into three lighter charged leptons ($\ell, \ell' = \mu, e$) can proceed via three classes of one-loop diagrams: the box diagram and the Z and γ penguin diagrams, as reported in Fig. 1.5 [23], whose amplitude can be written as

$$i\mathcal{A}_\gamma = -\frac{\alpha G_F}{\sqrt{2\pi}}(\bar{\ell}\gamma^\mu P_L L)(\bar{\ell}'\gamma_\mu \ell') \sum_i U_{\ell i} U^*_{\ell' i}\, x_i, \tag{1.38}$$

and

$$\begin{aligned} i\mathcal{A}_Z = {} & \frac{\alpha G_F}{\sqrt{2\pi}} \sum_i U_{\ell i} U^*_{\ell' i}\, x_i \,(3 + \log x_i) \\ & \times \left[(\bar{\ell}\gamma^\mu P_L L)(\bar{\ell}'\gamma_\mu \ell') - \frac{1}{2\sin^2\theta_W}(\bar{\ell}\gamma^\mu P_L L)(\bar{\ell}'\gamma_\mu P_L \ell') \right], \end{aligned} \tag{1.39}$$

$$\begin{aligned} i\mathcal{A}_{\rm BOX} = {} & \frac{\alpha G_F}{2\sqrt{2\pi}\,\sin^2\theta_W}(\bar{\ell}\gamma^\mu P_L L)(\bar{\ell}'\gamma_\mu P_L \ell') \\ & \times \sum_i U_{\ell i} U^*_{\ell' i}\, x_i \,(1 + \log x_i), \end{aligned} \tag{1.40}$$

where G_F and α are the fine structure and Fermi constants, respectively, U is the PMNS matrix and the factor x_i indicates the ratio m_{ν_i}/M_W. Predictions on the $\Gamma(L \to \ell\ell\ell)$ decay amplitude are obtained in the physical limit (PL) approach embedding the hierarchy $m_{\nu_i} \ll m_\ell, m_L \ll M_W$ and keeping the full dependence

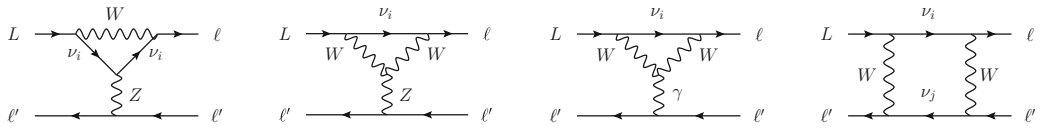


Figure 1.5 One-loop diagrams providing contribution to the CLFV processes $L \to \ell\ell'\ell'$ in the frame of the SM including massive neutrino oscillations. From left to right: Z penguins, photon penguin and box diagram. [23]

on the external and internal particle kinematics. In this frame and for the normal ordering of neutrino masses the decay amplitude is expressed as

$$\begin{aligned}\frac{\Gamma(L \to \ell\ell\ell)}{\Gamma_0} = \frac{3\alpha^2}{16\pi^2}\left|\sum_{i=2}^{3} U_{\ell i}U_{Li}^{*}\frac{\Delta m_{i1}^2}{M_W^2}\right|^2 \\ \times\left[\log^2 x_L + 2\log x_L - \frac{1}{6}\log x_\ell + \frac{19}{18} + \frac{17}{18}\pi^2\right. \\ \left. - \frac{1}{\sin^2\theta_W}\left(\log x_L + \frac{11}{12}\right) + \frac{3}{8\sin^4\theta_W}\right],\end{aligned} \tag{1.41}$$

where $x_L = m_L^2/M_W^2$ and $x_\ell = m_\ell^2/M_W^2$. Equation 1.41 depends only on the neutrino mass splitting and the decay amplitude is power suppressed by $|\sum_i U_{\ell i}U_{Li}^{*}x_i|$, leading to very suppressed branching ratios of the order of 10^{-55}, as outlined in Table 1.1 for the $\tau \to 3\mu$ and other LFV decays.

Table 1.1 Branching ratio of relevant charged LFV processes predicted in the PL for the normal (inverted) neutrino mass order [23].

	Branching ratio ($\mathcal{B}r$) [$\times 10^{-55}$]
$\tau \to \mu\mu\mu$	5.8 (5.8)
$\tau \to \mu ee$	3.8 (3.8)
$\mu \to eee$	2.9 (4.6)
$\tau \to eee$	0.33 (0.19)
$\tau \to e\mu\mu$	0.21 (0.12)

1.2.2 Gates to new Physics through $\tau \to 3\mu$ decays

The values of the branching ratios of LFV processes, obtained within the SM including the neutrino masses (Tab. 1.1), are far beyond the physics reach of current experiments, which nowadays are able to achieve sensitivities of the order of 10^{-8}. However, many extensions of the SM, referred to as Beyond the Standard Model (BSM) theories, predicts enhanced LFV branching ratios, much higher than the SM prediction and close to the experimental sensitivity.

Define CLFV as a process that conserves the total lepton number L but violates the symmetry $U(1)_{L_\mu - L_\tau} \otimes U(1)_{L_\mu + L_\tau - 2L_e}$. Hence violating decays can be divided in

different *groups* depending on whether they violate $L_e - L_\mu$, $L_\mu - L_\tau$, $L_e - L_\tau$ ecc. A few examples for each group are collected in Table 1.2. A somewhat trivial but crucial point is that if *one* process of a given group is observed, then all processes of the same group are allowed. However, this just proves that only one linear combination of lepton numbers is broken while others may still be conserved. Only the observation of at least two CLFV decays of two different groups would prove that CLFV exists and that the flavor group $U(1)_{L_\mu - L_\tau} \otimes U(1)_{L_\mu + L_\tau - 2L_e}$ is fully broken [24]. More specifically if a violation $\Delta(L_\alpha - L_\beta) = 2$ is observed then an additional violation of the kind $\Delta(L_\alpha - L_\gamma) = 2$ or $\Delta(L_\alpha + L_\gamma - 2L_\beta) = 2$ must be observed to fully break the flavor symmetry.
Nowadays, the most stringent exclusion limits are set on processes involving muons and electrons ($\Delta(L_e - L_\mu) = 2$), nevertheless, to fully constrain the lepton flavor symmetry other processes must be explored. The purpose of this work is to investigate LFV in $\tau \to 3\mu$ decays, which are relatively easy to explore at colliders, due to the clean signature and high identification efficiency for muons, and provide an orthogonal insight into the lepton flavor symmetry puzzle.

Table 1.2 CLFV with conserved L and B, omitting CP conjugate processes, with the relative current best exclusion limit at 90% CL. Processes are separated in independent group that partially violate the flavor symmetry $U(1)_{L_\mu - L_\tau} \otimes U(1)_{L_\mu + L_\tau - 2L_e}$ [24].

Group	Process	Current exclusion limit at 90% CL
$\Delta(L_e - L_\mu) = 2$	$\mu \to e\gamma$	4.2×10^{-13} [25]
	$\mu \to eee$	1.0×10^{-12} [26]
$\Delta(L_e - L_\tau) = 2$	$\tau \to eee$	2.7×10^{-8} [27]
	$\tau \to e\mu\mu$	2.7×10^{-8} [27]
	$\tau \to e\gamma$	3.3×10^{-8} [28]
$\Delta(L_\mu - L_\tau) = 2$	$\tau \to \mu ee$	1.8×10^{-8} [27]
	$\tau \to \mu\mu\mu$	2.1×10^{-8} [29]
	$\tau \to \mu\gamma$	4.4×10^{-8} [28]
$\Delta(L_\mu + L_\tau - 2L_e) = 6$	$\tau \to ee\bar{\mu}$	1.5×10^{-8} [27]
$\Delta(L_\tau + L_e - 2L_\mu) = 6$	$\tau \to \mu\mu\bar{e}$	1.7×10^{-8} [27]
$\Delta(L_e + L_\mu - 2L_\tau) = 6$	$\mu e \to \tau\tau$	-

Several BSM models predict sizable branching ratios for the $\tau \to 3\mu$ decay, up to $\mathcal{O}(10^{-9})$. LFV decays can be induced by the introduction of additional right-handed heavy Dirac or Majorana neutrinos, the latter arising, for instance, in the context of supersymmetric extensions of the SM [30; 31; 32]. Within supersymmetry (SUSY) frameworks, the $\tau \to 3\mu$ decay can also receive sizable contributions from Higgs-mediated processes through the inclusion of a supersymmetric seesaw mechanism [33]. Additional potential sources of LFV appear in models extending the Higgs sector with an $SU(2)$ triplet scalar field with unit hypercharge. The presence of such an extended Higgs sector naturally allows for small neutrino masses and flavor-violating Yukawa

interactions with SM fermions [34]. In these scenarios, the predicted branching ratio $\mathcal{B}r(\tau \to 3\mu)$ can reach values of the order of 10^{-7}. Some SM extension theories introduce additional gauge symmetries holding at energies above the TeV scale [35]. Significant enhancement of the $\tau \to 3\mu$ decay rate can arise from LFV interactions mediated by massive Z' bosons generated through the SSB of these extended symmetry groups (see Fig. 1.6). For some of these model, the predicted $\tau \to 3\mu$ branching ratio can reach up to 10^{-9}, approaching the current experimental sensitivity.

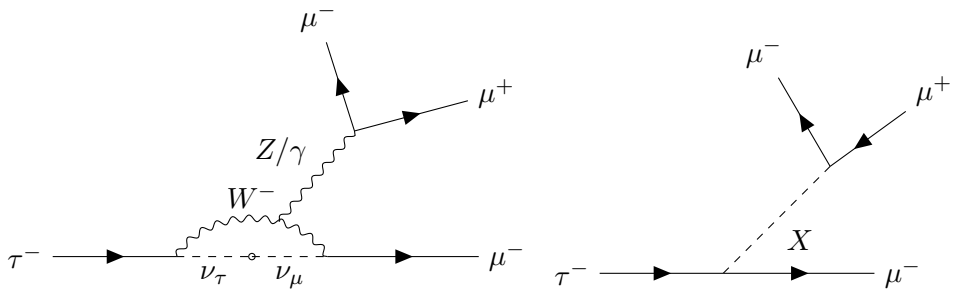


Figure 1.6 Feynman diagrams of the $\tau \to 3\mu$ SM decay through neutrino oscillation (left) and through a new Physics process involving the unknown particle X (right).

1.3 Experimental scenario for the $\tau \to 3\mu$ decay

As outlined in the previous sections, LFV decays of charged leptons are predicted to have a negligible probability within the SM, for the same reason they are powerful and experimentally clean probes of the known theory and possibly gates to uncover new Physics (NP). The focus of this work is on the search of $\tau \to 3\mu$ decay representing one of the golden channels to test lepton flavor symmetry in the charged sector at colliders and B-factories.

The great potential of this decay channel is testified by the large experimental effort put in its investigation by various collaborations. The main difference between their researches is the way τ leptons are produced in the various experiments. B-factories like Belle and BaBar exploit the large amount of τ leptons produced at the $\Upsilon(4s)$ resonance (Sec. 1.3.1). At hadron colliders, instead, τ leptons are produced via heavy flavor (HF) decays, i.e. from D and B mesons and via vector bosons decays. The LHCb experiment explores only the former production mode of τ leptons (Sec. 1.3.2). The ATLAS experiment investigates τ from vector bosons to cope with the reduced efficiency on the HF signal due to the high QCD contamination in the low momentum of the τ leptons (Sec. 1.3.3). The current experimental scenario is summarized in Table 1.3. The best limit has been set by the Belle II collaboration which set an exclusion limit on the $\tau \to 3\mu$ branching ratio of 2.1×10^{-8} at 90% confidence level (CL).

In this context the CMS experiment instead is able to efficiently detect τ leptons from both HF and vector boson decays reaching competitive performance with respect to other experiments.

Table 1.3 Current experimental status for the search of $\tau \to 3\mu$ decay. Observed (Obs.) and expected (Exp.) refer to the 90% CL exclusion limits on the branching fraction $\mathcal{B}r(\tau \to 3\mu)$.

Collaboration	Process	Luminosity [fb^{-1}]	Obs. (Exp.) [$\times 10^{-8}$]	Ref.
Belle II	$ee \to \tau\tau$	424	1.9	[29]
Belle	$ee \to \tau\tau$	782	2.1	[36]
CMS (Run 2)	$D, B \to \tau X$, $W \to \tau\nu$	131 (13 TeV)	2.9 (2.4)	[37]
Babar	$ee \to \tau\tau$	468	3.3 (4.0)	[38]
LHCb	$D, B \to \tau X$	3.0 (7-8 TeV)	4.6 (5.6)	[39]
ATLAS	$W \to \tau\nu$	20.3 (8 TeV)	39 (38)	[40]

1.3.1 B-factories

B-factories are high-luminosity electron–positron (e^+e^-) colliders specifically designed to produce large numbers of B mesons, enabling precision studies of their properties and decays. Operating at energies corresponding to the $\Upsilon(4S)$ resonance (10.48 GeV), i.e. just above the threshold for $B\bar{B}$ pair production, these experiments provide an ideal environment to investigate CP violation, flavor physics, and rare processes in the quark sector. The electron-positron collision energy also corresponds to a large cross section (0.9 nb) for $\tau^+\tau^-$ production, making the experiments at B-factories an excellent source of τ leptons in a clean environment.
The general methodology for $\tau \to 3\mu$ searches consists in treating separately the two τ leptons of the event. One is required to decay into SM final states (tag side) and the other one into the LFV final state (signal side). To exploit the missing energy expected in the tag side and reduce background contamination, each event is divided into two hemispheres, one containing the signal τ and the other containing the tag τ.

At present, the best experimental sensitivity on the $\tau \to 3\mu$ decay is achieved by Belle and Belle II Collaborations, which collected, respectively, 719 million (782 fb^{-1}) and 389 million (424 fb^{-1}) $e^+e^- \to \tau^+\tau^-$ events at the asymmetric-energy e^+e^- KEKB and SuperKEKB colliders [41; 42]. The Belle II analysis represents a complete re-implementation of the previous Belle search, featuring a significantly improved signal efficiency through the use of inclusive tagging, which encompasses a wider range of τ decay modes, and machine learning techniques for event selection. Thanks to these improvements, the signal efficiency increased from 7.6 % to 20.4 % in the latest results [29; 36].
In $\tau \to 3\mu$ decays, the invariant mass $M_{3\mu}$ of the reconstructed muons should be consistent with the τ mass, except for events affected by final-state radiation (FSR) from the τ or its decay products. The energy E^* of the τ center-of-mass frame is expected to be half of the e^+e^- center of mass energy, with corrections arising from initial-state radiation (ISR), beam effects and FSR. Consequently, the energy difference $\Delta E_{3\mu} = E^*_\tau - \sqrt{s}/2$ should be close to zero. Signal candidates for $\tau \to 3\mu$ events are studied in the two-dimensional plane ($\Delta E_{3\mu}$, $M_{3\mu}$) where signal events

are expected to cluster around $M_{3\mu}$ close to the τ mass (1.78 GeV) and $\Delta E_{3\mu}$ near to zero (Fig. 1.7). The final limit is derived in the elliptical region of this plane containing 90% of the signal. Since no evidence of signal was found, the Belle II Collaboration reported an observed upper limit of

$$Br(\tau \to 3\mu)_{\text{Belle II}} < (2.1^{+5.1}_{-2.4} \pm 0.4) \times 10^{-8}, \tag{1.42}$$

where the first uncertainty is statistical and the second is systematic.

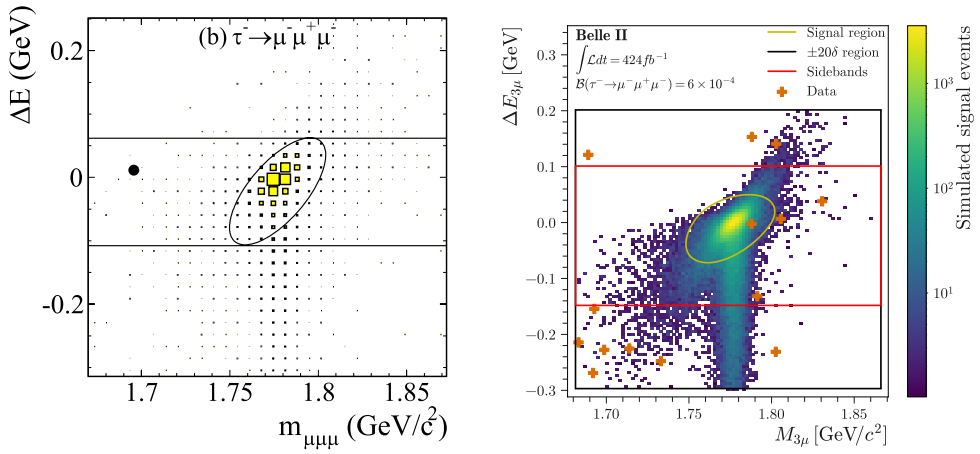


Figure 1.7 Scatter plot of selected events in the ($\Delta E_{3\mu}$, $M_{3\mu}$) plane for Belle (left) [36] and Belle II results (right) [29]. On the left (right) plot data are represented by the black dot (orange cross) and signal is represented by the yellow boxes (color-filled area). The ellipses represent the signal region where the sensitivity is calculated.

A similar strategy was employed in one of the first searches for the $\tau \to 3\mu$ decay was performed at the BaBar experiment at the SLAC (Stanford Linear Accelerator Center) National Accelerator Laboratory [43], using an integrated luminosity of 468 fb^{-1} and setting an exclusion limit of $3.3{\times}10^{-8}$ at 90% CL [38].

1.3.2 LHCb

The first search for the $\tau \to 3\mu$ decay at a hadron collider was conducted by the LHCb experiment using proton-proton collision data at a center-of-mass energy of 7 TeV (1.0 fb^{-1}) and 8 TeV (2.0 fb^{-1}) [39]. The analysis takes advantage of the large inclusive τ production cross section at the LHC, as τ leptons are predominantly produced in b and c hadrons decays. Using the LHCb measurements for the $b\bar{b}$ and $c\bar{c}$ production cross sections, together with the branching ratios $b \to \tau$ and $c \to \tau$, the inclusive τ cross section is estimated to be 85 μb at 7 TeV.

In the LHCb approach, the $\tau \to 3\mu$ signal is searched in the invariant mass distribution of the three muons which would form a narrow resonance around the τ mass. The selection criteria are optimized for the $\tau \to 3\mu$ signal signature. In addition, $D_s^{\pm} \to \phi(\mu\mu)\pi^{\pm}$ events are selected using a strategy similar as for the signal and serve as normalization and calibration channel. To discriminate between signal and background, three classifiers are employed: one based on the geometrical and

kinematic properties of the final-state tracks and the reconstructed τ candidate, one exploiting particle-identification information associated with the tracks, and one using the reconstructed mass of the τ candidate.
The main background sources in the $\tau \to 3\mu$ signal region arise from heavy-meson decays yielding three muons in the final state and from processes with one or two genuine muons combined with misidentified hadrons or with additional muons produced elsewhere in the event. Also, a significant fraction of the background originates from decays involving undetected final-state particles.
No significant excess of events over the expected background was observed. An upper limit on the branching fraction was therefore set as

$$Br(\tau \to 3\mu)_{\text{LHCb}} < 4.6\ (5.0) \times 10^{-8}, \tag{1.43}$$

where the observed (expected) bounds are calculated at 90 % CL.

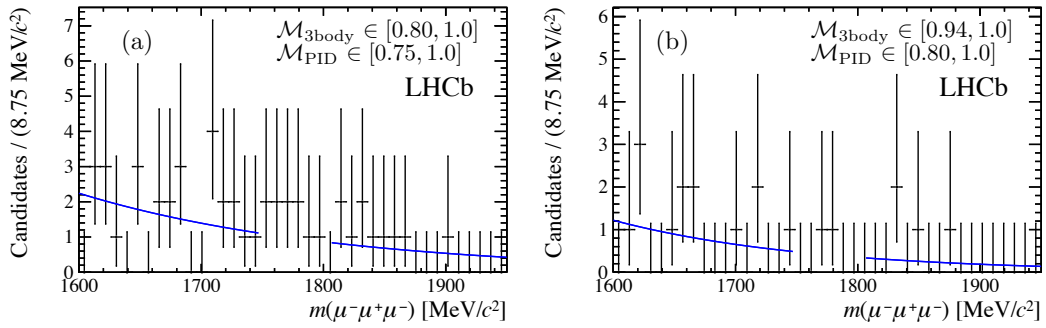


Figure 1.8 Invariant mass distributions and fits to the mass sidebands in (a) 7 TeV and (b) 8 TeV data for $\tau \to 3\mu$ candidates in the most sensitive category of [39].

1.3.3 ATLAS

The ATLAS collaboration searches $\tau \to 3\mu$ decays using τ leptons produced via $W \to \tau\nu$ decays in proton-proton collisions data collected at a center-of-mass energy of 8 TeV, corresponding to 20.3 fb^{-1} [40].

The analysis strategy consists in selecting events with three muons compatible with a common vertex (p_T (μ) > 2.5 GeV). A loose event selection is applied to collect a high-quality sample of candidates with invariant mass within 1 GeV from the τ mass. The loose sample of events is then analyzed with a Boosted Decision Tree (BDT) trained on a background sample, i.e. the data in the 3μ mass sidebands, and the signal simulated sample. The final selection combines more stringent requirements on the $\tau \to 3\mu$ candidates properties and a cut on the per-candidate BDT score. A search is performed for an excess of events at the τ lepton mass above the expected background level. The signal region (SR) is defined as a 2σ region around the τ lepton mass with σ corresponding to the three-muon mass resolution measured on the signal MC sample (Figure 1.9). No events are observed in the signal region for the final selection while $0.193 \pm 0.131(\text{syst}) \pm 0.037(\text{stat})$ background events are expected. This results in an upper limit of

$$Br(\tau \to 3\mu)_{\text{ATLAS}} < 37.6\ (39.4) \times 10^{-8}, \tag{1.44}$$

where the observed (expected) bounds are calculated at 90 % CL.

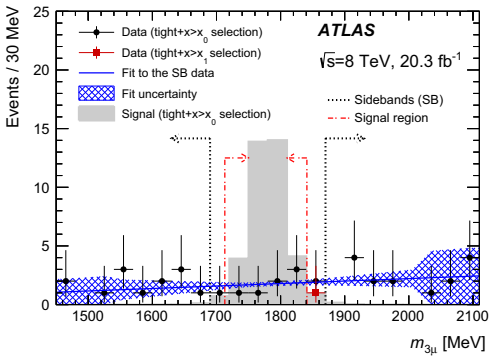


Figure 1.9 The 3μ mass distribution of data in the full mass spectrum for a loose cut on the BDT score (solid black circle) and for the final tight selection on the BDT (solid red circle). The blue line shows the result of a fit to the distribution in data, while the hatched area shows the fit uncertainty due to the SB range definition, the BDT cut location and the fit function choice. The solid gray area shows the signal shape (obtained from MC simulation), normalized to the area of the data for the selection [40].

1.3.4 CMS Run 2

The CMS experiment performed a search for LFV decay $\tau \to 3\mu$ using the full Run 2 dataset, corresponding to a total integrated luminosity of 131 fb^{-1}. The overall CMS sensitivity results from the combination of two analyses: the first exclusion limit, obtained analyzing 33 fb^{-1} of proton-proton collision data collected at a center of mass energy of 13 TeV during 2016 [44], and the improved sensitivity achieved with the larger dataset collected between 2017 and 2018 [45].

CMS is the only LHC experiment that exploits both the HF and vector-boson production modes in two orthogonal analyses. Both analyses target the distinctive experimental signature of three muons originating from a common secondary vertex. The main source of background arises from combinatorial candidates, i.e. τ lepton candidates reconstructed from three unrelated muons. In both analyses the background is modeled using data-driven techniques based on the sidebands of the three-muon invariant-mass distribution. BDTs are employed to optimize the signal selection and to enhance the sensitivity with different implementation in the HF and W analyses. Since the HF channel is more prone to contamination from hadrons misidentified as muons, a dedicated muon reconstruction quality BDT is trained to suppress this background source. Conversely, the W channel analysis BDT leverages the $\tau \to 3\mu$ candidate isolation and the missing energy related observables which are characteristics of the $W \to \tau\nu$ topology. Figure 1.10 shows some of the most discriminating variables used in the two channels, comparing the expected $\tau \to 3\mu$ signal with the data-driven background distributions.

Events are divided into orthogonal categories based on the relative mass resolution of the three-muon candidate. The branching fraction $\mathcal{B}(\tau \to 3\mu)$ is extracted through a simultaneous maximum-likelihood fit to the invariant mass distribution of the

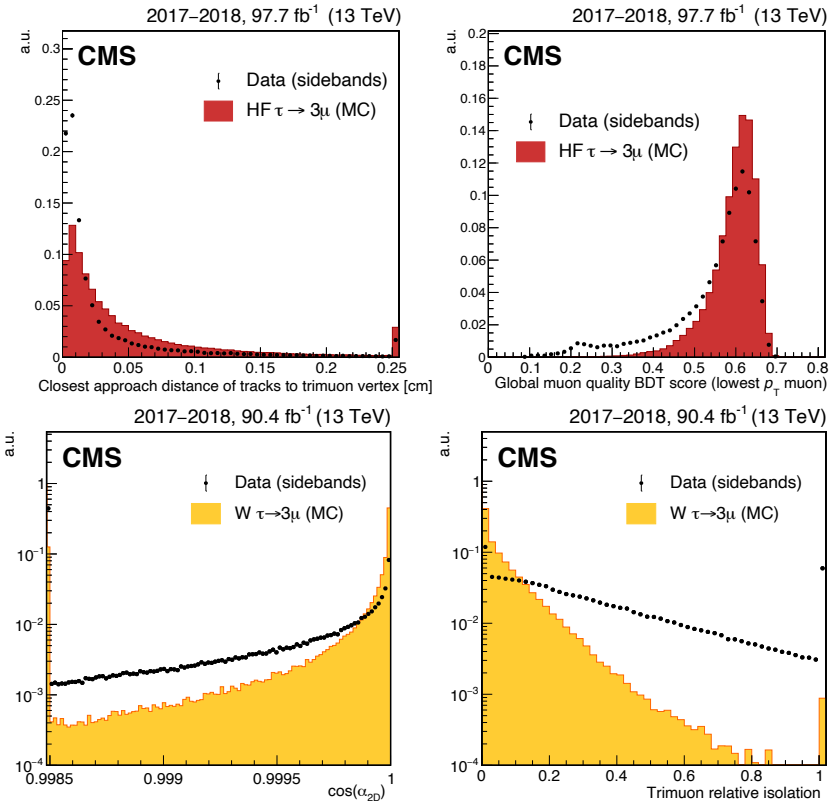


Figure 1.10 Signal and background distributions for observable used in the heavy-flavor (top) and in the W channel (bottom) analysis for th e search of $\tau \to 3\mu$ at CMS using the 2017 and 2018 data. [37]

three muons across all categories in the two channels. The results are shown in Figure 1.11.

No significant excess of events is observed, and an observed (expected) upper limit is set at 90% confidence level as reported below:

$$Br(\tau \to 3\mu)_{\text{CMS}} < 2.9\ (2.4) \times 10^{-8}. \tag{1.45}$$

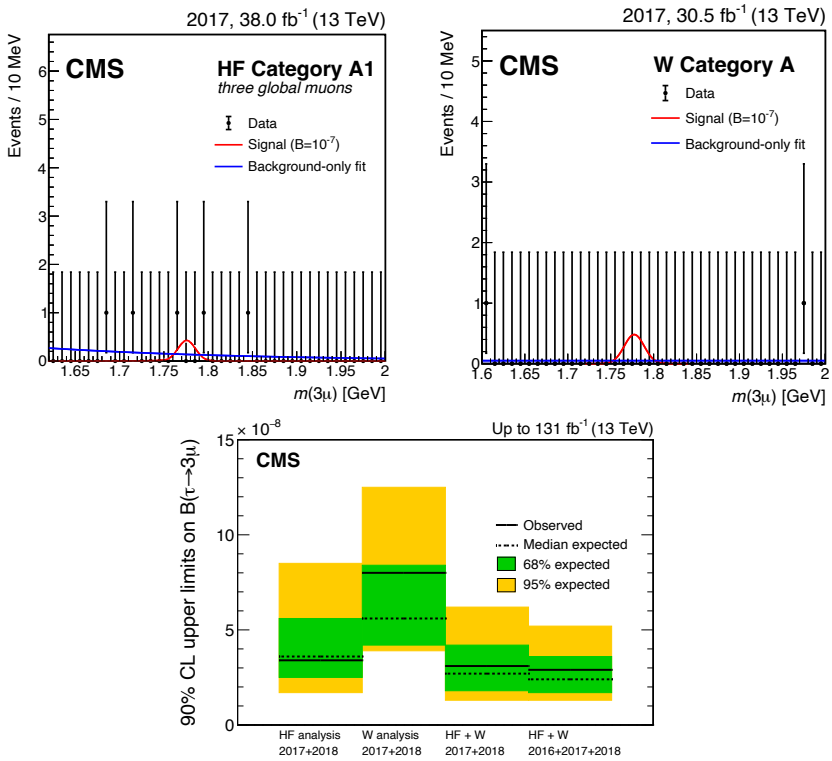


Figure 1.11 Three muon invariant mass distribution in the highest sensitivity category of HF (top left) and W (top right) channel. The red solid line represents expected signal from simulation, assuming $\mathcal{B}(\tau \to 3\mu) = 10^{-7}$, and the blue solid line represents the background model obtained for the fit to the mass sidebands. On the bottom plot the expected and observed upper limit on $\mathcal{B}(\tau \to 3\mu)$ [45].

1.3.5 Novel opportunities for LFV in $\tau \to 3\mu$ at CMS Run 3

The CMS Run 3 dataset opens unprecedented opportunities to expand the W channel analysis. A major step forward is the implementation of an improved trigger strategy enhancing the signal acceptance, specifically for the W channel. The worth of this effort is testified by the fact that despite, the W yield being four orders of magnitude lower than that of the HF, the clean signature and high selection efficiency make the contribution significant: it contributes up to 25% of the total sensitivity in the combined result. In parallel, the larger dataset and refined reconstruction techniques enable a more accurate treatment of the dominant backgrounds. The analysis can expand further beyond the $W \to \tau\nu$ production channel into the broader vector boson sector, for instance exploiting $Z \to \tau\tau$ events as an additional source of τ leptons. This expansion not only increases the available signal yield but also strengthens the robustness of the search program, positioning CMS as one of the leading experiment in this research field.
This research will be the core topic of this thesis work.

Chapter 2

The CMS detector at the Large Hadron Collider

This chapter provides an overview of the experimental environment in which the search for lepton flavor violation in the $\tau \to 3\mu$ decay is performed. Section 2.1 introduces the LHC, built at the Conseil Européen pour la Recherche Nucléaire (CERN) laboratory, where proton–proton collisions are produced at the highest energies ever achieved in a laboratory. The main features of the CMS detector are described in Section 2.2 and the algorithm for the particle reconstruction is then described in Section 2.3. For the purpose of this work, particular emphasis is placed on the detector components and algorithms relevant to muon identification and reconstruction. Finally, Section 2.4 presents the architecture of the CMS trigger and data-acquisition systems and discuss the special data streams, data scouting and data parking, that play a crucial role in extending the experiment's sensitivity to low-mass signatures during Run 3, including the search for LFV in $\tau \to 3\mu$ decays.

2.1 The Large Hadron Collider

The LHC is the world's largest and most powerful particle accelerator. It measures 27 km of length and it is capable of producing symmetric proton-proton (pp) collisions up to a center of mass energy $\sqrt{s} = 13.6$ TeV [46]. The LHC can also accelerate and collide heavy ions at an average center-of-mass energy of 5.36 TeV per nucleon. It has been installed in the underground tunnel previously occupied by the Large Electron Positron (LEP) collider, active until 2000 and remains the latest addition to the CERN particle accelerator complex, represented in Fig. 2.1.

The LHC beams are crossed in four interaction points where experiments are located. The A Toroidal LHC ApparatuS (ATLAS) [47] and CMS (Section 2.2) are general-purpose detectors, designed for a large program of physics analyses. They mainly analyze data from pp interactions exploiting different technical solutions for the detector designs. The A Large Ion Collider Experiment (ALICE) physics program mainly focuses on heavy-ion collisions, addressing the physics of the quark-gluon

plasma at extreme values of energy density and temperature in nucleus-nucleus collision [48]. The Large Hadron Collider beauty (LHCb) experiment is optimized for studies on matter-antimatter asymmetry, CP violation and flavor physics through analyses with peculiar focus on involving b quarks [49].

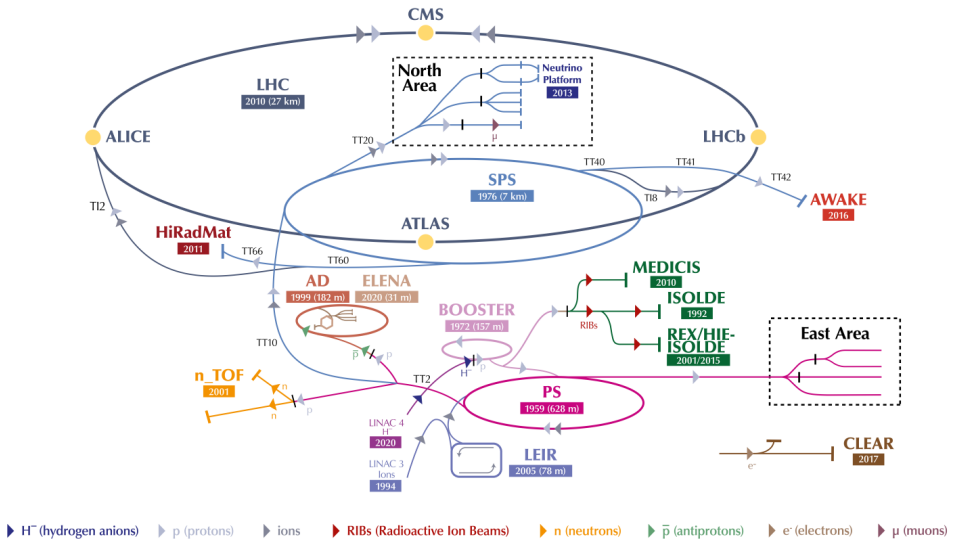


Figure 2.1 The CERN accelerator complex including LHC and its pre-accelerating system.

The LHC timeline, represented in Fig. 2.2, is divided into three periods of data taking separated by shutdowns periods. The first research run (Run 1) took place between March 2010 and February 2013 at an initial center-of-mass energy of 7 TeV, increased to 8 TeV from 2012. Analyzing the data collected in the Run 1 the collaborations of the ATLAS and CMS experiments discovered the Higgs boson in 2012.
After a shutdown of about three years for planned upgrades, the LHC Run 2 took place between 2016 and 2018, during which the ATLAS and CMS experiments recorded an integrated luminosity of about 140 fb^{-1} of pp collisions at a center-of-mass energy of 13 TeV. As detailed in Sec. 2.1.1, luminosity is a crucial indicator of an accelerator performance: it is proportional to the number of collisions that occur in a given amount of time. The second long-shutdown was extended until 2022 because of Covid-19 pandemic. The LHC is now towards the end of its Run 3, which will last until mid 2026. For the first time the LHC is producing proton-proton collisions at center-of-mass energy of 13.6 TeV and delivered about 316 fb^{-1} by the end of 2025, almost double the Run 2 integrated luminosity.
The new Run 3 dataset, in combination with the Run 2 dataset, will improve the searches for very high-mass resonances, the observation of rare SM physics processes (like the production of multiple vector bosons), and the characterization of the Higgs boson properties by differential analyses of its cross section.

After the conclusion of the Run 3 scientific program another long shut-down is scheduled until 2030 when the High Luminosity LHC (HL-LHC) program will start. The objective is to increase the integrated luminosity by a factor of 10 beyond the

LHC's design value, to sensibly expand the possibility of accessing rare SM and BSM signatures. This requires a substantial upgrade of the CMS subdetectors to be able to cope with the unprecedented data taking conditions. Details about the target design and the technical challenges deriving from such ambitious goal are better discussed in the Appendix A.

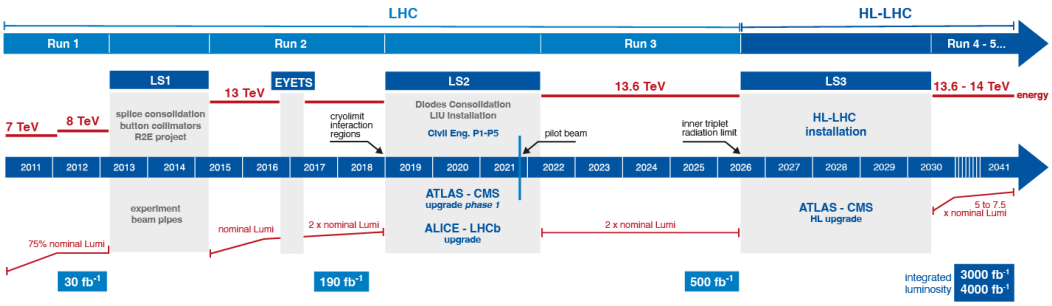


Figure 2.2 Timeline of the LHC operations since the beginning of Run 1 until the first HL-LHC collisions.

2.1.1 Proton-proton collision dynamics

The protons injected into the LHC undergo sequential acceleration steps, in different pre-accelerator facilities before being injected in the largest ring at the nominal energy. Protons are extracted from a hydrogen source and first accelerated by the linear accelerator LINAC 4 to an energy of 160 MeV. They are then injected into the Proton Synchrotron Booster (PSB), where their energy is increased to 2 GeV before being transferred to the Proton Synchrotron (PS). In the PS, the protons are further accelerated up to 26 GeV. The final stage of the pre-acceleration chain is the Super Proton Synchrotron (SPS), a circular accelerator with a radius of about 1 km, where the proton energy reaches 450 GeV. The protons are then ready to be injected into the LHC, where they are accelerated to the final energy of 6.8 TeV per beam. The protons circulate in the LHC ring in two separate beam pipes, moving in opposite directions. To maintain them on their circular trajectories, the LHC employs 1232 superconducting dipole magnets made of Niobium–Titanium, each 14.2 m long and cooled to 1.9 K using superfluid helium. Each magnet generates a bending magnetic field of about 8.3 T, oriented in opposite directions in the two beam pipes to steer the counter-rotating proton beams.
Protons are injected in the LHC ring in bunches. One orbit is divided into a total of 3564 time windows, each 25 ns in duration (bunch crossing slots) and potentially containing a colliding proton bunch. The actual collision rate depends on the number of colliding bunches and the structure of the filling scheme, which can be varied upon the physics and machine needs. In Run 3 the number of colliding bunches increased from 2300 in 2022 up to 2448 in 2025.

In hadron colliders, the actual particles participating in interactions and responsible for producing the final states are the partons, i.e. the quarks and gluons that originate inside the proton and carry a fraction of its total momentum. The proton's quantum numbers are defined by its valence-quark content (uud), while gluons and

additional quark–antiquark pairs (sea quarks) continuously form inside the proton as color-singlet states. The total momentum of the colliding partons is not known a priori and therefore cannot be used as a fixed kinematic constraint. However, for protons moving collinear along the beam axis, the total transverse momentum is known and equal to zero. The internal structure of the proton is described by the parton distribution functions (PDFs), $f_i(x, \mu^2)$, which express the probability density of finding a given parton carrying a fraction x of the proton's longitudinal momentum when probed at resolution scale μ^2 [21]. Since partons are confined inside hadrons, PDFs cannot be derived from first principles within perturbative QCD. Instead, they are extracted from experimental data under the assumption of QCD factorization (such as deep inelastic scattering and Drell–Yan processes), and evolved with Q^2 according to the DGLAP equations [50]. As illustrated in Fig. 2.3, the dominant contributions come from the valence quarks and gluons, which have the largest PDFs values.

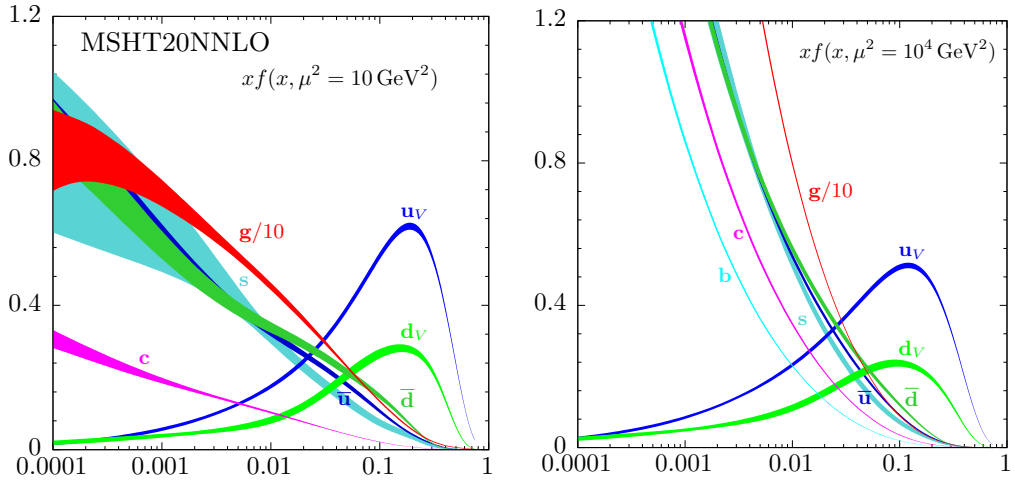


Figure 2.3 Unpolarized parton distribution density functions obtained at scales μ^2=10 GeV^2 (left) and μ^2=10 GeV^2 (right) [21].

Besides the collision energy, one of the most important quantities used to characterize the performance of a collider is the instantaneous luminosity ($\mathcal{L}$), measuring the number of interactions the collider can produce per unit time. It gets particularly relevant for searches for rare processes with small production cross sections as it is defined as the proportionality factor between the event rate and the production cross section (σ_p):

$$\frac{dR}{dt} = \mathcal{L} \cdot \sigma_\mathrm{p}. \tag{2.1}$$

Protons in the LHC are grouped into bunches of about 10^{11} protons each with an orbit period of about 90 μs. Assuming the particle distribution in each bunch is approximately Gaussian, the instantaneous luminosity can be written as

$$\mathcal{L}(t) = \frac{N_1(t) N_2(t) f N_b}{4\pi\sigma_x\sigma_y} \cdot \mathcal{F}(t), \tag{2.2}$$

where f is the revolution frequency, N_b is the number of bunches per beam (typically 2448 in 2025), N_1 and N_2 are the numbers of protons per bunch, and σ_x and σ_y (about 20 μm) represent the transverse beam sizes. The factor $\mathcal{F}$ accounts for the small crossing angle (about 120 μrad) of the two beams at the interaction points, which slightly reduces the bunches overlap region, thus the luminosity. During one proton fill the beam intensity reduces because of the interactions, hence the N_i dependence on time, although the collision angle can be adjusted to partially compensate the luminosity loss. During Run 3, the LHC achieves an instantaneous luminosity of up to 2.5×10^{34} cm^{-2}s^{-1}, thanks to advanced beam engineering that maximize the effective overlap of the colliding bunches, thus increasing the value of $\mathcal{F}$. Integrated over time, the integrated luminosity quantifies the total number of pp interactions collected over a given period.

At the LHC, many collisions occur simultaneously within a single proton–proton bunch crossing. Another crucial parameter is the pileup (PU), which quantifies the average number of interactions occurring per bunch crossing. It is proportional to the instantaneous luminosity and the inelastic proton–proton cross section ($\sigma_{\text{inel}} \approx$ 60 mb):

$$\mu(t) = \frac{\sigma_{\text{inel}} \cdot \mathcal{L}(t)}{f N_b}. \tag{2.3}$$

In recent years, the LHC has exceeded its design luminosity performance, delivering more than 100 fb^{-1} of integrated luminosity per year. The LHC was originally designed to produce an average of about 25 pileup interactions per bunch-crossing. However, during Run 2, it far exceeded this value reaching average pileup up to 70 in Run 3, as shown in Fig. 2.4.

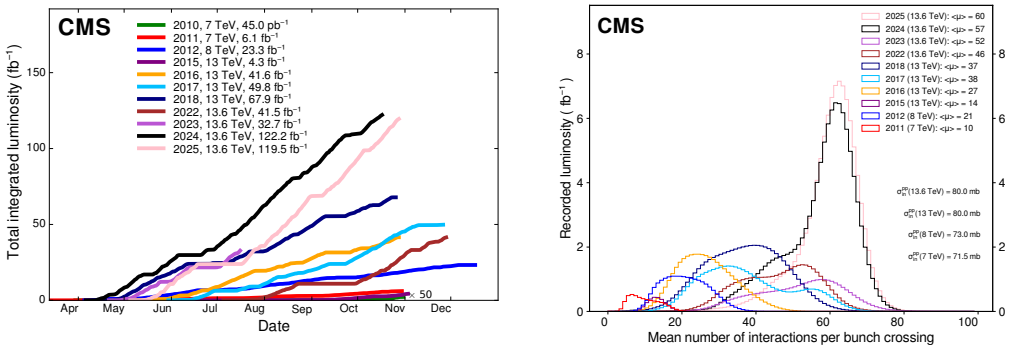


Figure 2.4 Left: Comparison of the cumulative luminosity versus day delivered to CMS experiment for pp collisions, at nominal center-of-mass energy in the Run 2 and Run 3 years of data-taking. Right: AveragePU distribution for pp collisions in Run 2 and Run 3 reordered by the CMS experiment. The overall mean values and the minimum bias cross sections are also shown [51].

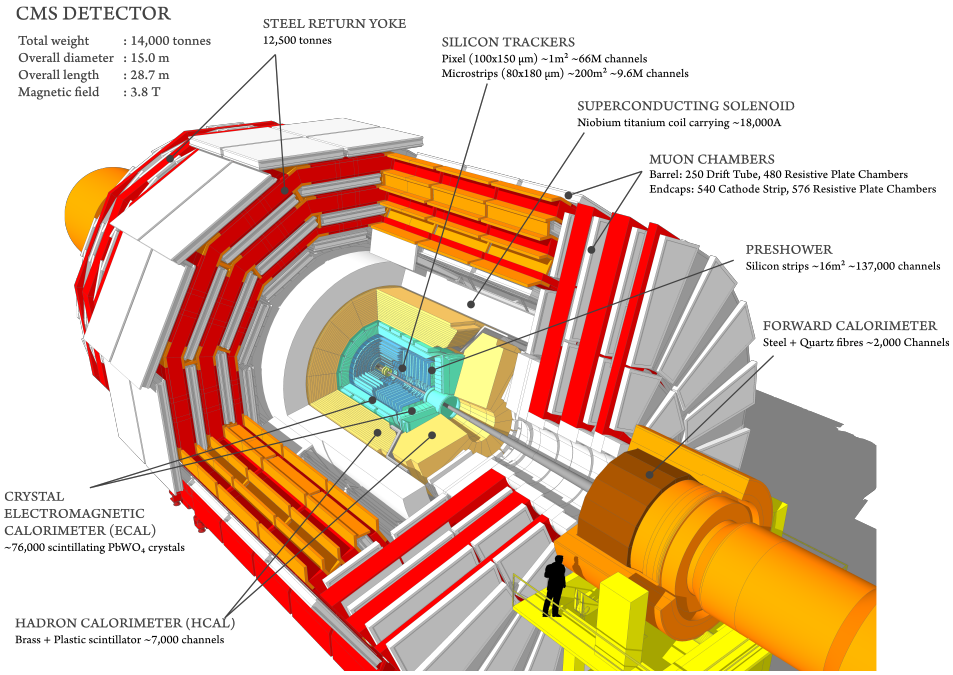


Figure 2.5 View of the CMS experiment and its subdetectors.

2.2 The Compact Muon Solenoid

The Compact Muon Solenoid experiment [52] is a general-purpose detector, placed in one of the LHC collision points. Its scientific program targets a wide range of physics analyses: from the search for the Higgs boson, to the precision measurements of SM phenomena, to searches for physics beyond the Standard Model. The detector has a cylindrical geometry, 21 m long, 15 m wide and 15 m high, and consists of multiple concentric layers of subdetectors, each one optimized for specific measurements or for detecting particular types of particles, as represented by the a cutaway view showed in Fig. 2.5. Starting from the interaction point and moving outward, the CMS subdetectors and components are ordered as follows. First there is the tracker, where the charged particle trajectories and the interaction vertex are reconstructed, electrons and photons are then absorbed in the electromagnetic calorimeter which detects their energy. Similarly charged and neutral hadrons can initiate their hadron shower in the electromagnetic calorimeter and get eventually absorbed by the hadronic calorimeter. The tracker and the calorimeters are enclosed within the solenoid magnet. Muons and neutrinos traverse the detector with little or no interactions. Muons are then detected in the most external detectors, namely the muon system, while neutrinos escape undetected. The lateral part and the circular faces of the CMS cylinder are also referred to as the barrel and endcaps, respectively.

The conventional reference frame for CMS is right-handed and centered in the interaction point (IP). It sees the z axis parallel to the beam line with counter clockwise direction, the y axis directed towards the surface and the x axis pointing to LHC ring center, as represented in Fig. 2.6. The angular variables are, with respect to the beamline, ϕ, azimuth angle, for the xy-plane and θ, polar angle, for

the zy-plane. The pseudorapidity, η, defined as

$$\eta = -\ln(\tan\frac{\theta}{2}), \tag{2.4}$$

is preferred to θ because, in the massless particle hypothesis, which is applicable in most of LHC hard interactions, it tends to the rapidity which is defined as follows:

$$y = \frac{1}{2}\ln(\frac{E - p_z}{E + p_z}), \tag{2.5}$$

where $c = 1$, E is the particle energy and p_z is the particle momentum p, projected along the beam axis. Transverse energy and momentum (p_T and E_T) are therefore defined as $p_T = p \cdot \sin\theta$ and $E_T = E \cdot \sin\theta$. The rapidity goes to 0 when the longitudinal component of the momentum goes to 0, and to $\pm\infty$ when the momentum is purely longitudinal.

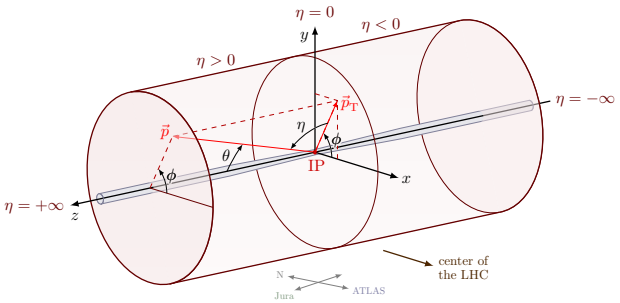


Figure 2.6 CMS coordinate system.

2.2.1 Solenoid magnet

The detector is built around a cylindrical solenoid magnet working as a coil of superconducting cable, cooled to -268.5°C, that generates a uniform magnetic field of 3.8 T at its center. The CMS magnetic field is used to bend the trajectory of the charged particles, hence to measure their momentum. In order to ensure good momentum resolution also in the forward region of the detector, a large enough length/radius ratio is also demanded. The length of the solenoid is then 13 m, while its diameter is 5.9 m, enough to contain the tracker and the calorimeters. In the external part of the solenoid the magnet field is returned by a saturated iron yoke, which also works as mechanical support of the detector. Figure 2.7 shows the designed magnetic field intensity and direction on a longitudinal section of the CMS detector.

2.2.2 Tracker

The CMS tracker [54] is the innermost component of the detector, located closest to the beam pipe and the interaction point. It consists of multiple layers of silicon detectors designed to reconstruct the trajectories of charged particles and the primary and secondary interaction vertices. The curvature of the particles' paths, measured

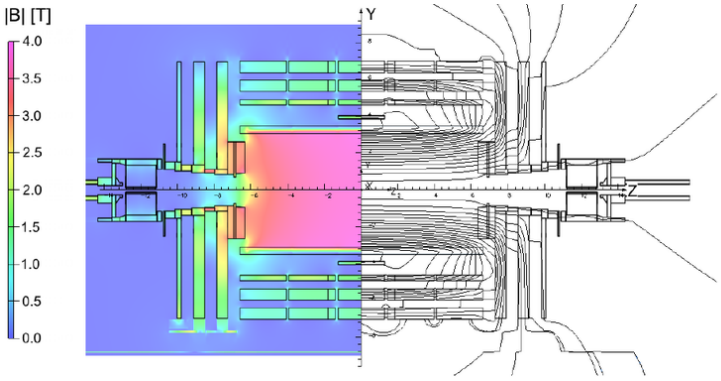


Figure 2.7 Design values of the magnetic field $|B|$ (left) and field lines (right) for a longitudinal section of the CMS detector, for a central magnetic flux density of 3.8 T. Each field line represents a magnetic flux increment of 6 Wb [53].

through successive hits in the silicon sensors, allows for precise determination of their momenta. The tracker is divided into two subsystems: the Pixel Detector and the Silicon Strip Detector.

The Pixel Detector is positioned at the very core of CMS, where it faces the highest particle flux. It comprises four cylindrical barrel layers and three endcap disks made of hybrid pixel sensors with a typical pixel size of 100 $\times 150 \mu m^2$. The small pixel dimensions provide excellent spatial resolution, essential for distinguishing nearby tracks and accurately reconstructing vertices. The pixel size results from a trade-off between spatial resolution and the number of required readout channels. The pixel detector extends up to a transverse radius of 16 cm, while the endcap disks are placed within 52 cm of the interaction point. Figure 2.8 reports a schematic representation of the Pixel Detector.

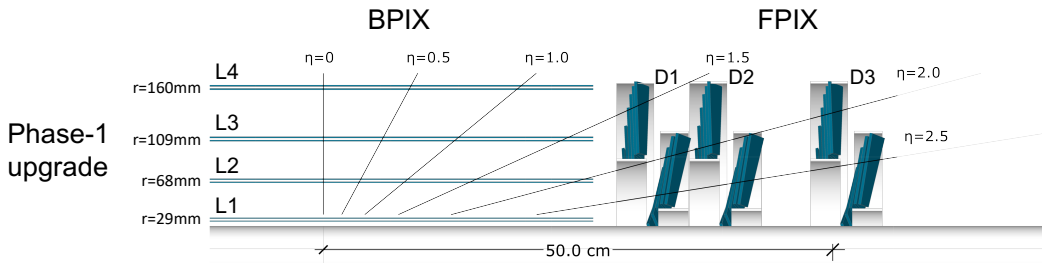


Figure 2.8 Schematic view of the CMS pixel detector, the gray shadow below represents the beam pipe edge [52].

The Silicon Strip Detector surrounds the Pixels and it is composed of four inner barrel (TIB) layers enclosed by two inner endcaps (TID), each made of three small disks. More externally the outer barrel (TOB), surrounding both the TIB and the TID, consists of six concentric layers. Finally two endcaps (TEC) close off the tracker on either end. The system extends up to a transverse radius of 1.3 m, covering the pseudorapidity region $|\eta| < 2.5$. The silicon strips have thicknesses between 320μm and 500 μm, with a pitch varying from 122 μm to 205 μm depending on the detector

region.

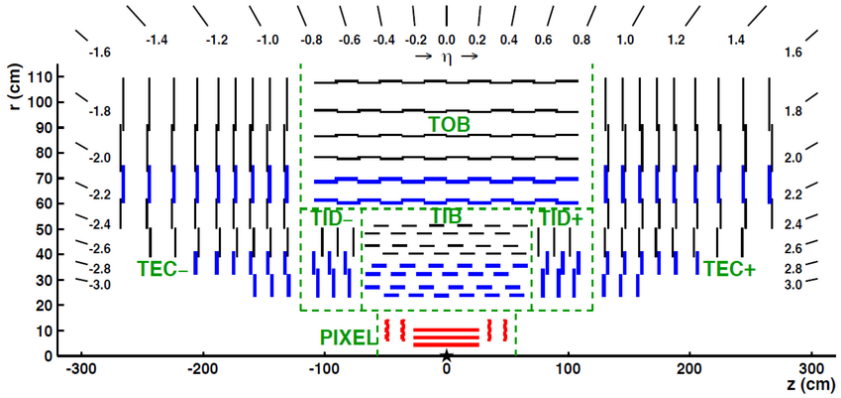


Figure 2.9 Schematic view of the upper half of the CMS inner tracker [55].

As charged particles traverse the tracker, the pixel and strip sensors generate tiny electric signals that are amplified and read out. Silicon technology was chosen for its high granularity, radiation hardness, and redundancy, enabling robust pattern recognition and efficient track reconstruction. The middle and outer tracker regions employ strips with coarser segmentation, as the particle density decreases with increasing radius. The tracker's material budget must be carefully controlled to minimize the degradation of particle energy measurements in the calorimeters. The thickness of the tracker corresponds to approximately 0.4 radiation lengths (X_0) at $\eta = 0$, increasing to about 1 X_0 at $|\eta| = 1.6$, and decreasing again to around 0.6 X_0 at $|\eta| = 2.5$. The spatial resolution of the tracker is about 10–15 μm, leading to a momentum resolution better than 2% (6%) in the barrel (endcap) for muons with 20GeV$< p_T <$100GeV, and better than 10% for transverse momenta up to 1 TeV.

2.2.3 ECAL: the electromagnetic calorimeter

The energy of photons and electrons is measured by the homogeneous and hermetic elecromagnetic calorimeter (ECAL) [56]. The CMS hadronic calorimeter (HCAL) is made of lead tungstate ($PbWO_4$) scintillating crystals. The high density (8.28 g/cm^3), short radiation length (8.9 mm), and small Moliere radius (22 mm) of the $PbWO_4$ enabled the construction of a compact calorimeter with fine lateral granularity and radiation resilient. The oxygen in its crystalline form makes the $PbWO_4$ highly transparent and it produces fast scintillation light, when electrons and photons pass through it, with about 80% of the light collected within 25 ns. However, the relatively low light yield requires the use of photodiodes with an intrinsic gain, able to operate inside the high magnetic field of CMS. The ECAL comprises a central barrel and two endcaps, a schematic view of the detector geometry and components is represented in Fig. 2.10.

The barrel, about 6 m long and with an inner radius of about 1.3 m, covers the pseudorapidity region $|\eta| < 1.48$. It consists of 36 elements, known as supermodules (SM), arranged in two rings of 18, each covering half the length of the barrel. The

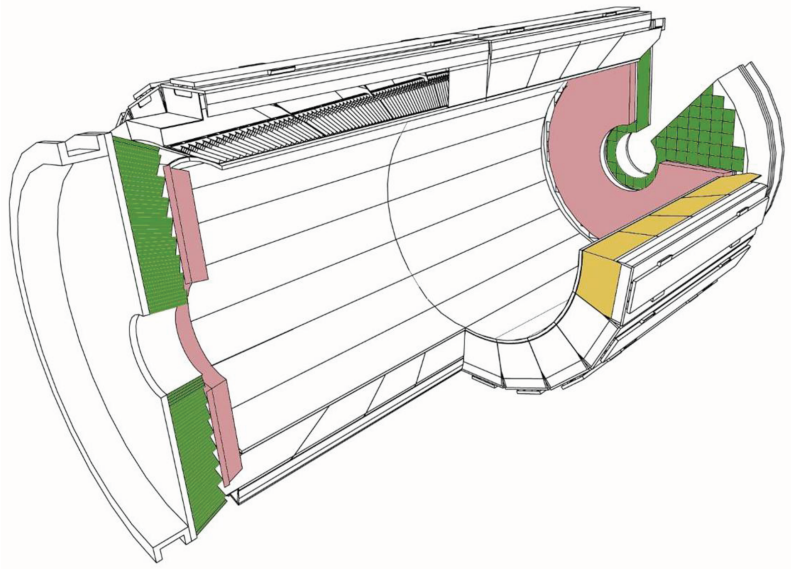


Figure 2.10 A cutaway view of the CMS ECAL. The barrel supermodule is highlighted in yellow, the endcap calorimeter and preshower are highlighted in green and pink, respectively.

62 100 crystals in the barrel are arranged in a quasi-projective geometry, oriented toward the nominal interaction point, with their axis tilted by approximately 3° in both η and ϕ. The crystals come in 17 different shapes, depending on their position along η. In the barrel, the emitted light is detected with avalanche photodiodes (APD) operating at a gain of 50, corresponding to a bias voltage between 350 and 420 V.

The endcap region is located at a distance of 3.14 m from the interaction point and covers the absolute pseudorapidity range from 1.48 to 3.0. Each endcap is made of two Dee-shaped sections hosting $PbWO_4$ which are placed according to an off-pointing pseudo-projective geometry. Because of the high radiation levels in the endcaps all materials used in this region must tolerate very large doses and neutron flux, therefore vacuum phototriodes (VPTs) are used for the scintillation light detection. In front of the ECAL endcaps, closer to the interaction point, the preshower covers a pseudorapidity range from $|\eta|$ between 1.65 and 2.61. Its main function is to discriminate between single γ and π^0 decaying in collimated photons. It contains lead converters followed by detector planes of silicon strips with a pitch of below 2 mm. The impact position of the electromagnetic shower is determined with an accuracy typically of 300 μm at 50 GeV. The fraction of energy deposited in the preshower (typically 5% at 20 GeV) decreases with increasing incident energy and it is used to correct the energy measured in the crystals.

The ECAL barrel energy resolution was measured at beam tests before the beginning of CMS Run 1, resulting in the following function of the incident electromagnetic particle energy (E), expressed in GeV:

$$\frac{\sigma}{E} = \frac{0.028}{\sqrt{E}} \oplus \frac{0.12}{E} \oplus 0.30\%, \tag{2.6}$$

where the $\oplus$ indicates the quadratic sum and the numerical values represent, respectively, the stochastic, the noise and the constant term [57]. In ECAL operation during Run 3 the relative energy resolution is measured to range between 1.5% and 4% in the barrel and between 4% and 6% in the endcaps using $Z \to ee$ events (see Fig. 2.11).

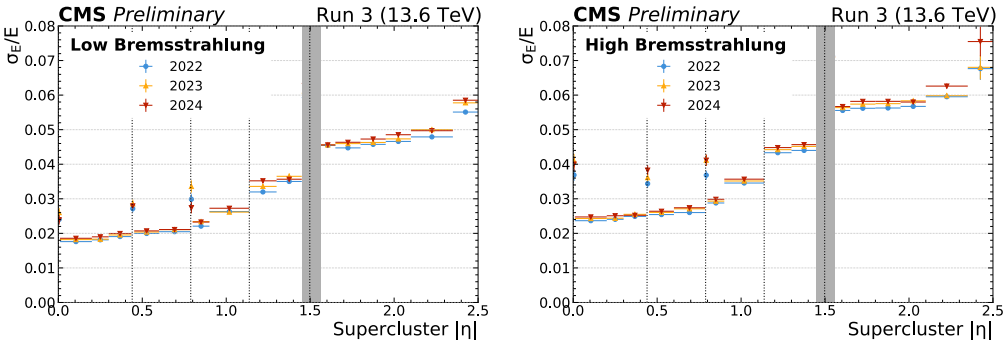


Figure 2.11 ECAL resolution measured in Run 3, using $Z \to ee$ events, for electrons having low bremsstrahlung emissions (left) and for an inclusive selection of electrons (right). The horizontal bars show the bin width [58].

In view of the upcoming HL-LHC operations the CMS ECAL will undergo a full upgrade of its readout electronics. A dedicated discussion of the prototype test and characterization is presented in Appendix A.

2.2.4 HCAL: the hadronic calorimeter

The HCAL [59] measures the energy of hadrons produced in pp collisions in combination with the ECAL. The design of HCAL originates from the necessity of containing the hadronic shower within the limited space between the ECAL and the solenoid magnet, and as opposed to electromagnetic calorimeter, the HCAL is a sampling calorimeter formed by layers of absorbing material alternating to layers of active material.
It is composed of four major subdetectors: the hadron barrel (HB), the hadron endcap, the hadron forward (HF) and the hadron outer (HO) calorimeters. The HB and HE cover the pseudorapidity regions $|\eta| < 1.392$ and $1.305 < |\eta| < 3.0$, respectively. The HO provides a measurement of the shower tails in the region $|\eta| < 1.26$, located outside the solenoid magnet, and the HF covers $3.0 < |\eta| < 5.2$. A schematic view of the HCAL detector is reported in Figure 2.12.

The HB and HE primarily use brass as absorber, except for the inner and outer layers of HB, which are constructed from steel. Brass has been chosen because it has a reasonably short interaction length, good mechanical properties and is non-magnetic, so it does not interfere with the CMS magnetic field. Plastic scintillators layers are used as active material to measure the energy deposits and that produces a rapid light pulse, collected by wavelength shifting optical fibers for fast transmission. The scintillation light is then detected by silicon photo-multipliers (SiPMs) that were installed between the 2017 and 2018, replacing the pre-existing hybrid photodiodes,

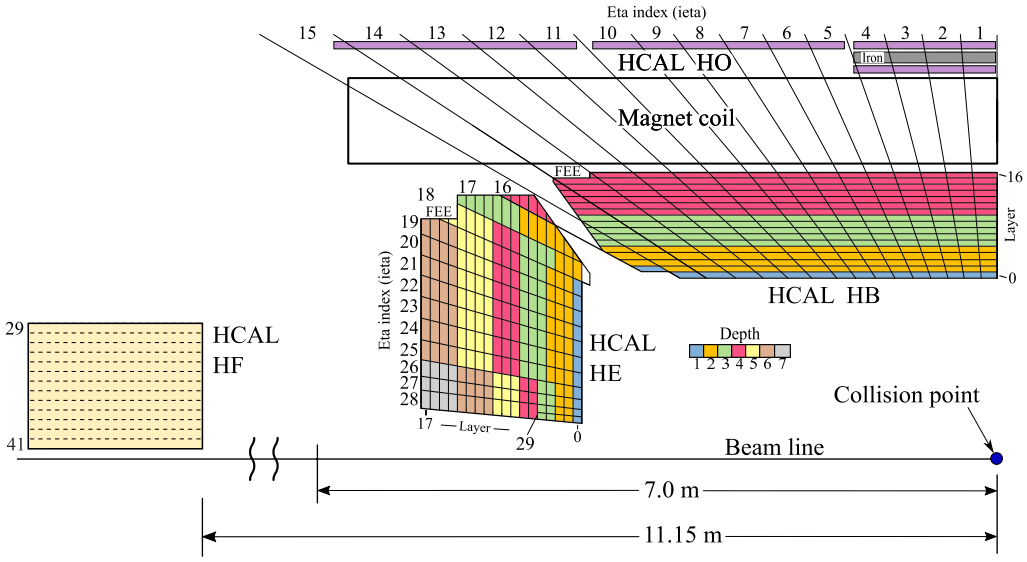


Figure 2.12 The longitudinal and transverse HCAL segmentation for Run 3. Within a tower, layers with the same color are routed to the same SiPM. [52].

as part as the HCAL Phase 1 upgrade.
The HF is a 1.65 m long sampling calorimeter with steel absorber and elements which produce Cherenkov light collected by photomultiplier tubes.
Being the particle energy measured only in the active layers a sampling calorimeter generally has lower energy resolution than an homogeneous calorimeter, but it is also more compact in order to absorb the radiation from hadronic showers.

The relative energy resolution for charged pions for the combined ECAL and HCAL, including barrel and endcap, is measured at beam tests as

$$\frac{\sigma}{E} = \frac{84.7\%}{\sqrt{E}} \oplus 7.6\%, \tag{2.7}$$

where the noise term is negligible (see Fig. 2.13). The HCAL energy resolution is much larger than the ECAL resolution (Eq. 2.6), therefore it dominates in the combination of HCAL and ECAL measured energies.

2.2.5 Muon system

As the name Compact Muon Solenoid suggests, muon detection is a central feature for the CMS experiment. Muons are the only charged particles that pass through the calorimeters without being absorbed and get detected by the muon system [60], embedded in the iron yoke, where the return flux of the magnetic field is of about 1.5 T. The return field bends muons in the opposite direction with respect to their bending in the tracker. These systems provide efficient detection of muons over a large range of pseudorapidity. The location in the magnetized steel behind the calorimeters and solenoid ensures a low probability of penetration to the muon detectors by particles other than muons and neutrinos.

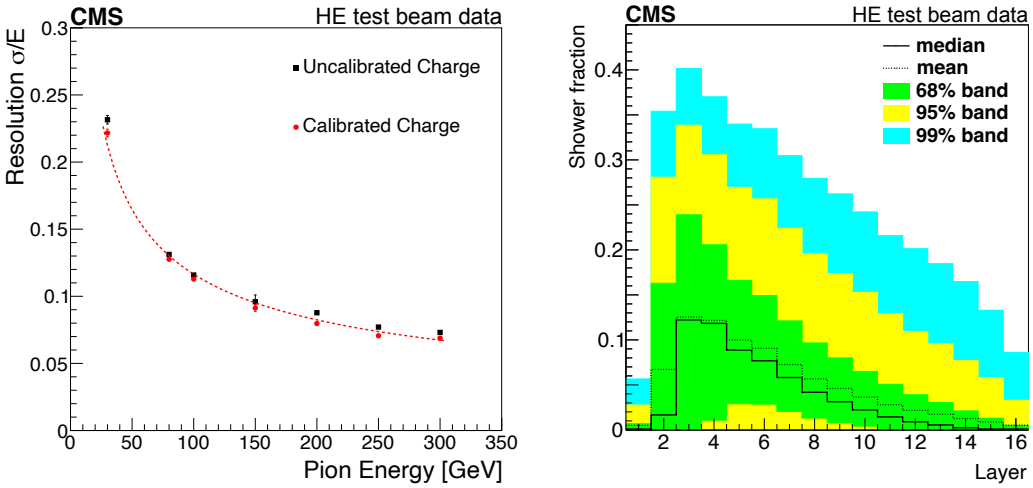


Figure 2.13 Left: HCAL energy resolution as a function of the pion beam energy measured at test beams prior to the data taking period, shown with (red bullets and dashed line) and without (black bullets) the channel calibrations. Right: longitudinal shower profile with bands containing 68%, 95%, and 99% of all events for each layer are shown. [52].

The muon system uses different technologies in different regions of the detector, as reported in Fig. 2.14: the drift tubes (DTs), the resistive plate chambers (RPCs), the cathode strip chambers (CSCs) and the gas electron multiplier (GEM) chambers.

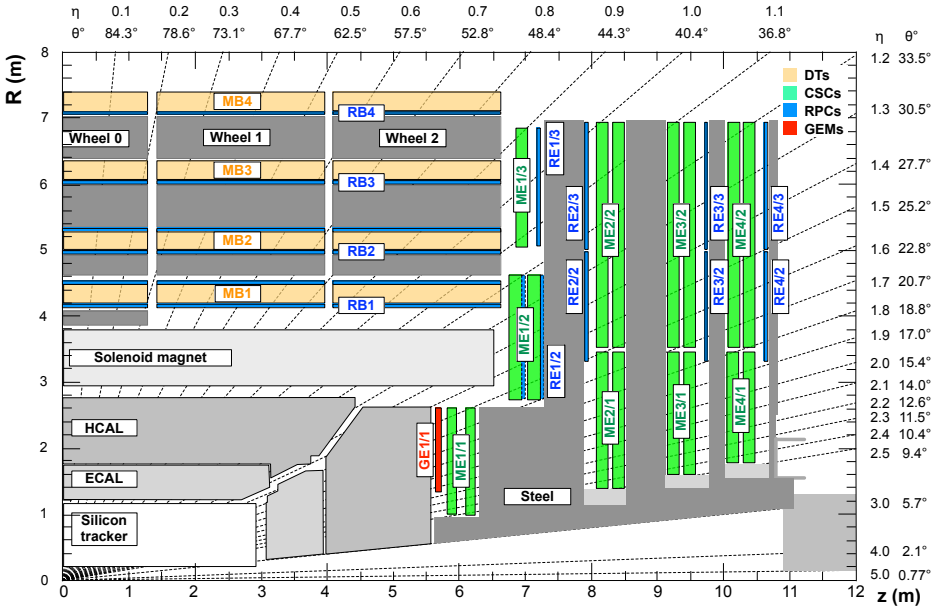


Figure 2.14 Longitudinal view of a quarter of the muon system in the zR-plane

The DTs cover the barrel region at $|\eta| < 1.2$, and are organized in four concentric shells, namely stations (MB1-4). Each station consists of eight layers of 2×2.5 m^2

drift containing drift tubes with a wire in a gas mixture of Argon (85%) and CO_2 (15%). When a muon passes through the single DT volume it ionizes the gas mixture and the released charges drift towards the wire (55 μm ns^{-1}). This mechanism allows for the position measurement of the muon with a spatial resolution of 100 μm (150 μm) in the r-ϕ (z) direction and time resolution around 2 ns.

The CSCs are multi-wire proportional chambers with a finely segmented cathode strip readout, instrumenting the endcaps in the region $0.9 < |\eta| < 2.4$. Each CSC module consists of six layers filled with Argon (45%) and CO_2 (50%) and CF_4 (10%) and each one having a plane of radial cathode strips and a plane of anode wires running perpendicular to the strips. The strips run radially in order to measure the muon position in the bending plane, the one perpendicular to the colliding beams axis, while the anode wires are oriented in azimuth and provide a coarse measurement in the radial direction. The CSCs are designed to be resilient to the magnetic field inhomogeneities in the forward region, maintaining an overall position resolution between 40-150 μm.

The RPCs are installed in both barrel and endcaps covering a pseudorapidity region below 1.6. They provide a complementary and trigger dedicated detector system with an excellent time resolution ($\sigma_t < 1.5$ ns) at a cost of a worse spatial resolution (0.8-1.2 cm), which is however compensated by the DTs and CSCs. Thanks to this excellent timing they provide a precise bunch-crossing assignment to the reconstructed muon tracks. RPCs consist of two parallel plates, a positively-charged anode and a negatively-charged cathode, both made of a very high resistivity plastic material and separated by a gas volume and operating in avalanche mode.

The GEM detectors were installed before the start of Run 3 in the region covering $1.55 < |\eta| < 2.18$ to enhance the track reconstruction and trigger capabilities of the endcap muon spectrometer, in view of the HL-LHC period. At present the first of the three foreseen GEM rings (GE1/1) is installed inside CMS, the others will be included during the long shutdown precedent to the beginning HL-LHC. Each endcap GE1/1 detector consists of 36 double-layered triple GEM chambers located just in front of the first CSC station. At the position of the first muon station muons experience the largest bending and due to increasing background rates at high $|\eta|$ the trigger rate in this region is large and difficult to mitigate. The insertion of the GEM chambers aim to improve the muon momentum reconstruction in the forward region to reduce the fake rate and pileup effects. Thus mitigating the L1 trigger rate thanks to excellent spatial and timing resolution (250–500 μm and <10 ns per layer) and to the GEM technology which is able to withstand rates up to 1 MHz/cm^2.

2.3 Event reconstruction

The layered structure of the CMS apparatus enables the identification of particles and the reconstruction of their kinematic properties by combining the information from the various subdetectors. The highly segmented tracker, the fine-grained electromagnetic calorimeter, and the hermetic hadron calorimeter, together with the strong and homogeneous magnetic field and the precise muon spectrometer, make CMS particularly well suited for the Particle Flow (PF) reconstruction algorithm [61].

The PF reconstruction provides a global description of each event in terms of distinct physics objects: charged and neutral hadrons, electrons, isolated photons, muons, and missing transverse energy. The PF algorithm is based on a set of relatively simple principles: charged particles are tracked and reconstructed in the inner tracker, electromagnetic particles are absorbed in the ECAL, hadrons deposit part of their energy in the ECAL before being absorbed in the HCAL, and muons and neutrinos traverse the inner detector with minimal interaction, leaving the steel magnetic yoke almost unperturbed. Muon trajectories are measured in the outer muon system, while neutrinos, which do not interact with the detector material, are inferred from an imbalance in the total transverse energy. A significant improvement in event reconstruction is achieved by correlating the signals from all detector layers and combining this information to determine the properties of the individual particles, as illustrated in Figure 2.15. In the next sections the event reconstruction via the PF algorithm is generally described with a particular focus on the muon reconstruction, for the purposes of this work.

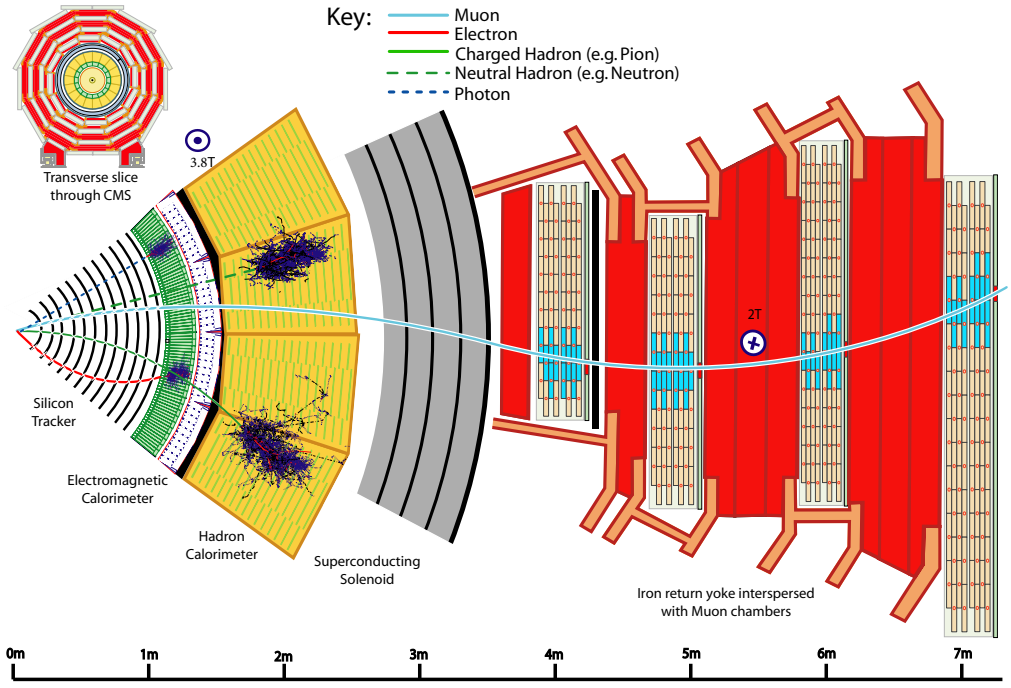


Figure 2.15 Representation of the particles interaction inside the CMS detector [61].

2.3.1 The particle flow algorithm

A given particle is generally expected to produce different interactions across the various CMS subdetectors. Those interactions are grouped in PF elements of three kinds: reconstructed trajectories of charged particles in the inner tracker (PF tracks), energy clusters in the calorimeters (PF clusters), and tracks reconstructed in the muon spectrometer. The particle reconstruction then proceeds through a linking algorithm that connects nearby PF elements from different subdetectors,

grouping them into PF blocks. Finally, high level physics objects are identified and reconstructed within each block.

PF tracks are reconstructed using a combinatorial track finder based on a Kalman filter, applied through several successive iterations, an approach referred to as iterative tracking. Each iteration consists of four main steps:

- **Seed generation:** a few hits in the inner tracker, compatible with a charged-particle trajectory, are used to initialize the trajectory extrapolation;
- **Pattern recognition:** a Kalman filter extrapolates the seed trajectory along the expected path, searching for additional hits in subsequent tracker layers to be associated with the particle's track;
- **Track fitting:** the particle's parameters, e.g. origin, transverse momentum, and direction, are determined from the full set of associated hits;
- **Track selection:** quality criteria are applied to the track fit χ^2 and to its compatibility with one of the reconstructed vertices. These criteria are adapted according to the track p_T, $|\eta|$, and number of hits.

The main differences among the various iterations lie in the seed generation strategy and in the quality criteria applied to the reconstructed tracks. The early iterations prioritize easily detectable tracks, such as high-p_T tracks originating close to the interaction point, to efficiently identify prompt signatures. After each iteration, the hits used in accepted tracks are masked and excluded from subsequent iterations.

The PF clustering is performed separately in the ECAL barrel and endcaps, and in the HCAL barrel and endcaps. A dedicated clustering algorithm was developed for the PF event reconstruction, aiming to achieve high detection efficiency even for low-energy particles and to separate nearby energy deposits. The clustering serves multiple purposes:

- detect and measure the energy and direction of stable neutral particles, such as photons and neutral hadrons;
- separate neutral particles from charged hadrons;
- reconstruct and identify electrons together with their associated bremsstrahlung photons;
- improve the energy measurement of charged hadrons whose track parameters are not precisely determined, e.g. for low-quality or high-p_T tracks.

Cluster building begins with the identification of seed cells, defined as calorimeter cells with an energy (E_T) larger than a predefined seed threshold and greater than their neighboring cells energy. Topological clusters are then formed around the seeds by aggregating adjacent cells (sharing at least one corner) with energy above a cell threshold set to twice the noise level. Finally, an expectation–maximization algorithm based on a Gaussian mixture model is applied within each topological

cluster to reconstruct the energies and the positions of the cluster component, assuming that the energy deposits in the individual cells of the topological cluster arise from Gaussian energy deposits centered in the seed cells.

Each particle typically gives rise to several PF elements across the various CMS subdetectors, which are ultimately connected through a linking algorithm to reconstruct the physics objects.
The linking algorithm tests pairs of elements that lie within a limited region in the (η, ϕ) plane, and the quality of each link is quantified by the angular distance between the two. The algorithm then produces PF blocks, in which elements are connected either directly or indirectly through common intermediate elements. Muons are reconstructed independently in the tracker and the muon system, as described in Section 2.3.2. Electrons are reconstructed with a dedicated tracking algorithm that accounts for bremsstrahlung radiation and are subsequently matched to PF clusters. Charged hadrons result from linking a track with associated PF cluster in both the ECAL and the HCAL, while PF clusters not associated with any track are identified as isolated photons.
A link between a track and a calorimeter cluster is established by extrapolating the track from its last measured hit to the ECAL, at a depth corresponding to the expected maximum of a typical longitudinal electron shower profile, and to the HCAL at a depth corresponding to one interaction length. The link distance is defined as the distance between the extrapolated track position and the cluster position in the (η, ϕ) plane. When multiple HCAL clusters are linked to the same track, or multiple tracks are linked to the same ECAL cluster, only the link with the smallest distance is retained. Cluster-to-cluster links are also established between HCAL and ECAL clusters when the position of a cluster in the more granular calorimeter (preshower or ECAL) lies within the envelope of the cluster in the less granular one (ECAL or HCAL).

Within each PF block, particles are identified in a hierarchical order. After each particle is reconstructed, its associated PF elements are removed from the block before proceeding. The reconstruction sequence prioritizes the muons followed by the electrons and energetic isolated photons. After removing the tracks with a transverse momentum uncertainty larger than the calorimetric energy resolution, the remaining elements in the block are cross-identified as charged hadrons, neutral hadrons, or photons originating from parton fragmentation, hadronization, or jet components.

2.3.2 Muon reconstruction and identification

The CMS muon spectrometer provides excellent performance in muon identification and reconstruction over the full detector acceptance, with an efficiency of about 99%. The upstream calorimeters absorb most other particles, except neutrinos, thereby ensuring high purity.

Muon tracking combines information from the muon system and the inner tracker to achieve precise momentum measurements. Depending on how the signals from these two subdetectors are combined, the high-level reconstructed muon objects are classified into three categories:

- **Standalone muons** are reconstructed using only information from the muon chambers. Energy deposits in the DT and CSC detectors are clustered to form track segments, which are then used as seeds for Kalman-filter-based tracking within the muon spectrometer, collecting all DT, CSC, and RPC hits compatible with the trajectory. To improve the momentum resolution, a beam-spot constraint can be applied in the fit.

- **Global muons** are built starting from a standalone muon track, matched to a inner tracker track, referred to as the inner track. The standalone and inner tracks are propagated onto a common surface, typically the innermost muon station, and the most compatible pairs is selected. For each matched pair, a combined track fit using all hits from both tracks is performed, again using a Kalman filter, to form the global-muon track. The global fit improves the momentum resolution for muons with p_T >200 GeV.

- **Tracker muons** are seeded by tracks with $p_T > 0.5$ GeV and total momentum exceeding 2.5 GeV. These tracks are extrapolated to the muon system surface and identified as tracker muons if they match at least one muon segment within a radius of 3 cm, or if the ratio of the track–segment distance to its uncertainty is smaller than 4.

Global and tracker muons sharing the same inner track are merged into a single candidate. Standalone-muon tracks have a worse momentum resolution and a higher contamination from cosmic muons compared to global and tracker muons. The global-muon reconstruction is designed to achieve high efficiency for muons traversing more than one muon detector station, i.e. typically for muon with $p_T > 10$ GeV. At lower momenta, multiple scattering in the steel return yoke prevents muons from crossing several detector layers. In this region, the tracker-muon reconstruction is more efficient, as it requires only a single matched segment in the muon system.

If a hadronic shower is not fully contained in the HCAL and reaches the muon system, it can be misidentified as a muon, a phenomenon referred to as punch-through. To reduce such contamination, the PF muon reconstruction applies a set of quality requirements to muon tracks. Isolated global muons are first identified by examining tracks and calorimeter deposits within an angular distance $\Delta R < 0.3$. The scalar sum of the track p_T and calorimeter energy deposits must not exceed 10% of the muon p_T. For non-isolated global muons, a tighter selection, described later in this section, is applied, along with the requirement that at least three matching track segments are found in the muon detectors.

Based on the reconstruction information and track properties, PF muon candidates are classified according to the identification algorithm (ID) they satisfy: Loose-ID, Medium-ID, Tight-ID and Soft-ID. The different IDs are not mutually exclusive. The algorithm which are relevant for the present analysis are the Medium and Tight IDs:

- **Medium-ID** identifies either global or tracker muons with additional requirements on track quality. The inner track must have a fraction of valid hits above 80%. Global muons global-track $\tilde{\chi}^2$ must be below 3 and other requirements

are imposed to limit the track kink and maximize the inner track-muon segment compatibility. This selection maintains high efficiency for both prompt and muons produced in HF decays.

- **Tight-ID** selects global muons with $\tilde{\chi}^2$ of the global fit smaller than 10, at least one muon chamber hit included in the fit, and at least two muon stations are included in the muon segment. These requirements target the punch-through suppression. In addition more than 10 silicon hits, including at least one pixel hit and compatibility with a primary vertex within 2 cm are required for the inner track. These significantly reduce the efficiency for in-flight decays.

2.4 Trigger and data acquisition

Each proton–proton collision corresponds to an event in which the signals from the subdetectors are reconstructed into particles, and high-level physics information about them is stored for later analysis. The online event selection is performed by the CMS trigger system, a fundamental component that ensures the quality of the experiment's physics performance. Its purpose is to select only those events with potentially interesting physical content to be recorded on disk. At LHC energies, the pp interaction cross section is about 100 mb and is largely dominated by soft scattering processes, resulting in low-p_T signatures. The most relevant processes targeted by the CMS physics program, however, typically have cross sections below 1 nb. The trigger system is therefore designed to retain only events with a significant amount of energy, making fast decisions based on prompt signals from the CMS subdetectors. The CMS trigger operates in two consecutive decision levels: the Level 1 Trigger (L1T) and the High Level Trigger (HLT) detailed in Sec. 2.4.1 and 2.4.2, respectively.
Beyond the physics motivations, there are also technical constraints that require limiting the number of events that can be written to disk. The typical event size during Run 3 is approximately 1 MB, saving all the events would demand data throughput of about 40 TB s^{-1}, far beyond modern storage capabilities. Since Run 1, CMS has been trying to overcome such technical limitation developing novel trigger strategies featuring reduced event content (scouting) or delayed reconstruction (parking), as outlined in Sec. 2.5. These efforts aim to extend the experiment's physics reach toward low-energy signatures, potentially key to new physics discoveries, while remaining within technical limitations.

2.4.1 Level 1 trigger

The Level 1 trigger is an hardware-based decision system that processes events at the full LHC bunch-crossing rate of 40 MHz, reducing the output rate to 100-110 kHz. The L1T has a very limited time window, about 4 μs to determine whether an event should be retained or discarded. To enable such rapid decision-making, the L1T relies exclusively on coarse-grained detector information from the calorimeters and the muon systems. From these inputs, physics objects are reconstructed, along with calculations of energy sums. The physics object are forwarded to the Global Trigger (GT) that receives the detector information separately from the Calorimeter Trigger

System and from the Muon Trigger and produces the final Level-1 Accept (L1A) decision [62; 63; 64].

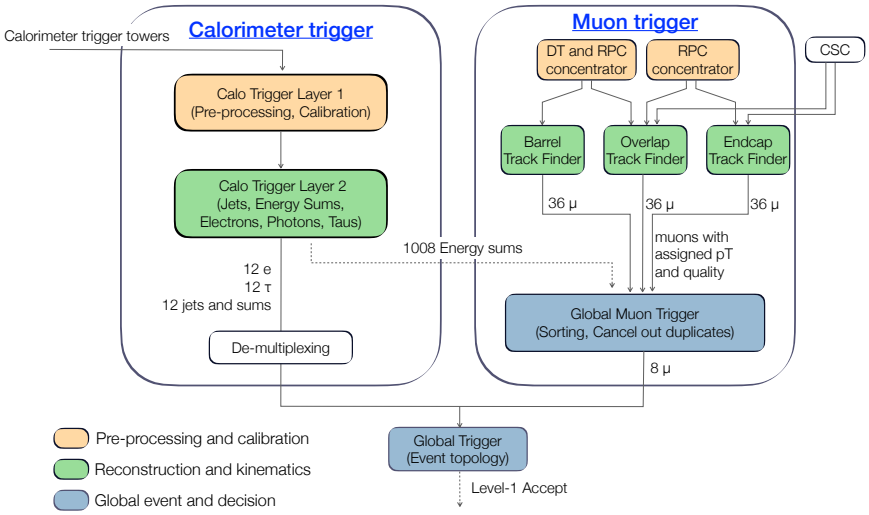


Figure 2.16 Simplified graphical representation of the L1 Trigger data flow.

The Calorimeter Trigger System receive as input energy sums of the trigger towers formed by the Trigger Primitive Generator (TPG) circuits on the ECAL and HCAL calorimeters. These deposits and their topology are processed together and used to generate preliminary physics objects such as photons, electrons, taus and jet candidates. The candidates and the transverse energy sums are forwarded to the global calorimeter trigger that selects the highest quality candidates for each category and transmits their information to the GT, together with the total transverse energy and total missing energy vector.

The muon trigger processes data from the RPCs, DTs and CSCs, with each L1 muon trigger systems having its own logic. Within DTs and CSCs, local trigger track segments are formed from the on-detector electronics and serves as trigger primitives (TPs). They are eventually sent to their own track finder (TF) unit which combines the segments to extrapolate the muon tracks measuring their p_T, accurate directional information is also provided by barrel DTs. The RPC strips are connected into trigger tracks by the pattern comparator trigger (PACT) to form trigger segments, the segment processors then connect the segments into tracks and calculate the p_T. Three muon TFs reconstruct muon tracks in three distinct pseudorapidity regions using TPs from the muon detectors. The barrel muon track finder (BMTF) receives inputs from DTs and RPCs in the barrel ($|\eta| \leq 0.83$), the overlap muon track finder (BMTF) uses DTs, CSCs, and RPCs in the overlap between barrel and endcap ($0.83 < |\eta| < 1.2$), while the endcap muon track finder (EMTF) takes inputs from CSCs, RPCs, and GEM in the endcaps ($1.2 \leq |\eta| \leq 2.4$). Each muon TF transmit up to 36 muons per bunch crossing to the global muon trigger, which sorts them, converts these tracks into the same η, ϕ and p_T scale, resolves duplicates and validates the muon sign. It then attempts to correlate the CSC and DT tracks with RPC tracks. It also correlates the muon tracks with an η-ϕ grid of quiet calorimeter towers to

determine if these muons are isolated. The final muons are sorted based on their initial quality, correlation and p_T and a maximum of eight muon tracks per bunch crossing is transmitted to the Global Trigger.
In Run 3 the BMTF algorithm was updated to improve the trigger efficiency for displaced muons. The new algorithm maintains efficiency above 90%, for displacement up to 40 cm, measured in the transverse plane and with respect to the beam-spot.

The final step of the Level 1 Trigger is the Global Trigger that, for every bunch crossing receives the trigger objects from the calorimeter and muon trigger and their associated information in terms of transverse energy E_T, transverse momentum p_T, position and direction and reconstruction quality. The event selection by the GT is prescribed by the trigger-menu. It is a sequence of paths, also referred to as L1-seeds, put in logical or. Each L1 seed requires that the event contains a specific signature in terms of physics objects and their energy. If an event is accepted, its data are transferred to a buffer for readout and subsequent processing by the HLT.

2.4.2 High Level 1 trigger

The HLT is a software-based algorithm running on a computing farm equipped with heterogeneous architectures, including both CPUs and GPUs, to exploit multi-threading and parallelism in event processing. The HLT workflow is organized around the concept of paths, each representing a sequence of algorithmic modules designed to reconstruct physics objects and apply selections targeting specific physics signatures. These steps are typically arranged in increasing order of complexity and reconstruction precision. The CMS HLT menu in Run 3 contains hundreds of paths, around 600 in 2022, covering both primary triggers used for physics analyses and auxiliary ones dedicated to calibration and efficiency measurements. The CMS data-acquisition system can sustain a throughput of about 10-15 GB/s, which constrains the HLT-accept rate to roughly 1.5 kHz. To regulate the output rate, some paths are prescaled, meaning that only a fraction of the events passing their selection criteria are recorded. This mechanism is mainly applied to calibration and low-energy paths. The HLT decision relies on a more complete event content compared to the L1T, with physics objects reconstructed using a simplified version of the PF algorithm optimized for timing constraints. To accelerate processing, the online PF reconstruction is applied only to limited detector regions, typically centered around the L1T objects that triggered the selection.

Until Run 2, online tracking was performed iteratively using hits recorded from the pixel and strip trackers, applying combinatorial Kalman Filter–based pattern recognition, similar to the offline PF reconstruction but with fewer iterations. In Run 3, the seeding and tracking strategy were significantly revised, now relying on a single global iteration seeded by pixel tracks reconstructed with the Patatrack algorithm [65]. Pixel tracks are required to have at least three pixel hits, transverse momentum p_T >0.3 GeV, and compatibility with a primary vertex candidate. The Run 3 online tracking demonstrates improved performance compared to that used in Run 2, achieving a substantially lower fake rate while maintaining high tracking efficiency, particularly for low-p_T tracks, as shown in Fig. 2.17. Online tracking algorithms are crucial for identifying and reconstructing muons at the HLT level.

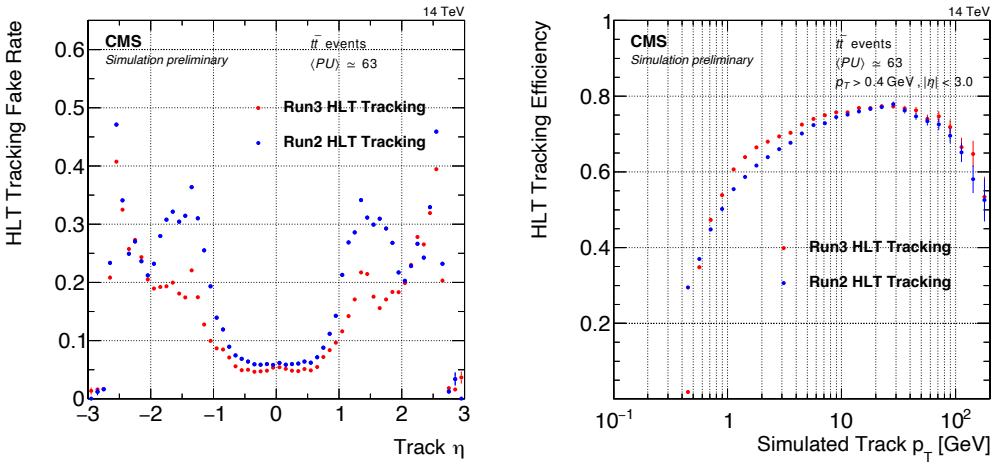


Figure 2.17 The fake rate as a function of the simulated track η (right) and the tracking efficiency as a function of the simulated track p_T (left) are compared for the Run 2 HLT tracking (blue) and the Run 3 HLT single-iteration tracking (red) [66].

Muon reconstruction in the HLT proceeds in two main stages, referred to as L2 and L3 reconstruction [67]. The L2 reconstruction is seeded by muon candidates that have passed the L1T selection and relies solely on hits in the muon systems. It is analogous to the standalone muon reconstruction described in Sec. 2.2.5, and refines the initial trajectory estimate provided by the L1T algorithm. The L3 reconstruction builds upon the L2 muons, performing an iterative track reconstruction to associate each L2 muon with an inner tracker track. Two complementary strategies are used:

- **Outside-In:** a track in the inner tracker is matched to an L2 muon, and a combined fit is performed. This method is conceptually similar to the global muon reconstruction in the offline PF algorithm.
- **Inside-Out:** the reconstruction starts from an inner tracker track, which is then extrapolated to the muon system and matched to a L1T muon, without performing a combined fit. This resembles the tracker muon reconstruction in the offline workflow.

In Run 3 the L2 reconstruction has been extended to include also hits from GEM detectors, further enhancing coverage in the forward region. After reconstruction, muon isolation is evaluated by considering the additional tracker tracks and calorimeter energy deposits around the muon.
A schematic representation of the detector regions involved in the HLT muon reconstruction and of its workflow is shown in Fig. 2.18.

2.5 CMS special data streams

The traditional paradigm for data acquisition and analysis at the LHC involves selecting proton–proton collision events online through a trigger system, storing them to disk in raw format, and subsequently reconstructing and analyzing them

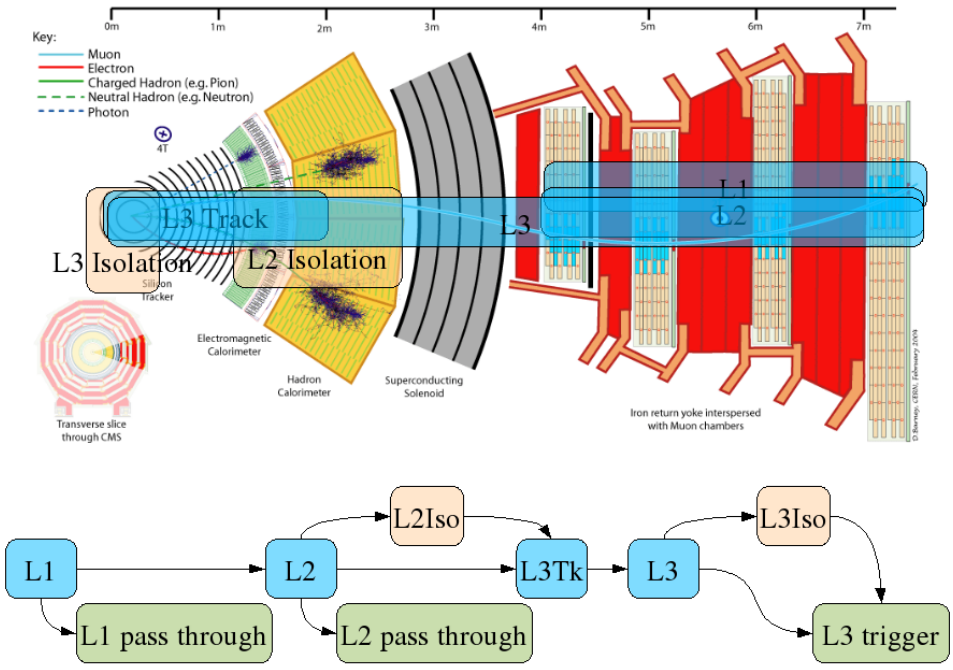


Figure 2.18 Schematic view of the online muon reconstruction in CMS.

offline. Already during Run 1 and Run 2, alternative trigger and data acquisition strategies were developed within the CMS Collaboration to extend the experiment's sensitivity to more challenging regions of phase space. These efforts have been significantly expanded in Run 3. The overall goal is to enhance the experimental sensitivity towards low energy signatures. Many new physics models predict the existence of low-mass particles with feeble couplings, while potential signs of new phenomena could be revealed by observing deviations from precise SM predictions in heavy-flavor decays. A notable example is the study of rare B-meson decays, which involve particles with momenta of only a few GeV. To maintain a manageable overall trigger rate, traditional data acquisition protocols typically require relatively high selection thresholds to suppress abundant SM background processes. As a result, potentially interesting signal events characterized by low energy and momentum may be inadvertently discarded.

The data-scouting and data-parking strategies were introduced to overcome two main limitations of the CMS data acquisition chain: the finite bandwidth available for writing data to permanent storage and the limited capacity for prompt (i.e., real-time) data reconstruction [68]. These techniques enable CMS to record and analyze a larger fraction of low-threshold events without saturating computing or storage resources. Throughout the years the CMS experiment has allocated increasing HLT rate for these special data-streams, as testified by the trends reported in Fig. 2.19. In the data-scouting the event content is stored on disk as it is reconstructed at the HLT, thereby reducing the data bandwidth through a minimal event format. Only high-level physics objects, such as jets, muons, electrons and photons, reconstructed online at the HLT are stored, while no raw detector data are saved for offline

reconstruction. The typical scouting event size is about 1.5kB, roughly ten times smaller than an event in the standard physics stream, allowing for a recording rate of up to 30kHz. The success of this approach relies on the excellent performance of the HLT reconstruction, which closely matches the quality of the offline algorithms and provides sufficient resolution to identify potential excesses or deviations indicative of new physics.
The data parking stream is better detailed in the next section.

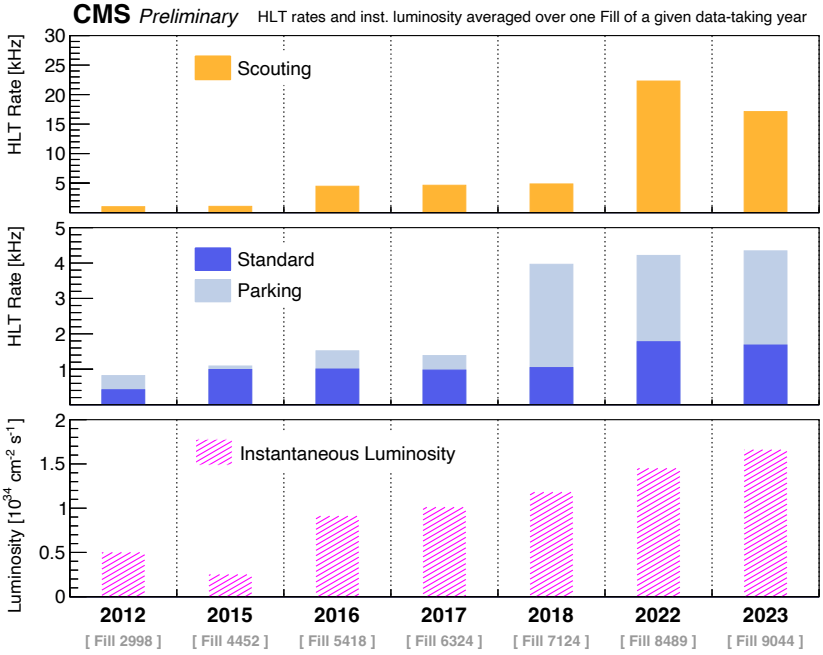


Figure 2.19 Evolution of the average output HLT-rates of the standard, parking and scouting data-streams during one fill of pp collisions for every data-taking year from 2012 to 2023. The average instantaneous luminosity delivered by the LHC during the given fill is also shown [69].

2.5.1 Parking

The data-parking streams overcome the limitations imposed by the finite data throughput when lowering trigger thresholds by deferring the offline reconstruction of the recorded events. Parking events retain the same data format as those in the standard physics stream. This strategy allows the event collection rate to be substantially increased beyond the capacity of the available computing resources for prompt reconstruction: data are transferred unprocessed, i.e. parked, to tape in raw format and reconstructed later, once sufficient computing resources become available.
Data parking was first introduced in 2012, however, a large-scale data-parking program was not launched until the end of Run 2 with the B-Parking. The primary physics motivation was to enable precision measurements of observables related to the so-called flavor anomalies, rare b-hadron decays that exhibit deviations from

Standard Model predictions [70]. In early 2018, the parking was used in combination with a new trigger strategy to select muons originating from b-hadron decays, thereby accumulating a high-purity sample of $b\bar{b}$ pairs, with lowest possible p_T. The B-Parking exploited the exponentially decaying luminosity profile during each fill: as the instantaneous rate decreased, additional bandwidth became available, allowing progressively looser L1 kinematic thresholds and thus increasing overall acceptance. This technique is referred to as the dynamic trigger turn-on.
In Run 3, the data-parking paradigm was completely redesigned and significantly expanded, benefiting from the enhanced in-situ computing resources of the CMS experiment, now equipped with both GPUs and CPUs. Since 2022, the LHC has been operating with updated machine parameters, adopting a luminosity-leveling mode in which $\mathcal{L}_{\text{inst}}$ is maintained at a constant value of approximately 2.0×10^{34} cm^{-2}s^{-1} and at fixed PU, for about six hours. This mode enables the rapid accumulation of large, uniform data samples under stable conditions favorable for analysis. After the leveling period, $\mathcal{L}_{\text{inst}}$ gradually decreases over an additional four-six hours. In the last period the activation of new paths occurs multiple times throughout the luminosity decay, maximizing the use of available bandwidth while keeping the L1 rate below 115 kHz. The total HLT output rate reaches up to 10-15 kHz, as represented in Fig. 2.20.

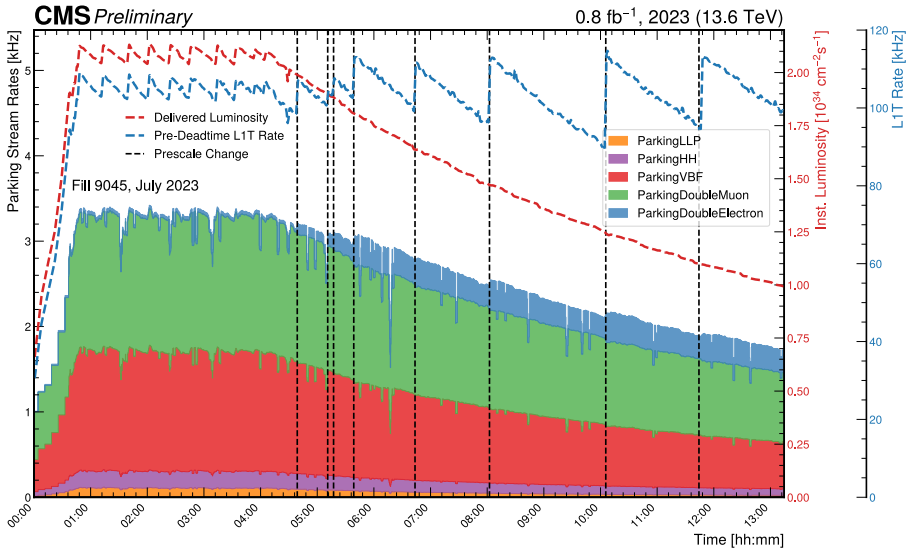


Figure 2.20 Rate of the parking streams in the HLT menu, LHC instantaneous luminosity (red dashed line), and L1T rate (blue dashed line) during a proton-proton fill in 2023. During the liminosity-deacy period looser double-electron triggers are progressively enabled (vertical black dashed lines), keeping the L1T rate high and increasing the HLT output rate of the corresponding parking stream [71].

Central to the Run 3 parking programs is the inclusive dimuon parking stream, which efficiently collect dimuon or multi-muon low-mass resonances with only a few L1 seeds undergoing the dynamic turn-on. This strategy stabilizes the data-taking conditions and simplifies offline analyses by reducing complications associated

with multiple trigger turn-ons. In the early years of Run 3 a dielectron parking stream was also active, targeting both flavor anomaly measurements and low-mass BSM searches. In 2024, the single-muon parking stream was introduced, partially reviving the concept underlying the 2018 B-parking initiative. These streams heavily rely on the dynamic trigger turn-on. Several additional triggers have since been incorporated into the data-parking streams, improving sensitivity across a broad range of physics processes beyond the scope of the B-physics. Dedicated parking streams now address specific signatures such as vector-boson fusion (VBF) Higgs boson production, Higgs boson pair production with τ leptons in the final state and BSM searches for long-lived particles (LLPs).

Chapter 3

Event selection for $\tau \to 3\mu$ reconstruction

This thesis presents the search for the lepton flavor violating process $\tau \to 3\mu$, reconstructed by selecting events with three muons from a τ lepton produced in $W \to \tau\nu$ and in $Z \to \tau\tau$ decays, using, for the first time, data collected by CMS during Run 3. The goal of the analysis is to measure the number of potential $\tau \to 3\mu$ signal candidates, consistent with a signal hypothesis, from the distribution of the τ lepton candidate invariant mass ($m_{3\mu}$), as the signal would manifest as a resonance around the τ mass (1.777 GeV).
For the first time τ leptons produced in Z decays are considered for the analysis, even though without a dedicated selection strategy. The signal extraction strategy discussed in this chapter is optimized to target the $W \to \tau(3\mu)\nu$ experimental signature, described by the diagram in Fig. 3.1, and its efficiency on $Z \to \tau(3\mu)\tau$ events will be discussed in the next chapters. The main τ lepton production channels in pp collisions are discussed in Sec. 3.1. The analysis strategy is optimized using the expected LFV signal modeled from simulation and applied on the data collected by the CMS experiment during the 2022 and 2023 years of data taking as discussed in Sec. 3.2. The online event selection relies on the combination of a double and triple muon trigger, and it is described in Sec 3.3. The experimental signature of the $W \to \tau(3\mu)\nu$ decay, and the set of criteria applied to the reconstructed final state particles in order to isolate potential signal events are discussed in Section 3.4 presenting the offline preselection strategy. It consists in reconstructing the $\tau \to 3\mu$ candidates by combining three muons that passed the online selection, with a constraint on their production vertex and applying a set of requirements on the muon kinematics and reconstruction quality. The W candidate is then reconstructed by combining the τ lepton with the missing transverse momentum of the event. After the preselection, the events are split into three categories for each year of data taking in order to maximize the combined sensitivity, as detailed in Sec. 3.5. Finally, in order to correct for possible discrepancies with data, a set of corrections is applied to the simulated samples of $W \to \tau(3\mu)\nu$ events and discussed in Sec. 3.6.

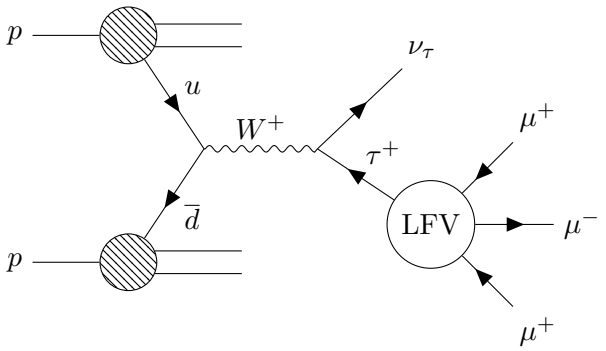


Figure 3.1 Feynman diagram of the $W \to \tau(3\mu)\nu$ decay.

3.1 Production of τ leptons at CMS

The dominant source of τ leptons at the LHC are hadrons containing c or b quarks, primarily D and B mesons, with a small contribution from HF baryons. Subdominant but important sources include vector boson decays, specifically $W \to \tau\nu$ and $Z \to \tau\tau$ decays. Table 3.1 reports the expected number of τ leptons produced in proton-proton collisions at $\sqrt{s}$ =13.6 TeV, for an integrated luminosity of 62 fb^{-1}. The expected yields are obtained from simulations using the event generator PYTHIA8 and with a total pp inelastic cross section of 80 mb [72].
Although τ leptons from HF decays dominate in terms of absolute production rate, they are significantly exposed to contamination by light hadrons, K and π mainly, reconstructed as muons. Due to the small mass difference between the parent meson and the τ, the final state muons are produced with high boost, predominantly in the forward direction, and with low transverse momentum. Consequently, a significant fraction lies outside the CMS detector acceptance (see Sec. 2.2.5), leading to a reconstruction efficiency of the order of 10^{-5}, including detector acceptance, for this production mode.
Differently, τ leptons from W and Z boson decays constitute only about 0.01% of the total τ yield, but benefit from a much higher reconstruction efficiency, around 20%. This arises from the fact that the final state muons have harder momentum spectra and are produced more centrally in the detector compared to those produced in HF decays. Furthermore, they are generally isolated, and in the case of W, they are associated with large missing transverse energy due to the energetic neutrino. Both isolation and missing transverse energy (MET) are powerful discriminants for background suppression. This makes the vector boson production mode a compelling channel to explore in the search for LFV $\tau \to 3\mu$ decays. A comparison of the production rates and kinematic properties for the different τ sources is shown in Fig. 3.2.

Considering the production rate, detector acceptance, and reconstruction efficiency, about 100 million τ leptons are expected to be reconstructed from both HF and vector boson decays in 62.6 fb^{-1} of pp collisions.

Table 3.1 Expected number of τ leptons produced at LHC at 13.6 TeV for an integrated luminosity of 62.6 fb^{-1} without considering the CMS detector acceptance. The yields are evaluated with leading order (LO) PYTHIA8, with total inelastic cross section of 80 mb.

Production	Decay		Yield ($\times 10^9$)
$pp \to c\bar{c} + X$	$D \to \tau\nu$	(95% D_s, 5% $D^{\pm}$)	7600
$pp \to b\bar{b} + X$	$B \to \tau\nu + X'$	(44% $B^{\pm}$, 45% B^0, 11% B_s^0)	3400
$pp \to b\bar{b} + X$	$B \to D(\tau\nu) + X'$	(98% D_s, 2% $D^{\pm}$)	1200
$pp \to W + X$	$W \to \tau\nu$		1.3
$pp \to Z + X$	$Z \to \tau\tau$		0.25

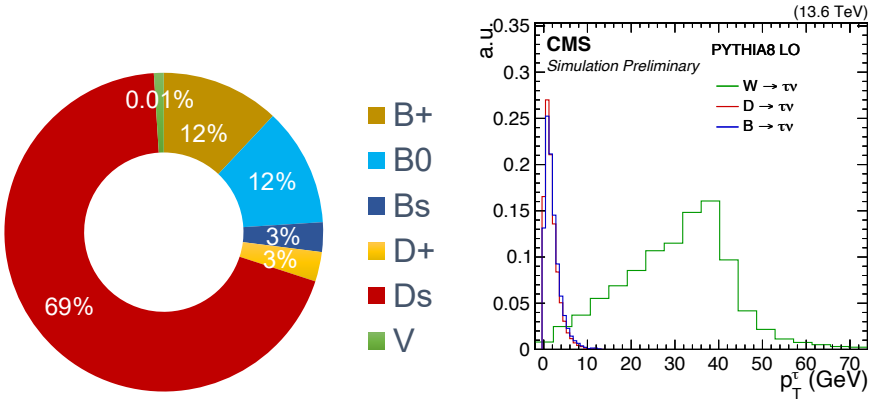


Figure 3.2 Left: Pie chart of τ lepton production modes in pp collisions. Right: generated τ p_T spectrum from PYTHIA 8 simulations, comparing the heavy-flavor and vector-boson decays.

3.2 Data and simulated samples

3.2.1 Simulated samples

The $\tau \to 3\mu$ signal in vector boson production channels is modeled using Monte Carlo (MC) events, generated with PYTHIA8 [72] and the CP5 tune [73] for the underlying event. The proton structure is described with the NNPDF3.1 PDF set [74]. The normalization of the signal is based on the W production cross-section measurement performed by CMS at 13.6 TeV [75], as detailed in Sec. 5.2.1.
Samples simulating $D_s^{\pm} \to \phi(\mu\mu)\pi^{\pm}$ decays, referred to as the control channel, are also used with the main purpose of validating the background rejection strategy and other aspects of the analysis. These events are generated with PYTHIA8, with the same settings as for the signal, while the decay chain is simulated using EVTGEN [76]. The $D_s \to \phi\pi$ decay is set to be a scalar decaying to a scalar and a vector, with prompt and non-prompt D_s (from b-hadron decays), while the decay $\phi \to \mu\mu$ is a vector to two leptons decay with radiative corrections.
In addition, simulated events for the SM process $W \to 3\mu\nu$ are used for background modeling, where in $W \to \mu\nu$ decays two additional muons originate from

a photon either from initial or final state radiation. Events are generated using MADGRAPH5_aMC@NLO [77] at next-to-leading order in perturbative QCD, using the NNPDF3.1 set of PDFs. PYTHIA8 with the CP5 tune is used for parton shower, fragmentation and hadronization.
In simulated events additional pp interactions occurring in the same or neighboring bunch crossings, referred to as pileup, are added to the simulation.
MC samples must reproduce the specific conditions of the CMS detector during data-taking. The detector response is included in the simulation using a detailed GEANT4-based description of CMS. Changes in the detector conditions during the run period are reflected in different MC campaigns to ensure consistency between data and simulation. In 2022, two distinct MC campaigns were produced to account for the shutdown of part of the ECAL endcap that occurred during the run. These are referred to as 2022preEE and 2022EE, representing conditions before and after the incident, respectively. In 2023, two further MC campaigns were generated, 2023preBPix and 2023BPix, to reflect a change in the barrel pixel tracker acceptance occurred during the year. It was checked that the $\tau \to 3\mu$ signal kinematics is negligibly affected by these detector changes within a given year. Therefore, unless specified, the 2022 and 2023 periods are treated without distinguishing between the sub-periods in the following discussion.

3.2.2 Double-muon parking dataset

The analysis presented in this work uses proton-proton collision data, at a center-of-mass energy $\sqrt{s} = 13.6$ TeV, collected with the CMS detector during 2022 and 2023. The two datasets correspond to an integrated luminosity of 34.5 fb^{-1} and 27.7 fb^{-1}, respectively, for a total of 62.2 fb^{-1}.

The analysis exploits the special double muon parking data-stream that was first introduced in the CMS data taking since the beginning of Run 3. The parking strategy enables the possibility to explore low-energy signatures without exceeding the data throughput limitation, as detailed in Sec. 2.5.1. Events collected with the parking strategy are not processed immediately by the global event reconstruction, instead, they are temporarily stored in local buffers at the CMS experimental site and reconstructed as soon as sufficient computing resources become available. Thanks to the excellent computing performances, during the 2022 and 2023, data were reconstructed almost promptly with minimal parking time of typically few days, as those of the standard physics streams.

3.3 Trigger selection

The events collected through the double-muon parking data stream are selected online by a collection of double and triple muon triggers targeting low mass resonances. The events entering the analysis presented here are selected online combining the requirements of an inclusive double muon HLT path and those of two triple muon paths specifically designed for the $\tau \to 3\mu$ signature. The path proper names are:

- `HLT_DoubleMu_4_3_LowMass`

- `HLT_Tau3Mu_Mu7_Mu1_TkMu1_IsoTau15`
- `HLT_Tau3Mu_Mu7_Mu1_TkMu1_IsoTau15_Charge1`

They have been active during the full 2022 and 2023 data-taking period. The first path, also referred to as the double-muon path in the following, requires the presence of at least two muons, and it was first introduced in the Run 3 HLT menu while the other two, inclusively referred to as the tri-muon paths, select events with at least three muons and are optimized to target the $W \to \tau(3\mu)\nu$ isolated signature specifically. They were inherited by the Run 2 menu. The online event selection for this analysis is obtained by the logical `OR` of the three paths. The tri-muon paths are more restrictive on the number of muons allowing for p_T threshold as low as 1 GeV for the sub-leading and trailing muon. Differently, in the double-muon trigger the requirement on the isolation and the number of online muon is relaxed compensating with tighter p_T thresholds on the two muons. A more detailed description of the trigger requirements is provided in the following sections. For a visual representation of the topological quantities entering the online selection, refer to the sketch in Fig. 3.3.

3.3.1 Double muon trigger

In Run 3, the inclusive double-muon parking trigger represents the most stable and productive investment within the CMS flavor and B-physics program under multiple aspects. Compared to the Run 2 trigger strategy, it ensures a more efficient use of the available trigger bandwidth and computing resources, owing to a general rationalization of the trigger path architecture. As a consequence, it provides a substantial gain in yields for b-hadron decay modes with double-muon final states, which were only marginally accessible during Run 2. For instance, the yield of $B^0 \to J/\psi K^0_S$ decays has increased by about an order of magnitude with respect to Run 2.

The `HLT_DoubleMu_4_3_LowMass` path receives contributions from many L1 seeds that require two oppositely charged muons, with only small changes between the 2022 and 2023 years of data taking at the level of L1 algorithms. Those associated with the highest rate can be divided in two groups for both years. The "central-η" algorithms impose a threshold on single muon $|\eta|$ without limiting the muon p_T, even if an implicit threshold of $p_T > 3$ GeV is required for the muons to reach the muon detectors in the central region. The "higher-p_T" algorithms require single muon p_T above 3 or 4 GeV with a further selection on the angular distance between the two muons. The main selections applied at the L1T level are detailed in Tab. 3.2.

At the level of the HLT the path requires the presence of at least two muons with p_T > 4 GeV and $p_T > 3$ GeV, respectively, both with absolute pseudorapidity $|\eta|$ below 2.5 and the double-muon system is required to have $p_T > 4.9$ GeV. To mitigate the contamination from PU muons, the distance of closest approach (DCA) of each muon with respect to the beam-spot must be smaller than 2.0 cm (see Fig. 3.3). Muons must be originating from the same interaction points, consequently their tracks are required to have $\mathrm{DCA}(\mu_1, \mu_2) < 0.5$ cm and the probability of a fit of their

Table 3.2 Selection applied at the level of the L1T by the main L1 seeds for the inclusive double-muon parking trigger, separately in the 2022 and 2023 year of data taking. Each line represents a different L1 seed.

Year	Selection on L1 muons
2022	no p_T requirement and $\|\eta\|(\mu_i)$<2.0 and $\|\Delta\eta(\mu_1,\mu_2)\|$ <1.6
	$p_T > 4$ GeV and $\Delta R(\mu_1,\mu_2)\|$ <1.2
2023	no p_T requirement and $\|\eta\|(\mu_i)$<1.5 and $\Delta R(\mu_1,\mu_2)\|$ <1.4
	$p_T > 3$ GeV and $\|\eta\|(\mu_i)$<2.0 and $\Delta R(\mu_1,\mu_2)\|$ <1.6
	$p_T > 4$ GeV and $\Delta R(\mu_1,\mu_2)\|$ <1.2

tracks to a common secondary vertex (SV) larger than 0.5%. The muons invariant mass is restricted to the interval between 0.2 GeV and 8.5 GeV. These requirements are summarized in Tab. 3.3.

Table 3.3 Selection applied at HLT to the online reconstructed muons by the double-muon parking trigger.

Target	Selection at HLT level
leading muon (μ_1)	$p_T > 4$ GeV and $\|\eta\| < 2.5$
	DCA(μ_1, BS) < 2.0 cm
sub-leading muon (μ_2)	$p_T > 3$ GeV and $\|\eta\| < 2.5$
	DCA(μ_2, BS) < 2.0 cm
muon pair	DCA(μ_1,μ_2) < 0.5 cm
	2.0 GeV < $m(\mu_1,\mu_2)$ < 8.5 GeV
	p_T $(\mu_1,\mu_2) > 4.9$ GeV
	double-muon vertex probability > 0.5%

3.3.2 Tri-muon triggers

The triple muon paths `HLT_Tau3Mu_Mu7_Mu1_TkMu1_IsoTau15_Charge1` and `HLT_Tau3Mu_Mu7_Mu1_TkMu1_IsoTau15`, have been deployed during LHC Run 2 to target specifically the $W \to \tau(3\mu)\nu$ isolated topology. The only difference between the two is that the former path contains an explicit requirement on the absolute charge of the tri-muon system ($|\sum_i q(\mu_i)| = 1$). They require the presence of two L3 muons with $p_T > 7$ GeV and $p_T > 1$ GeV and one tracker muon with $p_T > 1$ GeV. All muons must have $|\eta| < 2.5$. At L1 they are seeded by a set of single- double- and triple- muon seeds.
Some geometrical and kinematical requirements are introduced in order to select muons which are likely to come from the same interaction vertex in the detector as summarized in Tab. 3.4. At least two out of the three muons must have longitudinal distance $\Delta z(\mu_i,\mu_j)$< 0.3 cm, with each muon having DCA(μ, BS)< 0.5 cm, the pair invariant mass between 0.2 GeV and 1.9 GeV and $p_T > 2$ GeV. The τ candidates are

built from muon triplets and only those having $p_T > 15$ GeV, $|\eta| < 2.5$ and invariant mass between 1.3 GeV and 2.1 GeV pass the HLT selection. The three leptons are required to be isolated with absolute isolation smaller than 2.0 GeV and relative isolation smaller than 0.2. The isolation at HLT is calculated similarly as the one used offline (see Sec. 3.4.2), but only charged particles are considered.

Table 3.4 Selection applied at the level of the HLT on the online reconstructed muons by the tri-muon parking triggers.

Target	Selection at HLT level
leading muon (μ_1)	$p_T > 7$ GeV and $\|\eta\| < 2.5$ DCA(μ_1, BS) < 0.5 cm
sub-leading muon (μ_2)	$p_T > 1$ GeV and $\|\eta\| < 2.5$ DCA(μ_2, BS) < 0.5 cm
sub-leading muon (μ_2)	$p_T > 1$ GeV and $\|\eta\| < 2.5$ DCA(μ_3, BS) < 0.5 cm
at least one muon pair	$\Delta z(\mu_i, \mu_j) < 0.3$ cm 0.2 GeV $< m(\mu_i, \mu_j) <$ 1.9 GeV p_T (μ_i, μ_j) >2 GeV double-muon vertex probability $> 0.5\%$
muon triplet	p_T (3μ)>15 GeV and $\|\eta(3\mu)\| < 2.5$ $1.3 < m_{3\mu} < 2.1$ GeV

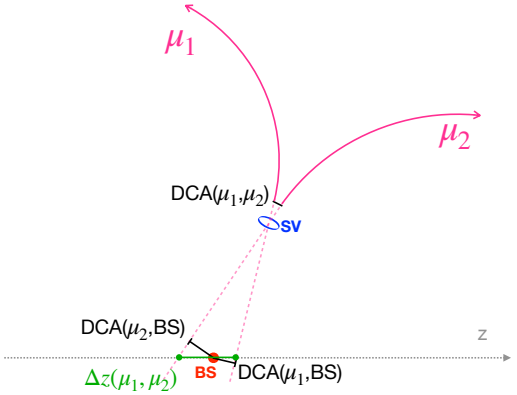


Figure 3.3 Sketch of the topological quantities used in the event selection at HLT-level.

3.3.3 HLT matching

A geometrical matching is performed to ensure that the offline selected muons correspond to the trigger objects, i.e. the L3 muons, which have fired the wanted HLT path. The matching is performed between offline muons and trigger objects

using the angular distance between the two, defined as

$$\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}, \tag{3.1}$$

ϕ being the azimuthal angle and η the pseudorapidity. Two objects are considered matched if ΔR <0.1. If multiple offline muons are matched to the same trigger object, the closest matching is kept.

3.4 Signal candidates reconstruction

In each event selected by one of the two HLT paths, potential LFV candidates ($\tau \to 3\mu$) are built by considering all possible combinations of three muons that lie within the CMS acceptance and are compatible with originating from a common SV. Each valid combination forms an eligible $\tau \to 3\mu$ candidate. The event missing transverse momentum is then combined with the four-momentum of the τ lepton to reconstruct the W boson kinematics in the transverse plane. This section provides a detailed description of the physics observables used in the analysis to identify and characterize potential $W \to \tau(3\mu)\nu$ candidates.

3.4.1 Muons

Events with at least three muons are selected. Muons must have passed the trigger selection as described in Sec. 3.3. If the double-muon trigger was fired, at least two muons of the triplet must be the trigger-muons, otherwise, if the triple-muon trigger was fired, all muons must be trigger muons. Only offline muons having $p_T > 3.5$ GeV and $|\eta| < 1.2$ or $p_T > 2.0$ GeV and $1.2 < |\eta| < 2.4$ are considered. All muons are required to pass the Medium-ID criteria described in Sec. 2.3.2.
It is also observed that around 65% of the $\tau \to 3\mu$ candidates have all three muons passing the Tight-ID. This algorithm selects PF global muons with additional quality requirements, such as a minimum number of hits in the inner tracker and impact parameter below 2 mm with respect to the primary vertex. These requirements are designed to suppress muons from decay in flight which are more likely to originate in hadronization. Particularly, the tight-ID flag of the trailing muon offers a valuable handle to discriminate signal candidates from background. The muon Tight-ID flag are given as input to the BDT discriminator which is used for the final selection, see Sec. 4.1.2. In Figure 3.4 the distribution of the muons p_T and the pairwise angular distance are shown for the signal using the 2022 simulated sample.

3.4.2 $\tau \to 3\mu$ candidates

Signal $\tau \to 3\mu$ candidates are built from any possible combination of three muons passing the aforementioned preselection. If the charge of the muon triplet is unitary and the invariant mass is smaller than 3.0 GeV, the triplet undergoes a kinematic fit to a common vertex in order to select only muons coming from the same point inside the detector. Given the non-negligible τ mean life, 30 ps [21] approximately, the lepton is likely to cover a significative distance, with respect to the inner tracker resolution, i.e. about 1 mm before decaying. Consequently, the signal events are

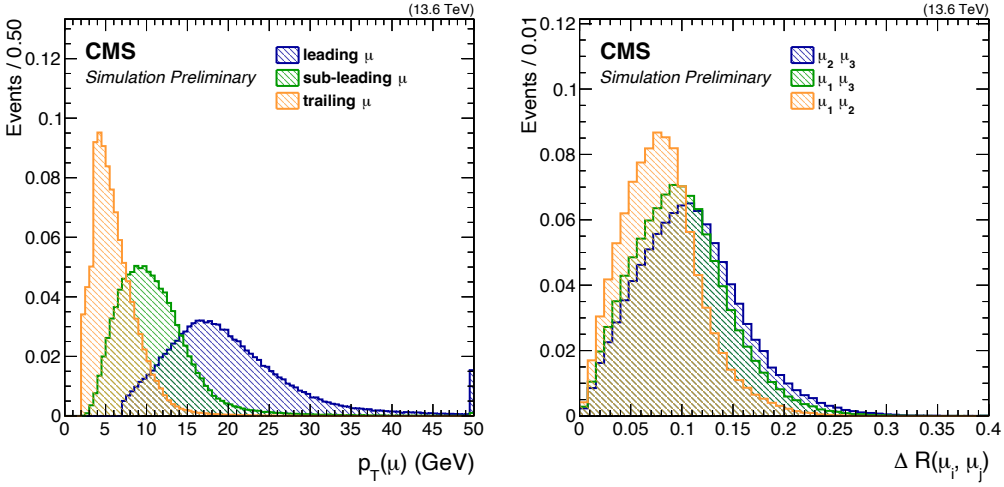


Figure 3.4 Reconstructed p_T (left) and the pairwise angular distance ΔR (right) for the three muons of the $\tau \to 3\mu$ candidates in signal MC samples for 2022 data-taking conditions.

expected to originate from a decay point which might be displaced with respect to the beam-spot, therefore referred to as the secondary vertex (SV). The kinematic vertex fit is based on the least means squared minimization of the distance between the input tracks and the common interaction point, with Lagrange multipliers to implement the kinematic constraint. If the common vertex is found, i.e. the fit converged, the resulting SV is associated with a fit probability, given by the χ^2 probability of the kinematic fit itself. The triplet is rejected, otherwise. Each muon track is re-fitted with the additional constraint that it originates from the SV and its kinematics is corrected accordingly. The re-fitted tracks parameters provide a better resolution on the signal $\tau \to 3\mu$ invariant mass, as shown in Fig. 3.5, and they are used in the rest of the analysis.

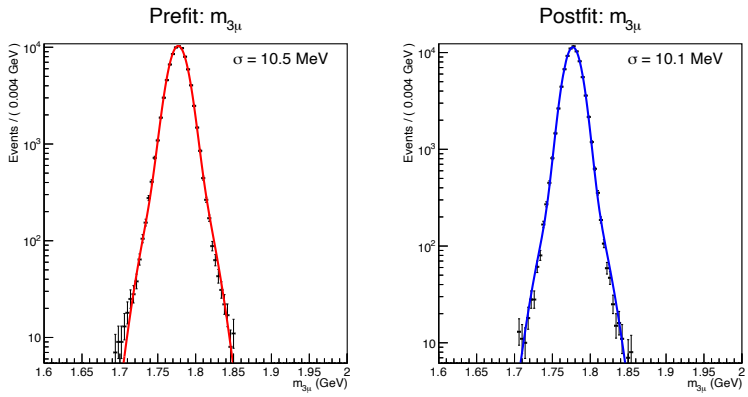


Figure 3.5 Reconstructed invariant mass distribution in simulation for the $\tau \to 3\mu$ signal using the pre-fit (left) and the corrected (right) muons kinematics, restricted to barrel region.

To reduce the contamination from the combinatorial background, namely three muons produced in unrelated processes, signal candidates are discarded if any of the three muons forms a "good double-muon vertex", with another oppositely charged muon in the event that is not part of the signal triplet. A double-muon vertex is considered "good" if the resulting fit probability is larger than 5% and the post-fit invariant mass of the two muons lies within 2σ from the nominal mass of the resonances listed in Tab. 3.5. Conversely, contamination from dimuon resonances reconstructed within the $\tau \to 3\mu$ candidates is suppressed using the procedure described in Section 4.3.

Table 3.5 List of vetoed resonances and values of the width considered for the 2σ exclusion. The central value is the nominal mass, and the width, in case of narrow resonances, corresponds to CMS experimental resolution.

Resonance	**Nominal mass [GeV]**	σ **[GeV]**
η	0.5479	0.030
$\rho(770)$	0.7753	0.075
ω	0.7827	0.030
ϕ	1.0195	0.030
J/ψ	3.0969	0.030
$\psi(2S)$	3.6861	0.030
Υ	9.4603	0.070
$\Upsilon(2S)$	10.0233	0.070
$\Upsilon(3S)$	10.3552	0.070
Z	91.1976	2.495

In addition, the selection applied at trigger level is reproduced offline, except for the isolation cuts because those computed offline are not well described by the ones calculated offline. The trigger selection depends on the HLT path which fired the event giving priority to the double-muon trigger. If this trigger fires, independently of the other considered paths, the requirements listed in Tab. 3.3 must be fulfilled by at least one double-muon pair in the $\tau \to 3\mu$ candidate. Otherwise, the tri-muon trigger requirements of Tab. 3.4 are applied to the offline muons. Muons triggered by the `HLT_DoubleMu_4_3_LowMass` path suffer from a large contamination from fakes, i.e. light charged hadrons like π and K which are identified as muons by the particle-flow algorithm. To mitigate this effect, which is more pronounced at low p_T, tighter selections are applied offline on the trigger muons, which reproduce some of the tri-muon trigger requirements. A leading-muon $p_T > 7$ GeV and trailing muon $p_T > 1.0$ GeV are required and the muon pair must have longitudinal distance $\Delta z < 0.7$ cm and invariant mass below 1.9 GeV. Finally, the $\tau \to 3\mu$ candidate is required to have $p_T > 15$ GeV. This reinforcement on the HLT emulation is 99% efficient on the simulated $\tau \to 3\mu$ candidates selected by the double-muon trigger.

The set of selection criteria described so far defines the preselection strategy, aimed at identifying events that potentially contain a muon triplet with the topology and kinematics expected for $\tau \to 3\mu$ decays. These requirements are summarized in

Tab. 3.6. A more specific selection is subsequently applied to optimize the signal sensitivity, as detailed in the following chapters. In particular, the next sections describe the kinematic and topological variables of the $W \to \tau(3\mu)\nu$ decays that provide strong discrimination power against background processes and are used as input to the multivariate analysis presented in Sec.4.1.2.

Particularly useful for this analysis is the muon triplet displacement significance in the transverse plane, defined as the magnitude of the SV displacement vector, projected on transverse plane, normalized to its uncertainty (L_{xy}/σ). The total τ momentum, given from the combination of the muon refitted tracks, is expected to be aligned with the displacement vector in signal events. Therefore, the angle $\alpha(\tau, \text{BS})$ between the direction of the τ momentum and the segment connecting the beam-spot (BS) and the SV, projected on the transverse plane has a large discriminating power with respect to background candidates. These quantities are visually defined in the sketch in Fig. 3.6 and they will be used in the next steps of the analysis. Finally, leptons from W decays tend to have little or no hadronic activity surrounding them as opposed to leptons produced in charm and beauty decays or to fake candidates. This aspect is quantified by the $\tau \to 3\mu$ isolation, defined as

$$I_{\text{abs}} = \sum p_T^{\text{charged}}(d_z < 0.2 \text{ cm}) + \max\Big(0, \sum p_T^{\gamma} - \Delta\beta \sum p_T^{\text{charged}}(d_z > 0.2 \text{ cm})\Big). \tag{3.2}$$

The first term is the sum of the transverse momenta of the charged tracks reconstructed in a cone of radius ΔR =0.4 around the τ direction, and compatible within 0.2 cm with primary vertex (PV) along the z coordinate. The three signal muons are excluded from the sum. The second term contains the sum of the photons transverse momenta, reconstructed in the same cone. The PU contribution is statistically removed by subtracting an amount proportional to the sum of the tracks transverse momenta considering only those tracks which are not compatible with the primary vertex and falling within a ΔR=0.8 cone around the τ candidate. The proportionality factor is $\Delta\beta$ =0.2, and it has been optimized to remove the taus identification efficiency dependency on PU. The correlation between pileup and each term appearing in Eq. 3.2 is shown in the left plot of Fig. 3.7, demonstrating that the total isolation is negligibly biased by the interactions multiplicity.
In this analysis the relative isolation is used as it is less dependent on the τ lepton kinematics:

$$I_{\text{rel}} = I_{\text{abs}}/p_T^{\tau}. \tag{3.3}$$

The right plot in Fig. 3.7 reports the signal efficiency, as a function of PU, for an isolation requirement $I_{\text{rel}} < 0.2$, i.e. the isolation requirement at HLT level.

3.4.3 W boson reconstruction

$W \to \tau(3\mu)\nu$ events are characterized by the presence of non-negligible MET due to the non-reconstructed neutrino. MET-related observables provide a valuable handle to discriminate signal from background events. The missing transverse momentum magnitude and direction in the transverse plane are determined by the particle-flow algorithm (Sec. 2.3.1), balancing the visible momentum of the collision. The PF MET is then corrected using the Pileup Per Particle Identification (PUPPI) algorithm,

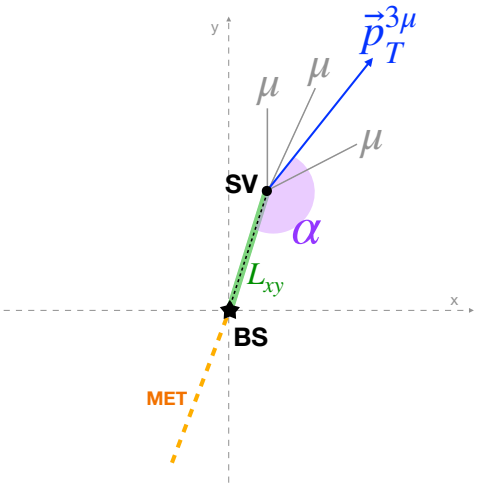


Figure 3.6 Qualitative representation of the $W \to \tau(3\mu)\nu$ decay topology as projected on the transverse plane in the CMS reference system. The secondary vertex displacement (L_{xy}) and the angle (α) between the displacement direction and the $\tau \to 3\mu$ momentum are highlighted in green and purple, respectively.

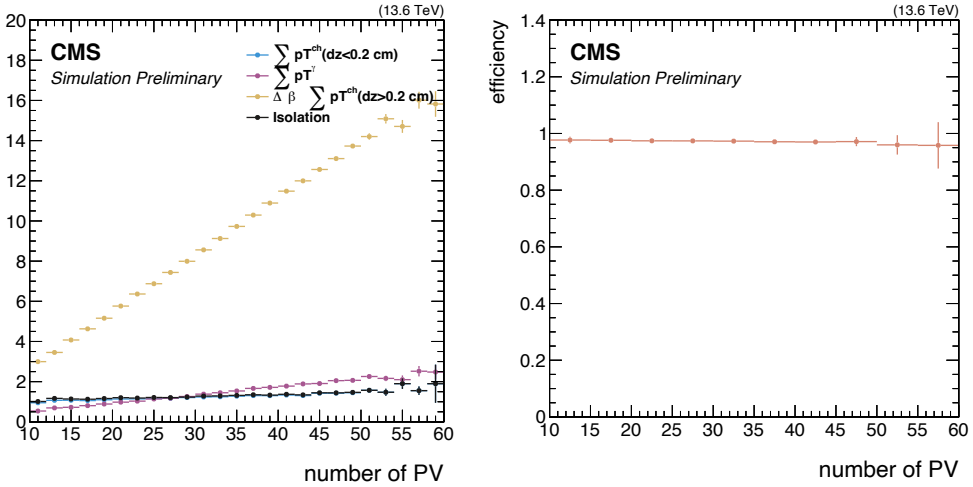


Figure 3.7 On the left, profiled distribution of the isolation components appearing in Eq. 3.2 and the total absolute isolation (black dots) and on the right, the efficiency of the $I_{\text{rel}} > 0.2$ requirement. In both figures the quantities are shown as a function of the number of reconstructed primary vertices.

which is a pileup mitigation technique that assigns weights to each particle in the event based on its probability to be produced by a PU interaction [78].
In the majority of the simulated $W \to \tau\nu$ decays, the neutrino is produced in the opposite direction with respect to the τ lepton, therefore, signal events generally have MET and $\tau \to 3\mu$ vectors in the opposite directions and with balanced p_T, as shown in Fig. 3.8. The neutrino momentum longitudinal component (p_z^ν) is inferred by solving the four-momentum conservation and assuming that the W boson is

produced on-shell:

$$m_W^2 = (P^\tau + P^\nu)^2, \tag{3.4}$$

$$0 = [(E^\tau)^2 - (p_z^\tau)^2](p_z^\nu)^2 - 2Ap_z^\tau p_z^\nu + (E^\tau p_T^\nu)^2 - A^2, \tag{3.5}$$

$$\text{where } A \equiv \frac{1}{2}(m_W^2 - m_\tau^2) + \vec{p_T^\tau} \cdot \vec{p_T^\nu}, \tag{3.6}$$

where p_T^ν indicates the MET of the event. Equation 3.6 has two solutions, both are used in the analysis as inputs to the BDT discriminator described in Sec. 4.1.2. Combining the $\tau \to 3\mu$ candidate and the MET in the transverse plane the transverse mass m_T is defined as

$$m_T = \sqrt{2p_T^\tau \cdot \text{MET} \cdot [1 - \cos\Delta\phi_{(\tau,\text{MET})}]}. \tag{3.7}$$

Since signal events are characterized by large MET, in the opposite dierection with respect to τ direction, $\tau \to 3\mu$ candidates are expected to have large transverse mass, above 40 GeV, differently with respect to combinatorial τ leptons.

If multiple $\tau \to 3\mu$ candidates within the same event satisfy the preselection requirements, only the one with the largest transverse mass is retained for the subsequent analysis steps. The fraction of events containing more than one reconstructed $\tau \to 3\mu$ candidate is 0.15% (0.23%) in the MC signal sample and 0.81% (0.81%) in data for the 2022 (2023) dataset, as shown in the left panels of Fig. 3.9. The right panels of the same figure indicate that, in events with multiple $\tau \to 3\mu$ candidates, the discarded candidates share two out of the three muons with the selected one in more than 97% of the cases. Finally, the rate of events in which the reconstructed $\tau \to 3\mu$ candidate is not the one produced at generation level is below 3% in both years of data taking.

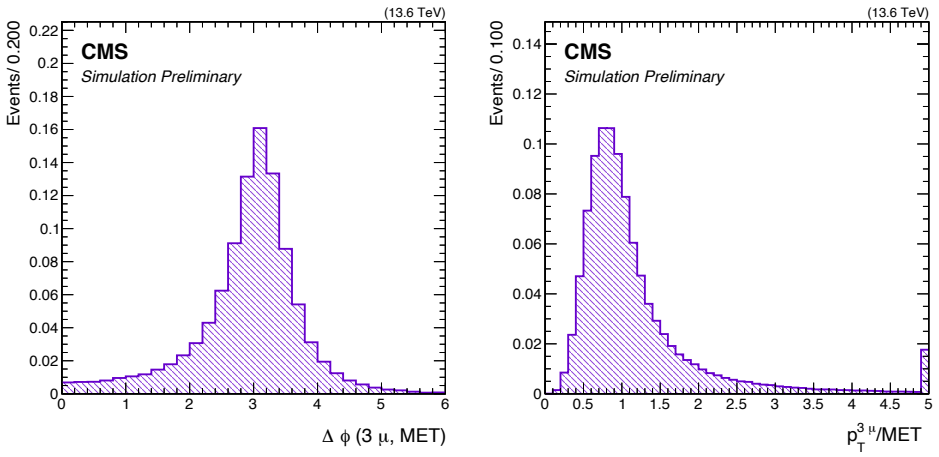


Figure 3.8 Angular distance $\Delta\phi$ in the transverse plane between the $\tau \to 3\mu$ candidate and MET vector (left) and ratio between the $\tau \to 3\mu$ p_T and the MET of the event in the signal simulated sample.

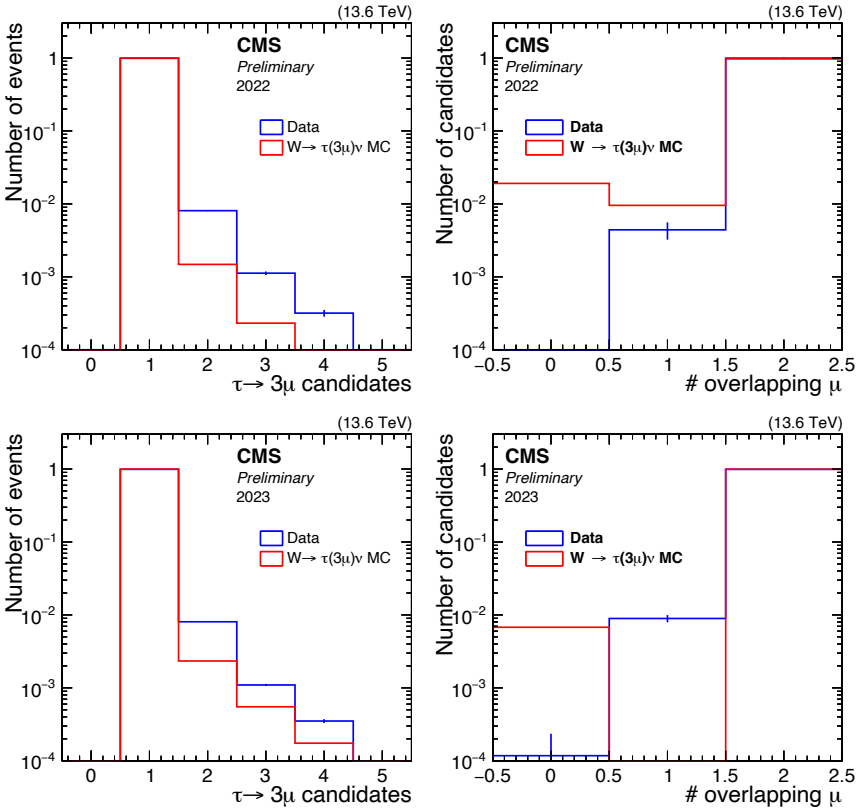


Figure 3.9 Per-event number of $\tau \to 3\mu$ candidates passing all the preselection requirements (left). Number of muons the discarded candidates share with the one kept in the analysis (right). MC (data) samples corresponds to the red (blue) lines, the top and bottom row plots refer to the 2022 and 2023 years, respectively.

3.4.4 Remarks on the preselection

The selection criteria outlined in the previous section are summarized in Tab. 3.6. They target the identification of a single $\tau \to 3\mu$ candidate in each event starting from muons which are selected online by either the double-muon parking trigger or by the tri-muon trigger. The two HLT strategies have a large overlap in terms of the number of selected $\tau \to 3\mu$ candidates: in around 76% of the events the τ lepton candidate passing the preselection is selected online by both the paths. The two strategies are also complementary as around 12% of signal events are selected by only one of the two triggers. Figure 3.10 shows the overlap and the orthogonality between the $\tau \to 3\mu$ MC events selected by the double- and tri-muon HLT paths, after the application of the preselection. The total efficiency on $W \to \tau(3\mu)\nu$ simulated events is 23.5%(22.7%) in 2022 (2023) samples (see also Tab. 5.1).

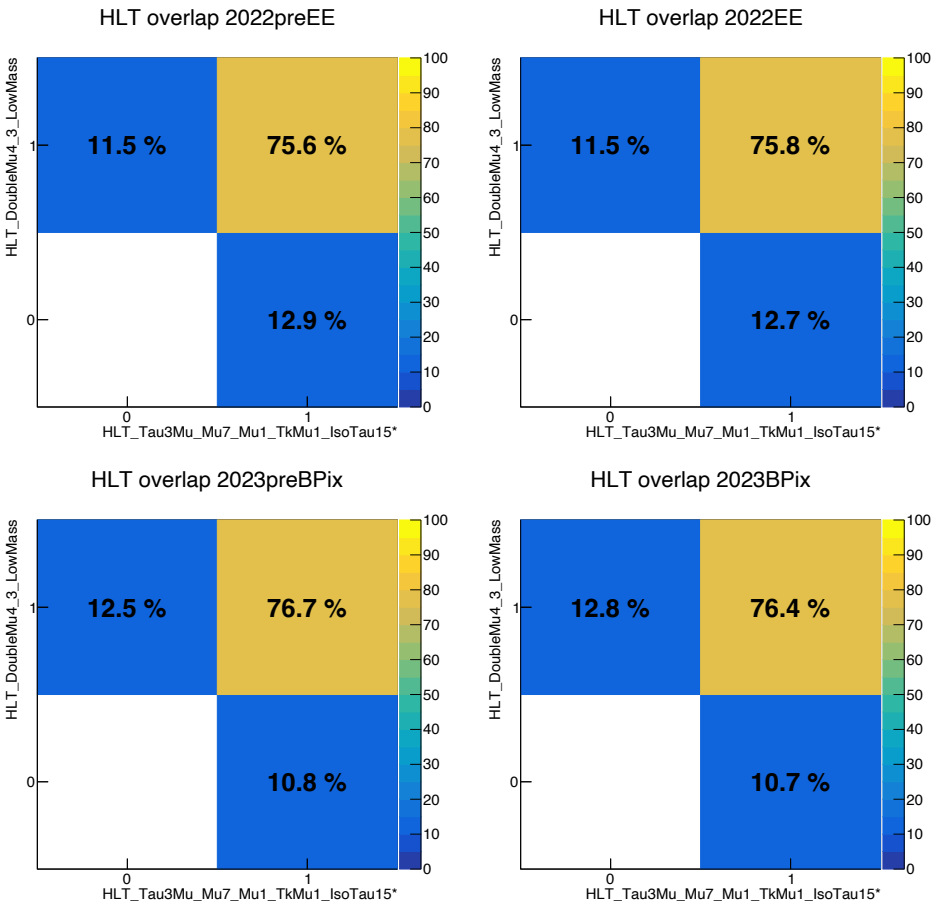


Figure 3.10 Events overlap and orthogonality between the `HLT_DoubleMu_4_3_LowMass` and `HLT_Tau3Mu_Mu7_Mu1_TkMu1_IsoTau15` * paths. Simulated events, with the offline emulation applied, are split by year - 2022 (top) and 2023 (bottom)- and by the different data-taking periods corresponding to different samples (left and right).

Table 3.6 Preselection requirements.

Target	Selection
Events	- double- or tri-muon HLT fired
Muons	- $p_T > 3.5$ GeV if $\|\eta\| < 1.2$ or $p_T > 2.0$ GeV if $1.2 < \|\eta\| < 2.4$ - pass the Medium-ID - matched to HLT muons
$\tau \to 3\mu$ cand.	- $\|\sum_i q(\mu_i)\| = 1$ - successful three-muons SV fit - no good double-muon vertex for any pair made of one muon of the $\tau \to 3\mu$ candidate and another opposite signed of the event - HLT selection of Tab. 3.3 (Tab. 3.4) re-applied to offline muons if double-muon trigger (did not) fired
$\tau \to 3\mu$ passing double-muon HLT	- $p_T(\mu_1) > 7.0$ GeV and $p_T(\mu_3) > 1.0$ GeV - at least 2 muons with $m_{\mu_i\mu_j} < 1.9$ GeV and $\Delta z(\mu_i\mu_j) < 0.7$ cm - $p_T(3\mu) > 15$ GeV

3.5 Events categorization

In order to improve the overall signal sensitivity, the $\tau \to 3\mu$ candidates passing the preselection criteria described in the previous section are divided into categories based on the pseudorapidity of the muon triplet total momentum, $|\eta_{3\mu}|$. The muon reconstruction performance is not uniform across the CMS detector, as the resolution on the triplet invariant mass degrades in the forward region, directly affecting the signal shape. In addition, the background composition varies between the central and forward regions. In particular, misidentified and combinatorial muons are more prevalent at high pseudorapidity. Therefore, different phase space regions are treated with dedicated selection criteria, optimized separately, in order to maximize the global sensitivity of the analysis. For each year of data taking three orthogonal categories are defined:

- **A** : $|\eta_{3\mu}| \leq 0.9$,
- **B** : $0.9 < |\eta_{3\mu}| \leq 1.8$,
- **C** : $1.8 < |\eta_{3\mu}| \leq 2.4$,

for a total of six categories. The event categorization used in this work is different from the one used in the CMS Run2 analysis, where the categorization was based on the $\tau \to 3\mu$ candidate relative mass uncertainty σ_M/M, where σ_M is the per-candidate uncertainty from the SV fit propagated to the triplet invariant mass. The Run 2 categories are defined as follows:

- **A'** : $\sigma_M/M \leq 0.7\%$,
- **B'** : $0.7\% < \sigma_M/M \leq 1.2\%$,
- **C'** : $\sigma_M/M > 1.2\%$.

The heatmap in Fig. 3.11 shows the strong correlation between the mass resolution σ_M/M and the pseudorapidity of the $\tau \to 3\mu$ candidates. The momentum resolution of individual muons deteriorates in the forward regions for the shorter lever arm in the less uniform magnetic field, the larger detector material, and the reduced detector coverage thus the resolution on the muon triplet reconstructed kinematics increases.

The Run 3 and Run 2 definitions are strongly correlated and both defines three categories with different resolution on the $\tau \to 3\mu$ invariant mass, hence different signal shapes. It was, nevertheless, observed that the Run2 categorization creates non-trivial background shaping in the $\tau \to 3\mu$ mass distribution, as shown Figure 3.12. Conversely, the event categorization using the triplet pseudorapidity proves to be more resilient to mass shaping, as shown by the plots in Fig. 3.13.

In the following, the pseudorapidity category of a given year is referred to as "Xyy," where "X" indicates the pseudorapidity region and "yy" are the last two digits of the year (e.g., A22 or C23, etc.).

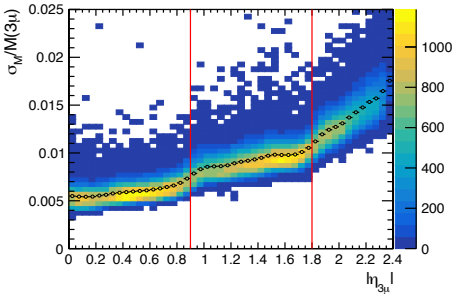


Figure 3.11 Relative mass uncertainty of the $\tau \to 3\mu$ candidates versus the absolute pseudorapidity, in the 2022 Monte Carlo signal sample. The plot shows that the two quantities are strongly correlated. The red lines represent the boundaries of the Run 3 event categorization.

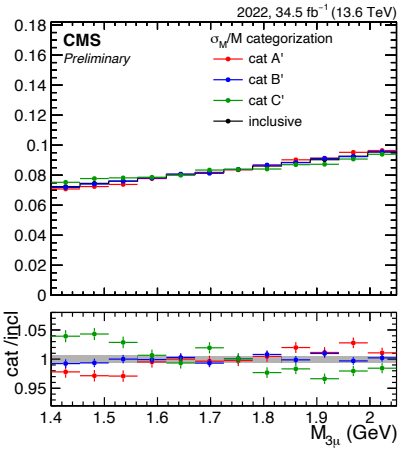


Figure 3.12 Reconstructed $\tau \to 3\mu$ candidates invariant mass in 2022 data after preselection and using the Run2 categories defined upon the relative mass uncertainty. Different background shapes are visible in the three categories. Distributions are normalized to unity.

It is, finally, proved that the chosen boundaries maximize the signal sensitivity studying the combined upper limit, as a function of the values of $|\eta_{3\mu}|$ at the boundaries separating categories A and B ($|\eta|_{\mathrm{AB}}$) and categories B and C ($|\eta|_{\mathrm{BC}}$). The results of the scan are reported in Fig. 3.15, the 2022 data and MC are used applying a loose cut on the BDT score (a detailed description of the analysis BDT is in Sec. 4.1.2). Since negligible changes are expected in the muon reconstruction the same categories definition is used for 2022 and 2023. Additionally, the impact on the total sensitivity of a larger number of categories is tested. Each category is split first in two, having 6 categories in total, and then in three subcategories, for a total of 9, without optimizing the internal boundaries in $|\eta_{3\mu}|$. No significant improvement is observed on the combined expected upper limit. Hence, it is concluded that the one presented is the optimal categorization for this analysis.

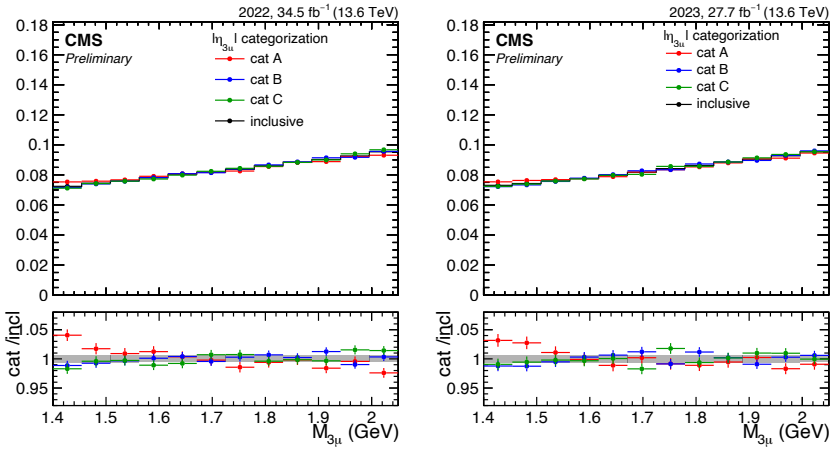


Figure 3.13 Reconstructed $\tau \to 3\mu$ candidates invariant mass in data (left: 2022; right: 2023), after the preselection, for the three categories defined on the τ pseudorapidity. No significant difference in the background shape is observed across the three categories. Distributions are normalized to unity.

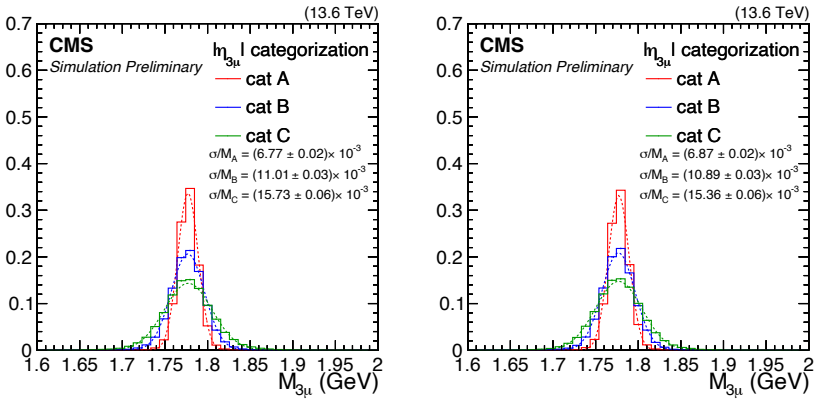


Figure 3.14 Signal shape in the three analysis categories for 2022 (left) and 2023 (right) $W \to \tau(3\mu)\nu$ simulated samples.

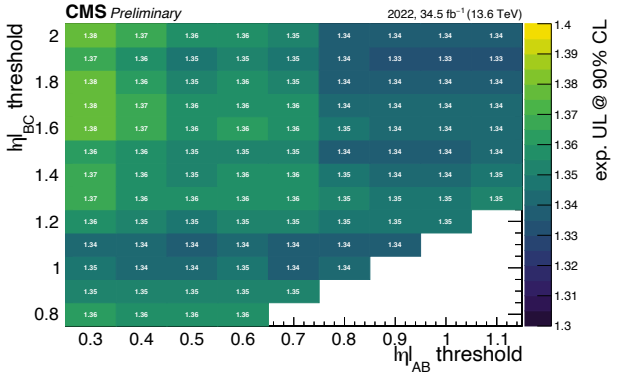


Figure 3.15 Combined expected exclusion upper limit at 90% CL as a function of the $|\eta|$ boundaries separating category A and B and category B and C. The thresholds chosen in the analysis $|\eta_{AB}| = 0.9$ and $|\eta_{BC}| = 1.8$ are in the region with better performance.

3.6 Simulation corrections

The analysis relies on MC samples to describe the signal processes and to compute selection efficiencies. It is essential to ensure that the simulations accurately reproduces the data taking conditions and the efficiency observed on data.
Several potential discrepancies between simulation and data are addressed through the application of event-by-event correction weights to the simulated events, referred to as scale factors. The general approach is to compare the same quantities in data and MC samples, the magnitude of the correction is defined as the discrepancy observed between data and MC. At each correction is associated a systematic uncertainty, which is taken into account in the final result. A detailed discussion on the systematics is presented in Sec. 5.4.
The derivation of the various corrections and their physical sources are discussed in the next sections.

3.6.1 Pileup

As discussed in Sec. 2.2, pileup refers to the multiplicity of proton-proton interactions per bunch-crossing. In simulations, the pileup distribution is defined prior to data taking and therefore does not necessarily match the conditions observed in data. Given an input pileup distribution, additional interactions besides the primary hard scatter are generated by sampling a value from this distribution, referred to as the *true* pileup. The number of pileup events, both in-time and out-of-time, is then drawn from a Poisson distribution with mean equal to the true pileup value. In data, the quantity equivalent to the true pileup in MC is calculated using luminosity measurements, as pileup is defined as in Eq. 2.3. Corrections are derived comparing the true pileup distribution in data and MC. Their impact is directly visible in the distribution of the number of reconstructed primary vertices, since the PV multiplicity scales with PU, with an average reconstruction efficiency of about 70% (Fig. 3.16). Good agreement between data and Monte Carlo is observed in 2022, while a shift between the two distributions is visible in 2023. This discrepancy was traced to the input pileup distribution used in 2023 MC samples, which assumes a proton-proton inelastic cross section of 80 mb, whereas data are more consistent with a value of 69.2 mb, as used in 2022. As a result the primary vertices multiplicity is larger in uncorrected MC than in data, while it becomes smaller after the corrections are applied, indicating an over-correction. It was verified, however, that the 2023 pileup corrections, although not optimal, do not have a significant impact on the analysis performance.

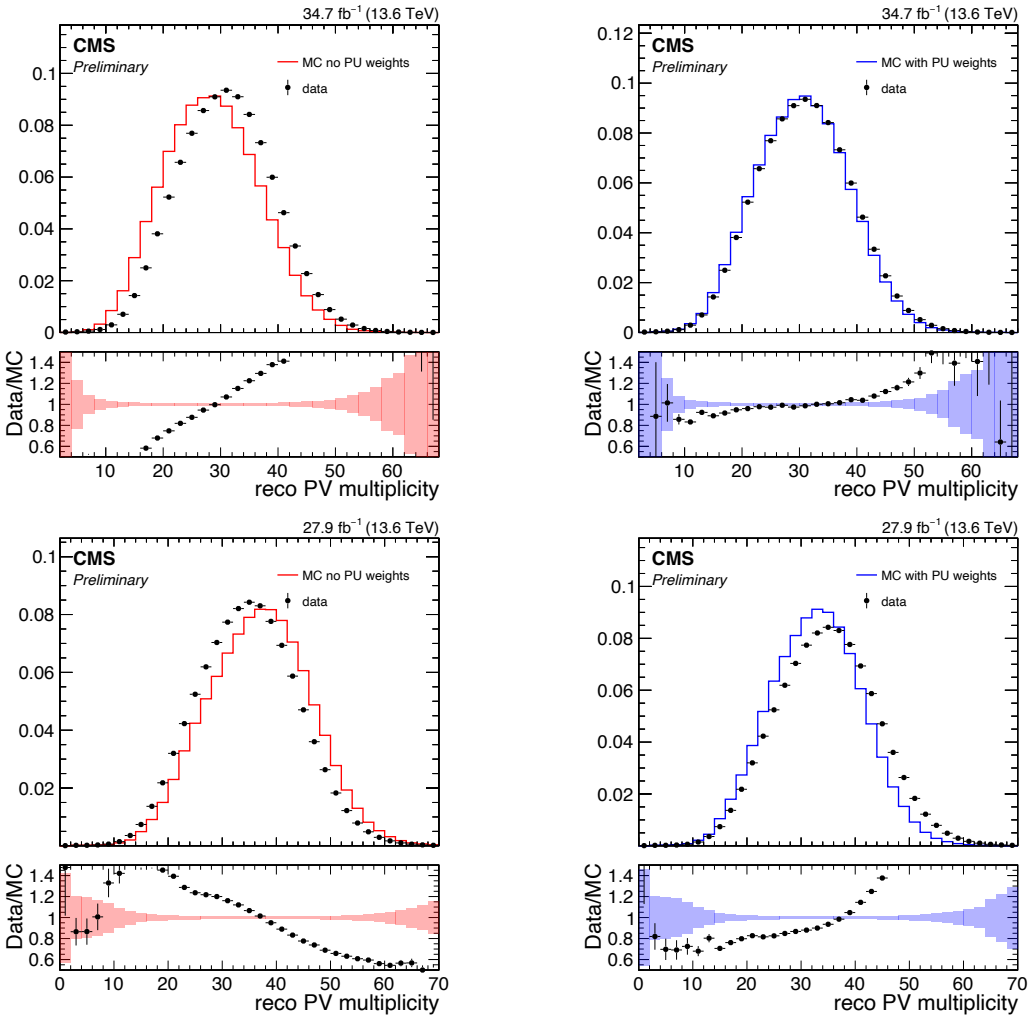


Figure 3.16 Distribution of the number of reconstructed primary vertices in data and $W \to \tau(3\mu)\nu$ simulation for 2022 (top row) and 2023 (bottom row). Plots on the left (right) report the comparison before (after) the PU correction factors are applied.

3.6.2 Muon ID efficiency

As described in Sec. 3.4.1, the muons entering the analysis are required to pass the medium muon identification criteria. The identification efficiency is measured using the tag-and-probe method, applied to reconstructed $J/\psi \to \mu\mu$ decays to match the same phase space region as the signal muons, which are relatively low in p_T.
For these studies a simulated sample of prompt and non-prompt $J/\psi \to \mu\mu$ events is used; data are taken from the CMS muon dataset, which is collected using several single-, double-, and triple- muon triggers and is the most unbiased dataset available for performance studies on muons.
The tag-and-probe method consists in reconstructing $J/\psi \to \mu\mu$ events, where one of the two muons, the *tag-muon*, is triggered by the `OR` of various single-muon triggers and satisfies a tight offline selection. The other muon, the *probe-muon*, is subject to looser requirements, to avoid biases on the efficiency measurement. The

tag-and-probe pairs are selected applying the requirements in Tab. 3.7. The medium identification efficiency is then defined as the fraction of tag-and-probe pairs in which the probe muon passes also the Medium-ID criteria:

$$\varepsilon_{\mathrm{ID}} = \frac{N_{J/\psi}^{\text{pass ID}}}{N_{J/\psi}^{\text{total}}}, \tag{3.8}$$

where the number of reconstructed J/ψ ($N^{*}_{J/\psi}$) is extracted from the fit to the di-muon invariant mass distribution for events passing the given selection.
The data to MC scale factors are given by the ratio of the efficiencies and are defined in bins of p_T and $|\eta|$ of the probe muon, as listed in Tab. 3.7. They are shown in Fig. 3.17 for 2022 and 2023. They are applied event by event to each of the three muons according to its p_T and pseudorapidity.

Table 3.7 Selection applied to the tag and probe muons and on the muon pair to select the $J/\psi \to \mu\mu$ events and evaluate the Medium-ID efficiency.

tag-muon	$p_T > 8$ GeV and $\|\eta\| < 2.4$ pass the Tight-ID criteria matched to a trigger object from a collection of single muon triggers
probe-muon	Tracker muon $p_T > 3.5$ GeV $\Delta z(\mu, \mathrm{BS}) < 0.5$ cm
muon pair	$\Delta R\ (\mu_1, \mu_2) > 0.3$ 2.7 GeV $< \mathrm{M}(\mu_1, \mu_2) <$ 3.5 GeV
$p_T \times \|\eta\|$ bins	[0, 3, 4, 5, 6, 7, 8, 9, 10, 30] GeV $\times$ [0, 0.9, 1.2, 2.1] [0, 3, 8, 30] GeV $\times$ [2.1, 2.4]

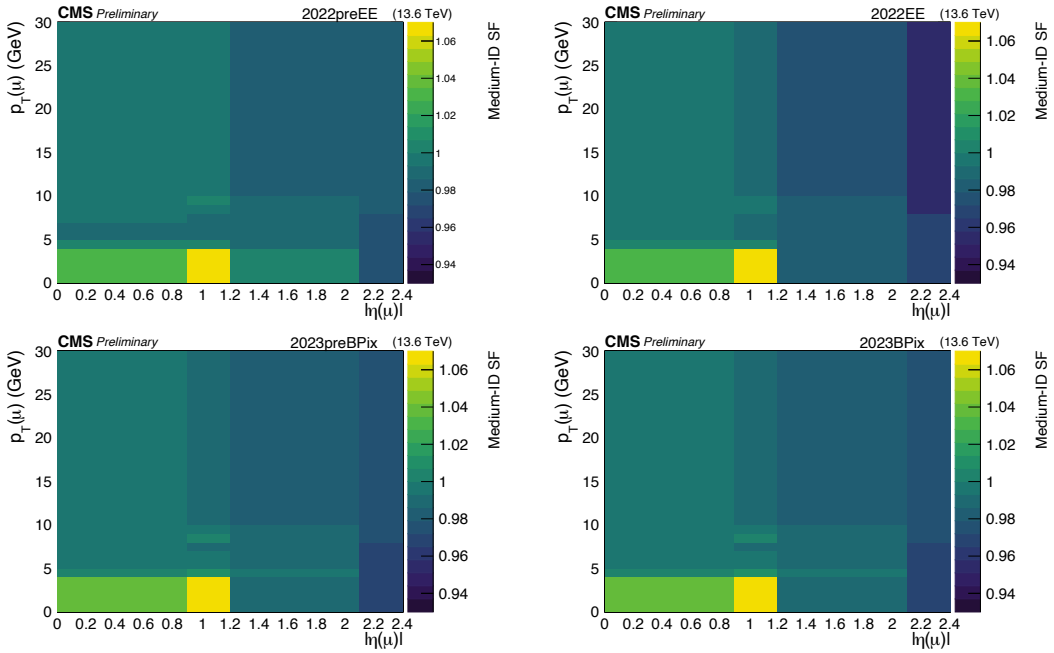


Figure 3.17 Muon medium-ID scale factors for 2022 (top) and 2023 (bottom), and for the two data-taking periods in which each year is divided (left and right).

3.6.3 Trigger efficiency

This analysis exploits the OR of the three HLT paths `HLT_DoubleMu_4_3_LowMass`, `HLT_Tau3Mu_Mu7_Mu1_TkMu1_IsoTau15_Charge1` and `HLT_Tau3Mu_Mu7_Mu1_TkMu1_IsoTau15` to select events online, as described in Sec. 3.3. For the sake of simplicity and given the large overlap between the paths (see Fig. 3.10), trigger efficiency corrections are calculated only for the di-muon path. No correction is defined for those events which only fired the tri-muon paths, instead a systematic uncertainty is applied. This will be described in Sec. 5.4.

The `HLT_DoubleMu_4_3_LowMass` efficiency is studied with the tag-and-probe method on $J/\psi \to \mu\mu$ events, independently for the 2022 and 2023 years of data taking. The data and MC dataset used for the study are the same as for the muon-ID efficiency. In each tag-and-probe pair, the tag muon is the one which fires a reference HLT path and the probe muon is used to measure the efficiency of the probe HLT path. The trigger efficiency is defined as the fraction of J/ψ candidates that passed the tag-selection in which the probe muon fires the probe trigger:

$$\varepsilon = \frac{N_{J/\psi}^{\text{pass HLT}}}{N_{J/\psi}^{\text{total}}}. \tag{3.9}$$

The `HLT_DoubleMu_4_3_LowMass` performance are evaluated factorizing the L1T and the HLT efficiency. The reference path to compute the L1 efficiency, ε(L1), is `HLT_Mu8`, only requiring the presence of a global muon with p_T >8 GeV. The probe path at L1 is `HLT_Mu0_L1DoubleMu`, a pass-through for the `HLT_DoubleMu_4_3_LowMass` L1

seeds, i.e. simply the logical OR of the L1 seeds with no requirements at HLT level. The HLT efficiency, ε(HLT|L1), is measured with HLT_Mu4_L1DoubleMu as reference path, the same L1 seeds pass-through with a lower threshold of 4 GeV on the muon p_T, required at HLT level. The probe HLT is the HLT_DoubleMu_4_3_LowMass. Consequently the total trigger efficiency is defined as

$$\varepsilon_{\mu_i,\mu_j} = \varepsilon_{\mu_i,\mu_j}(\text{L1}) \cdot \varepsilon_{\mu_i,\mu_j}(\text{HLT}|\text{L1}). \tag{3.10}$$

The trigger efficiency is studied in bins of the probe-muon p_T and in three regions of its absolute pseudorapidity $|\eta|$ corresponding to the barrel, overlap and endcap region of the muon system detector (see Sec. 2.2.5). The selection criteria for the $J/\psi \to \mu\mu$ candidates entering the study and the binning are detailed in Tab. 3.8. The J/ψ yield is measured from the bare number of events passing the above selections (cut-and-count approach). This choice was made because the selection is tight enough to make the combinatorial background negligible, especially in low statistics bins, hence the cut-and-count method leads to reliable results. The uncertainty on the MC efficiency is purely statistics while the efficiency in data is associated with both statistical and systematic uncertainties, summed in quadrature. We define the systematics on data as the absolute difference between the efficiency measured with cut-and-count approach and with the signal-plus-background fit to the $J/\psi \to \mu\mu$ candidates invariant mass. The plots in Fig. 3.18 report the data and MC efficiency for the two years of data taking.

Table 3.8 Selection applied to the tag and probe muons and on the muon pair to select the $J/\psi \to \mu\mu$ events entering the HLT-efficiency study.

tag-muon	$p_T > 8$ GeV and $\|\eta\| < 2.4$ passes the medium-ID requirements matched to the trigger object which fired the reference path (HLT_Mu8 or HLT_Mu4_L1DoubleMu)
probe-muon	$p_T >$2.0 GeV $\|\eta\| < 2.4$ passes the medium-ID requirements
muon pair	$\Delta z(\mu_1, \mu_2) < 1.0$ cm SV probability $> 0.05\%$ 2.9 GeV $<$ M$(\mu_1, \mu_2) <$ 3.3 GeV
$p_T \times \|\eta\|$ bins	[0,1,2,3,4,5,6,7,8,9,10,12,14,16,18,20,25,30,50] GeV $\times$ [0, 0.9, 1.2, 2.4]

The tag and probe method measures the efficiency for a di-muon pair to fire a di-muon trigger. For each $\tau \to 3\mu$ candidate, the efficiency that at least 2 out of the 3 muons fire the HLT_DoubleMu_4_3_LowMass trigger is defined as

$$\varepsilon(3\mu) = 1 - (1 - \varepsilon_{12}) \cdot (1 - \varepsilon_{13}) \cdot (1 - \varepsilon_{23}), \tag{3.11}$$

$$\text{where } \varepsilon_{\mu_i,\mu_j} \equiv \varepsilon_{ij}. \tag{3.12}$$

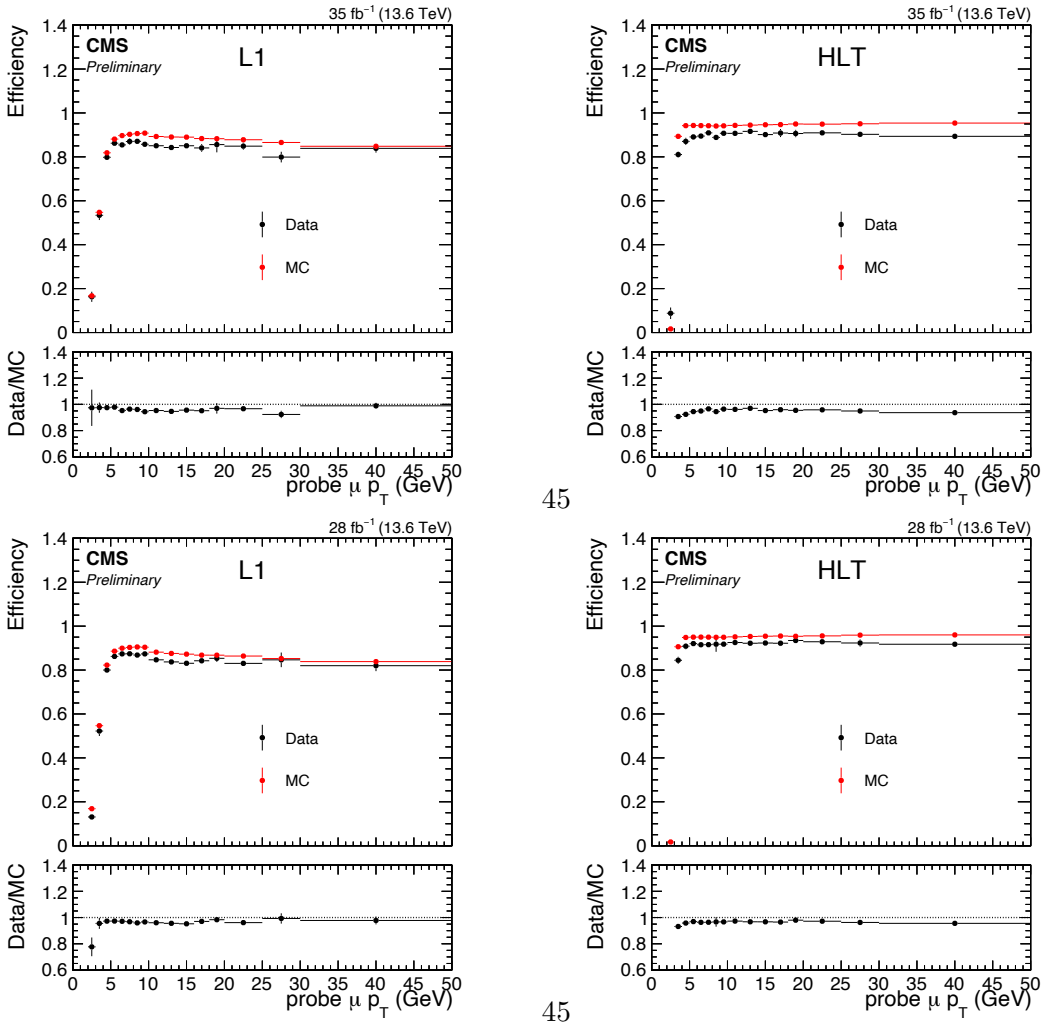


Figure 3.18 Data and MC efficiency as a function of the probe-μ p_T, inclusively w.r.t. $|\eta|$, for the L1 (left) and the HLT (right) component of the `HLT_DoubleMu_4_3_LowMass` trigger for 2022 (top row) and 2023 (bottom row) datasets. The bottom panel in each plot reports the data to MC efficiency ratio. For MC only statistical uncertainties are calculated, while the statistical and systematic uncertainty are added in quadrature in data.

Where ε_{ij} is measured as a function of the probe muon p_T and $|\eta|$ as described in Eq. 3.10. The efficiency scale factor is derived event wise as the ratio between the measured $\varepsilon(3\mu)$ efficiency in data and in MC. The plot in Fig. 3.19 reports the average trigger scale factor as a function of the τ lepton p_T.

3.6.4 Vector boson kinematics

The MC signal samples for the $W \to \tau(3\mu)\nu$ process are generated with PYTHIA8 with LO precision (see Sec. 3.2.1). However, the efficiency of the selection steps is expected to depend on the kinematic properties of the generated particles, especially those of the W boson.

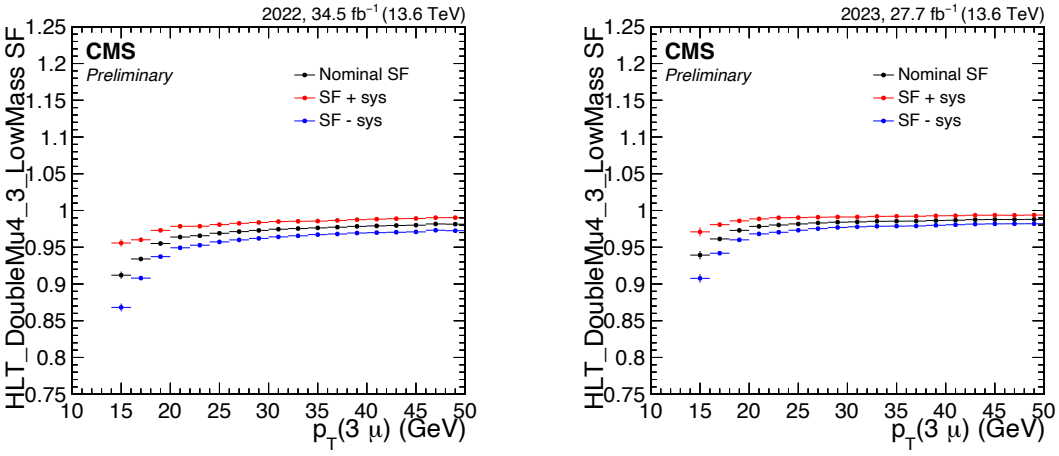


Figure 3.19 Profile of the trigger efficiency scale factor for $\tau \to 3\mu$ candidates selected by the `HLT_DoubleMu_4_3_LowMass` trigger as a function of the τ p_T. The nominal correction is shown with black points. Its value shifted up and down by 1σ is shown in red and blue dots, respectively.

To improve the description of the W boson kinematics we use MC samples produced using MADGRAPH5_aMC@NLO v.2.9.9 at next-to-leading order (NLO) precision with up to two outgoing partons at Born level. PYTHIA8 is used for the hadronization step, with the same configurations discussed in Sec. 3.2.1. The generated vector boson spectrum is compared between the MADGRAPH and PYTHIA8 samples in a p_T and η plane to derive the scale factors as the ratio between the NLO and the LO predictions. The scale factors maps are reported in Figure 3.20. The NLO re-weighting is applied event by event to the signal MC samples based on the kinematic properties of the generated W boson. In regions where the scale factor is not defined, due to the poor statistics in either the LO or NLO sample, the scale factor is set to 1. The plot in Figure 3.21 shows that the NLO re-weight brings an almost flat correction to the three muons η distribution, hence it does not affect the categories composition, while it makes the τ p_T spectrum harder.

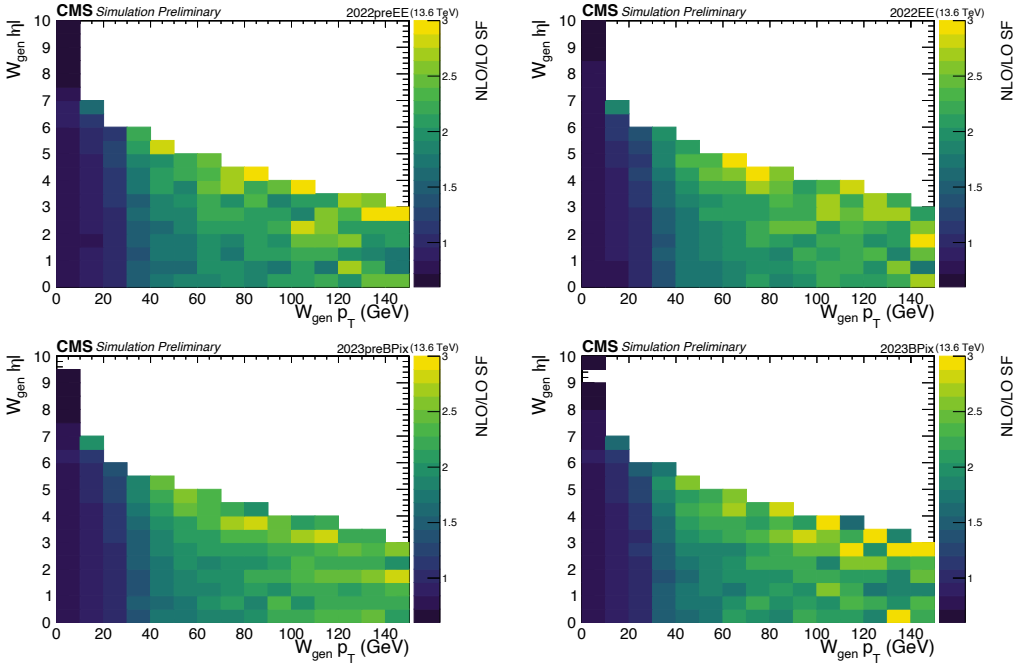


Figure 3.20 NLO/LO scale factors for $W \to \tau(3\mu)\nu$ signal simulated samples using 2022 (top) and 2023 (bottom) conditions and for the two data-taking periods in which each year is divided (left and right).

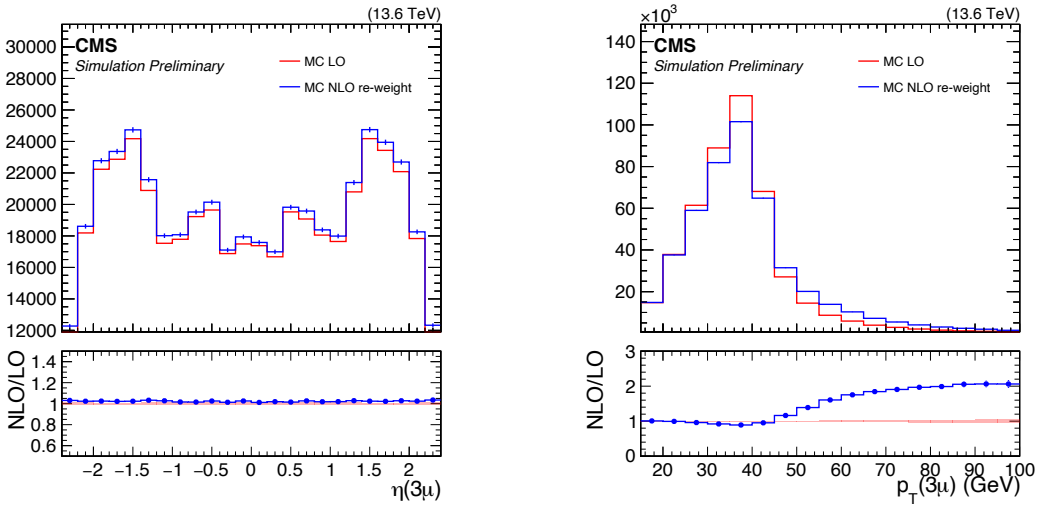


Figure 3.21 NLO re-weighting effect on the $\tau \to 3\mu$ candidates η and p_T in the 2023 $W \to \tau(3\mu)\nu$ MC signal sample.

Chapter 4

Background reduction strategy

As a consequence of the preselection described in Sec. 3.4, only a single $\tau \to 3\mu$ candidate is retained per event. However, this selection is not yet specific enough to isolate $W \to \tau(3\mu)\nu$ signal candidates in data, and a large fraction of the selected τ candidates still originates from background processes. A portion of the background arises from the non-resonant SM decay $W \to 3\mu\nu$, in which two additional muons are produced via photon emission in $W \to \mu\nu$ events. The dominant background contribution, however, is of combinatorial nature, consisting of events in which three unrelated muons are combined to form a $\tau \to 3\mu$ candidate.
To suppress these two types of background contributions, both a cut-based and a multivariate selection technique are employed and simultaneously optimized. The SM background is reduced using a cut-based selection, while the residual combinatorial background is reduced using a binary BDT classifier. The BDT algorithm developed for this analysis, together with its performance and validation tests, is presented in Sec. 4.1, whereas the specific strategies for suppressing the $W \to 3\mu\nu$ background is detailed in Sec. 4.2. The contamination from dimuon resonances within the muon triplets is discussed in Sec. 4.3. In Sec. 4.4, the potential contribution from specific background processes exhibiting a resonant structure around the τ lepton mass, not necessarily modeled by the data sidebands, is investigated. Finally, the selection strategy based on the BDT output is validated by comparing data and MC for $D_s^{\pm} \to \phi(\mu\mu)\pi^{\pm}$ events, which serve as a control sample, as discussed in Sec. 4.5.

4.1 Combinatorial background

In this analysis, the largest background arises from combinatorics, i.e. τ candidates built from unrelated muons. The composition, characteristics and kinematics of this background cannot be reliably derived from simulations, therefore the most effective modeling is through data-driven methods. The $\tau \to 3\mu$ candidates selected in data with three-muons reconstructed invariant mass in the tau mass sidebands region, defined as the disjoint intervals $m_{3\mu} \in [1.40, 1.74]$ GeV and $m_{3\mu} \in [1.82, 2.05]$ GeV, are used to estimate the expected background contribution in the signal region. Events in the data sidebands are most likely background events and provide

an empirical description of it, which is also, by definition, the most inclusive and unbiased available.
The final step for the $W \to \tau(3\mu)\nu$ selection is performed using a BDT algorithm, after the preselection is applied (see Sec. 3.4) and the candidates having L_{xy}/σ below 2.0 are removed to reduce the contamination from prompt processes like $W \to 3\mu\nu$, as described in Sec. 4.2. In the following sections, after a general introduction about BDTs, the architecture and performance of the BDT algorithm specifically used for the rest of the analysis, are described.

4.1.1 General concepts on Boosted Decision Trees

In high-energy-physics analyses like the one presented in this work, multivariate approaches generally achieve significantly better performances in classification problems of separating potential signal events from background, if compared to simpler cut-based techniques. This is because a machine learning (ML) algorithm is designed to exploit correlations between different observables in high dimensional spaces improving the overall performance.
In this analysis, a BDT algorithm is employed to distinguish $\tau \to 3\mu$ signal events from the inclusive background modeled upon the data sidebands, and it is implemented using the XGBoost(Extreme Gradient Boosting) library [79].
This section presents a description of the general BDT architecture designed for this analysis. BDTs are based on the elementary concept of a binary decision tree, i.e. a decision graph that takes a set of observables $\mathbf{x} = (x_1, x_2, ..., x_m)$ in input and attributes it to one of the categories of the classification problem. In this study case, for each $\tau \to 3\mu$ candidate the input set $\mathbf{x}$ is represented by the input features listed in Tab. 4.1 and the output categories are signal and background.

A single decision tree consists of a layered structure of decision nodes. At each node the events are split into two child-nodes, belonging to a deeper layer, according to a splitting criterion. In the i-th node, the j-th feature of a given event is evaluated. The event is passed down to one of the two child nodes depending on whether the variable x_j lies above or below the node-threshold $\theta_i(x_j)$. Events propagate through successive layers of the tree until they reach a terminal leaf node, which is associated with one of the classification categories. This architecture extends a simple cut-based analysis into a multivariate technique by continuing to analyze events that fail a particular criterion. Instead of discarding such events immediately, the decision tree checks whether additional criteria can still lead to a proper classification [80]. This concept is schematically illustrated in Fig. 4.1. The idea behind the BDTs is to construct ensemble of decision trees, i.e. an ensemble of a large number (N_T) of decision trees, each with its own structure. When the forest is evaluated on a given event, N_T independent binary classification responses are produced. The final classification is then obtained by combining the individual responses.

Each classifier is represented as a function $g_n(\mathbf{x}, \Theta_n)$ of the feature vector $\mathbf{x}$ and the set of node thresholds Θ_n [81] and the classification output $\hat{y}(\mathbf{x})$ is a linear combination of the single trees classifications. For instance, the simplest way to compute the forest response for a given event is from the average of the N_T classifiers outputs. In this way, the outcome is not binary but a continuous value ranging

between 0 and 1, representing the background and signal hypothesis, respectively and it is defined as

$$\hat{y}(\mathbf{x}) \equiv F(\mathbf{x}; g_1, ..., g_{N_T}) = \frac{1}{N_T} \sum_{n=1}^{N_T} g_n(\mathbf{x}, \Theta_n). \tag{4.1}$$

Random forests are significantly more powerful than single decision trees because as they are less sensitive to statistical fluctuations. It can be shown that the variance of the classification error, due to the finite size of the training set, decreases as the number of classifiers in the forest increases, according to

$$\mathrm{Var}(\mathrm{Err}[\hat{y}]) = \mathrm{Var}(g) \cdot \left(\rho(\mathbf{x}) + \frac{1 - \rho(\mathbf{x})}{N_T}\right), \tag{4.2}$$

where $\rho(\mathbf{x})$ is the correlation degree between the trees in the forest, built on the same learning set, and $\mathrm{Var}(g)$ is the variance of the predictions obtained using a single tree. The theorem resulting in the Eq. 4.2 demonstrates that as the ensemble gets arbitrary large, i.e. as $N_T \to \infty$, the variance of the ensemble reduces to $\mathrm{Var}(g)\rho(\mathbf{x})$. Being the correlation $\rho(\mathbf{x}) < 1$, the variance of an ensemble is strictly smaller than the variance of an individual model [82].

Finally, the model is trained on a representative dataset containing both signal and background events, namely the training set, using the supervised learning approach. For each event the model receives as input both the features set $(x_1, x_2, ..., x_m)$ and the correct classification (y), i.e. a binary label indicating whether the candidate is signal or background. Goal of the training is to determine the optimal value of forest free parameters Θ which are the split threshold for each node in the N_T trees.

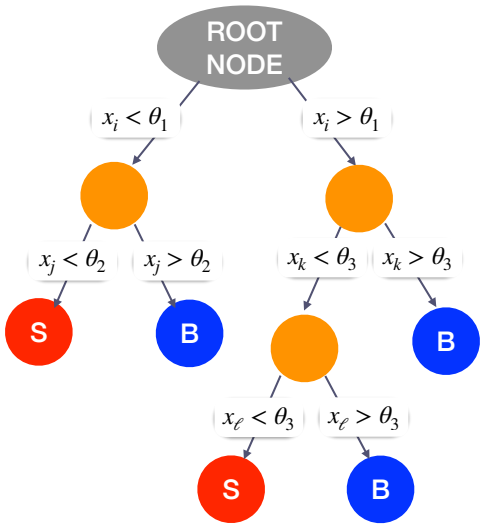


Figure 4.1 Graphical representation of a decision tree. The circles are the internal nodes, with their associated splitting criterion. The classification nodes are represented in red and blue for background and signal category, respectively.

The optimal configuration is found by maximizing a objective function which can be written in general form as

$$\text{obj}(\Theta) = L(\Theta) + \Omega(\Theta), \tag{4.3}$$

where $L(\Theta)$ is the training loss function, measuring how well the model predicts the training data, and $\Omega(\Theta)$ is the regularization term, that explicitly quantifies the model complexity and penalizes overly complex configurations during training. In the binary classification designed for this analysis the most convenient definition for the loss-function is the binary cross-entropy defined as

$$L(y_i, \hat{y}_i) = \frac{1}{N} \sum_{i=1}^{N} \left[y_i \cdot \log(p_i) + (1 - y_i) \cdot \log(1 - p_i) \right], \tag{4.4}$$

$$p_i = \frac{1}{1 + e^{-\hat{y}_i}} \in [0, 1], \tag{4.5}$$

where N is the number of candidates, y_i is the true class label and p_i is the probability of the i-th candidate to belong to the signal class, defined as the sigmoid function of the model prediction $\hat{y}_i$.
The training proceeds in steps, called epochs, following the boosting approach, whose general idea is to train single decision trees sequentially, where each new tree focuses on correcting the errors of the previous ones. One of the simplest applications of the boosting is to provide to one tree a large fraction of events that were misidentified from the forest after the previous iteration, thus improving the model accuracy. In this additive learning approach only the last-added tree parameters are optimized, while those of the earlier trees remain fixed. After the t-th epoch is completed the prediction on the objective function and the corresponding prediction on the i-th candidate can be expressed as

$$\text{obj}^{(t)} = \sum_{k=0}^{N} l(y_k, \hat{y}_k^{(t)}) + \sum_{n=1}^{t} \omega(g_n) = \text{obj}^{(t-1)} + \sum_{k=0}^{N} l(y_k, \hat{y}_i^{(t)}) + \omega(g_t), \tag{4.6}$$

with

$$\hat{y}_i^{(t)} = F(\mathbf{x}; g_1, ..., g_t) = \hat{y}_i^{(t-1)} + \alpha_t \cdot g_t(\mathbf{x}_i, \Theta_t), \tag{4.7}$$

where the $\omega(g_n)$ quantifies the n-th tree complexity, thus penalizing too complex trees. The learning procedure is stopped once the objective function fails to improve by more than a predefined amount, specified in the training configuration.
After the training, the classifier ensemble output is as a scalar value, ranging between 0 and 1, resulting from a linear combination of the N_T tree responses. Each tree is weighted with a parameter α_n whose value depends on the classifier accuracy obtained during the training. Therefore, higher is the weight, more relevant is the single classifier in determining the final response and the boosting method leverages more on the most accurate members of the ensemble.

Finally the ensemble output value, that will be referred to as the BDT-score in the following, can be interpreted as the probability for the given event to be signal:

$$\text{BDT score}(\mathbf{x}) = \sum_{n=1}^{N_T} \alpha_n \cdot g_n(\mathbf{x}, \Theta_n). \tag{4.8}$$

In addition to the free parameters, the BDT structure relies on hyperparameters: additional settings that are defined before the training. Their value can sensibly impact the performance of the algorithm and controls the bias–variance trade-off in the training. Some hyperparameters act on the definition of the regularization term $\Omega(\Theta)$ in the objective function of Eq. 4.5, thus directly controlling the complexity penalization.
ML algorithms are subject to underfitting and overfitting, which must be carefully controlled. The former occurs when the model is too simple to adequately describe even the training set, while the latter arises when the model is too complex, learning the training data too precisely and thus failing to generalize when applied to unseen data. This is often referred to as the bias–variance trade-off. The difficulty of training a ML algorithm is to simultaneously minimize both the bias, i.e. the difference between the model predictions and the true classification, and the variance, i.e. the variability of the model prediction for a given event [80].
To further prevent overtraining part of the training data are set aside during the learning phase, the BDT performance is then evaluated on both the training and the validation sets, using the Receiver Operating Characteristic Area under the Curve (AUC) metric. The objective function is further penalized by the discrepancy between the validation metric computed on the training and validation samples, thereby limiting bias reduction in favor of improved model generalization.

4.1.2 Analysis BDT architecture

The algorithm is trained with a signal sample of $W \to \tau(3\mu)\nu$ simulated candidates passing the preselection criteria and the SV displacement selection detailed in Sec. 4.2. The background sample consists of the $\tau \to 3\mu$ candidates reconstructed in data, passing the same selection as signal, and laying in the data sidebands region.

In simulated samples, the number of τ^+ and τ^- candidates are present in different proportion according to the W boson charge asymmetry in proton-proton collisions. The charge unbalance arises from the presence of two valence u-quarks and only one valence d-quark in the proton, hence larger is the probability to produce W bosons in $u\bar{d} \to W^+$ than $d\bar{u} \to W^-$ process, since the proton PDFs are sensibly larger for valence quarks than for sea quarks. The same unbalance is not present in data sidebands because the τ lepton candidates are not necessarily produced in W decays, but mostly from unrelated muons. Even though the charge asymmetry is a physical effect, it is not a LFV $\tau \to 3\mu$ signal feature per se, rather a feature of the W boson production. The BDT training should not correlate the probability for a τ to be a LFV signal candidate with its charge because it has no strong correlation with the signal likeness. Therefore, to prevent the $\tau \to 3\mu$ candidate charge from driving the algorithm learning, the negative charge candidates in the signal samples are weighted in the training with the ratio between the number of τ^+ and τ^- candidates, which is about 1.32. Such re-weight is applied only to train the BDT and the signal MC is not reweighted by the charge asymmetry in the rest of the analysis.
In the training the signal sample is also weighted with the NLO scale factors described in Sec. 3.6.4 to prevent the BDT from learning upon artificial differences due to the LO precision simulated in MC.

As described in the previous section, a BDT classifies events by taking, for each one, a set of observables as input features. These features must provide discriminating power between signal and background events. For the purposes of this analysis, a set of 19 input features with discrimination power between the $\tau \to 3\mu$ signal and the combinatorial background is selected. These variables describe the muon tracks topology and kinematics, the kinematics of the τ lepton and the W boson, as well as secondary-vertex quantities obtained from the kinematic fit. The complete list and description of the input features is provided in Table 4.1 and most of them have been extensively defined in Sec. 3.4. The distribution of the most discriminant input features are shown in Fig 4.2 for $W \to \tau(3\mu)\nu$ simulated samples and background from data sidebands.

Table 4.1 Features used as input to the BDT.

Feature	**Description**
Tight-ID(μ_i)	Boolean variable indicating whether the i-th muon passes the tight muon ID criteria. [Sec. 2.3.2]
$\Delta z(\mu_i, \mu_j)$	Longitudinal distance between two muons in the triplet, evaluated with respect to the PV position. [Fig. 3.3]
$p_T(3\mu)$	$\tau \to 3\mu$ candidate reconstructed transverse momentum.
$\lvert\eta(3\mu)\rvert$	Absolute $\tau \to 3\mu$ candidate pseudorapidity - floating values are mapped into 7 integer bins
$I_{\text{rel}}(3\mu)$	$\tau \to 3\mu$ candidate isolation normalized to its p_T. [Eq. 3.3]
SV L_{xy}/σ	Displacement significance, in the transverse plane, of the τ lepton SV with respect to the beam-spot. [Fig. 3.6]
SV probability	χ^2 probability associated to the kinematic fit to a common SV.
SV $\cos(\alpha)$	Cosine of the angle between the τ lepton momentum and the vector connecting the beam-spot and the SV in the transverse plane. [Fig. 3.6]
MET	Missing transverse momentum.
$p_T(\tau)$/MET	Ratio between the τ candidate transverse momentum and MET.
$\Delta\phi(\tau$, MET)	Azimuthal angular distance between the τ candidate and MET.
$m_T(\tau$, MET)	Transverse mass for the combination of the τ momentum and the MET in the transverse plane. [Eq. 3.7]
p_T(W)	W boson candidate p_T.
min p_z^ν	Minimum solution for the missing longitudinal momentum. [Eq. 3.6]
max p_z^ν	Maximum solution for the missing longitudinal momentum. [Eq. 3.6]

The BDT-classifiers share the same hyperparameters which are tuned using an inclusive training on the full training sample and maximizing the BDT classification power using the AUC metric. The optimization tool Optuna [83] is used, it performs

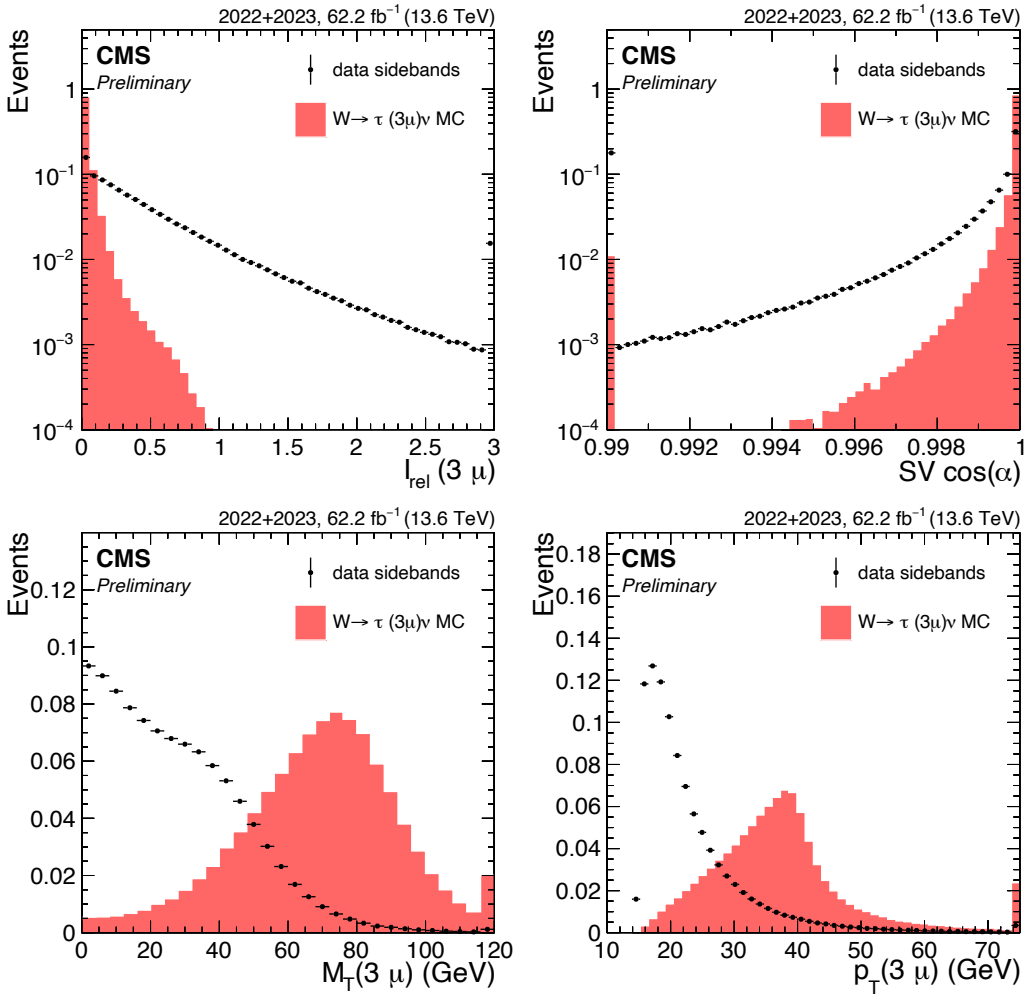


Figure 4.2 The distributions of four out of the nineteen input features of the analysis BDT for the $W \to \tau(3\mu)\nu$ MC (red histograms) and for data in the sidebands (black dots). From left to right top to bottom: the relative isolation, the cosine of the angle α, the transverse mass and the τ candidate transverse momentum.

a Bayesian search over the hyperparameters space. The hyperparameters optimal values and optimization range is summarized in Tab. 4.2.

In order to reduce the risk of overfitting bias the BDT is not applied to the same dataset used for its training. The risk in case of overfitting is that the classifier may simply memorize the features of the training events rather than learning the general patterns that distinguish signal from background. As a result, the performance obtained on the training sample would be too optimistic, while the true predictive power on independent data could be significantly worse. To obtain an unbiased estimate of the classifier's performance, the BDT must be evaluated on a statistically independent dataset. To limit the overfitting bias and, at the same time, to exploit the full available statistics, the analysis BDT is trained using a k-fold cross-validation technique. The combined signal and background training dataset is split into five subsets, or folds, and five distinct classifiers are trained. Each BDT is trained on four

Table 4.2 The hyperparameters of the analysis BDT are listed below. An asterisk (*) marks those whose value increases the model complexity [79]. The second and third columns indicate the value used in training and the range explored during optimization, respectively.

Parameter	Value	Range	Description
max depth*	5	[3, 7]	maximum depth of a single tree
estimators* ($\mathbf{N}_T$)	15000	[1k, 20k]	maximum number of trees, i.e. the maximum number of epochs
learning-rate	0.005	[10^{-3}, 1.0]	regulates the difference between weights of two subsequent tree
subsample*	0.6	[0.5, 1.0]	training data fraction randomly sampled to grow the tree of a given epoch
colsample-bytree*	0.7	[0.5, 1.0]	features fraction used to grow the tree of a given epoch
min. child weight	70	[10, 100]	lower bound on number of events in a node required to undergo further splitting
gamma	7.75×10^{-6}	[10^{-5}, 1.0]	minimum loss reduction required to make a further partition on a leaf node of the tree
reg-alpha	6.60×10^{-5}	[10^{-5}, 1.0]	L1 regularization parameter in the objective regulation term
reg-lambda	0.037	[10^{-5}, 1.0]	L2 regularization parameter in the objective regulation term

out of the five subsets, i.e. on 80% of the events, and then applied to the remaining 20%, which serves as the application set. This approach ensures that the BDT is always applied to an event sample entirely independent from the one it was trained on. Furthermore, 20% of the training set is used as validation set to evaluate the AUC metric during the training, controlling overfitting.

4.1.3 BDT performance

The BDT performances are quantified upon the discrimination power of its score between signal and background $\tau \to 3\mu$ candidates.

The distribution of the BDT output, after the training, is shown in Fig. 4.3. A first qualitative test for overtraining consists in comparing the BDT classification obtained during the trainings and the one used in the analysis. Each event is associated with a training score, defined as the average between the scores of the four BDTs using the given event in the training. The analysis score represents the BDT output used in the analysis. The training and the analysis distribution show good agreement both for the data sidebands and the signal samples without strong

indication for the presence of overtraining.
The BDT classification performance is quantified using the signal and background efficiency as a function of a selection working point, see Fig. 4.4. The Receiver Operating Characteristic (ROC) curves for the training and the application set are compared to check for overtraining, the performance on the training set are slightly better but the overtraining is under control. The per-category ROC curves and relative AUCs indicate signal-versus-background separation is more efficient in category A, corresponding to τ leptons in the central pseudorapidity region, and slightly degrades for forward τ candidates, belonging to category C (see Sec. 3.5).

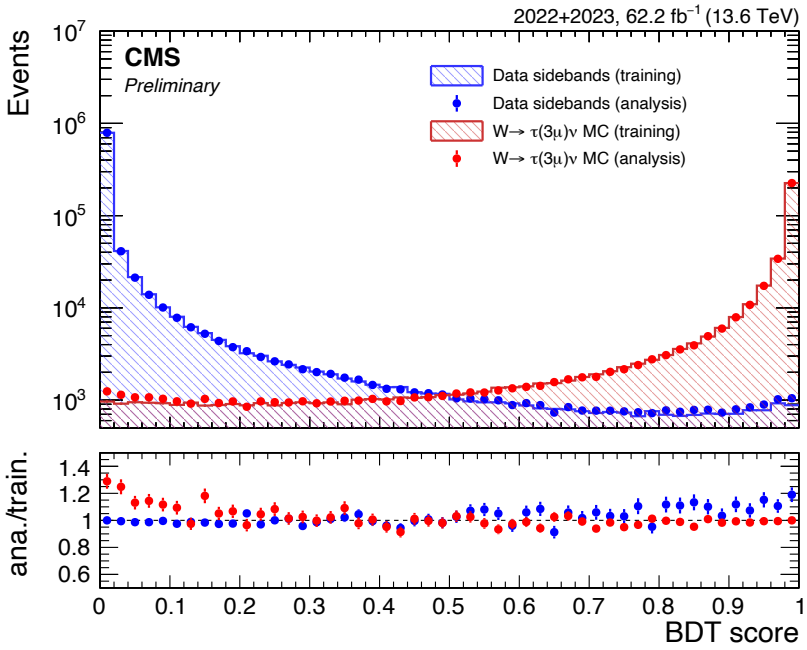


Figure 4.3 Distribution of the BDT score for signal (red) and data sidebands (blue) events comparing the training score (dashed-filled histograms) and on the analysis score (full markers).

The BDT relies on 19 observables to assign the signal probability to each event. A deeper understanding of the BDT algorithm is provided by the ranking of the input features according to their importance in the BDT decision-making. The input features ranking, according to their relative F-score, is shown in the left plot of Figure 4.5. The relative F-score is a measure of the predictive performance for a certain feature and it corresponds to the number of times a feature is used to split the data across all trees within the model divided by the total number of splits. The most relevant features, according to this metric, are those describing the decay topology and the SV fit quality. This definition actually correlates the importance of a feature with its cardinality, i.e. the number of values it can assume. Therefore it artificially assigns lower rank to boolean or integer features like the muon Tight-ID. The larger is the number of values the feature can assume, the more are the ways the variable can be split, the higher is the ranking. An alternative way to evaluate the feature importance consists to evaluate the impact on the model accuracy when the

values of an input feature are permuted across candidates in the sample, thus losing the correlation with the event true classification [84]. The larger is the decrease in performance when a given feature is permuted, the larger its importance. The permutation importance is reported in the left plot of Fig. 4.5. This metric weekly depends on the observable's cardinality, indeed the third-muon Tight-ID has much higher ranking compared to the F-score method. According to the permutation importance metric the most important features are the τ candidate isolation, the momentum alignment with respect to the displacement and the transverse mass.

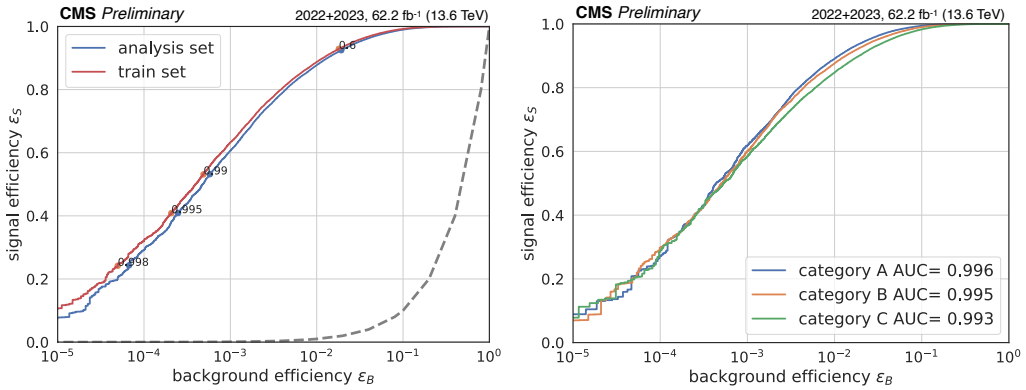


Figure 4.4 On the left, the ROC curves compared between the training set (red) and the analysis set (blue), the dashed gray line is the bisector representing the random classifier performance. On the right, the ROC curves for the analysis set splitting the events according to the analysis event-categories.

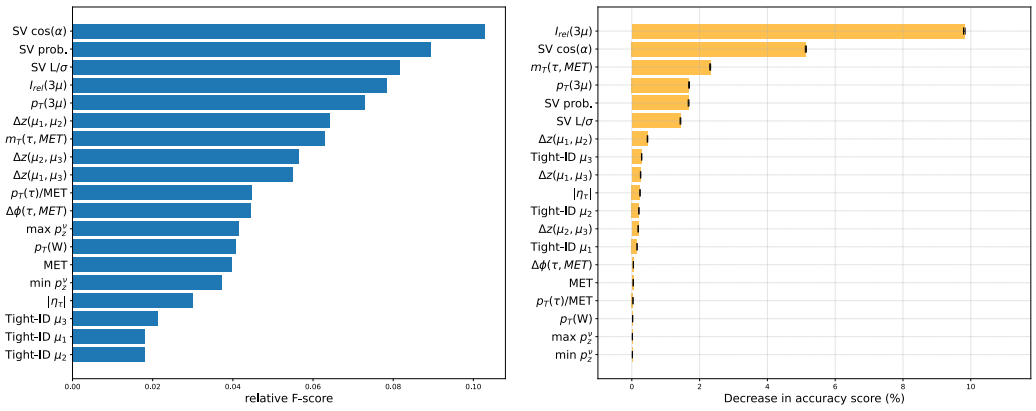


Figure 4.5 Features importance, expressed through the relative F-score (left) and through permutation importance (right).

4.1.4 BDT bias tests

Before using the BDT output in the analysis to isolate potential $\tau \to 3\mu$ signal candidates, it is crucial to test for the presence of effects that could bias the signal sensitivity estimation.

Since the analysis final goal is to fit the muon triplet invariant mass distribution, it is crucial to ensure that the a cut on the BDT score does not introduces fictitious shaping in its distribution, especially in the signal region, where they could create fake signals. In other words it must be checked that the BDT score does not correlate with the mass of the $\tau \to 3\mu$ candidate. This effect is investigated studying the linear correlation and the distance correlation [85] among the features, the BDT score and the invariant mass and no significant correlations are observed, looking at the last two rows of the matrices in Fig. 4.6.
Further mass-sculpting effects are investigated studying the BDT selection efficiency as a function of the $\tau \to 3\mu$ candidate reconstructed mass. This is studied using two background samples: the data sidebands and the $W \to 3\mu\nu$ MC. No shaping is observed, as demonstrated by the trends in Fig. 4.7 for different cuts on the BDT score. A further test has been done on signal samples generated changing the τ lepton mass between 1.65 GeV and 1.95 GeV and comparing the efficiency of a tight selection on the BDT score, i.e. above 0.990. The efficiencies are reported as a function the simulated τ lepton mass in Figure 4.8, and they are all compatible with the efficiency calculated at the nominal mass, within 3σ, showing no significant correlation between the selection efficiency and the muon triplet invariant mass. Figure 4.9 reports the BDT selection efficiency in simulated samples, for different values of the threshold on the BDT score, combining the $\tau \to 3\mu$ simulated samples at different τ masses. The three analysis categories are studied separately to disentangle effects due to the correlation between the BDT efficiency and the candidates pseudorapidity, showing no significant mass-sculpting effects.

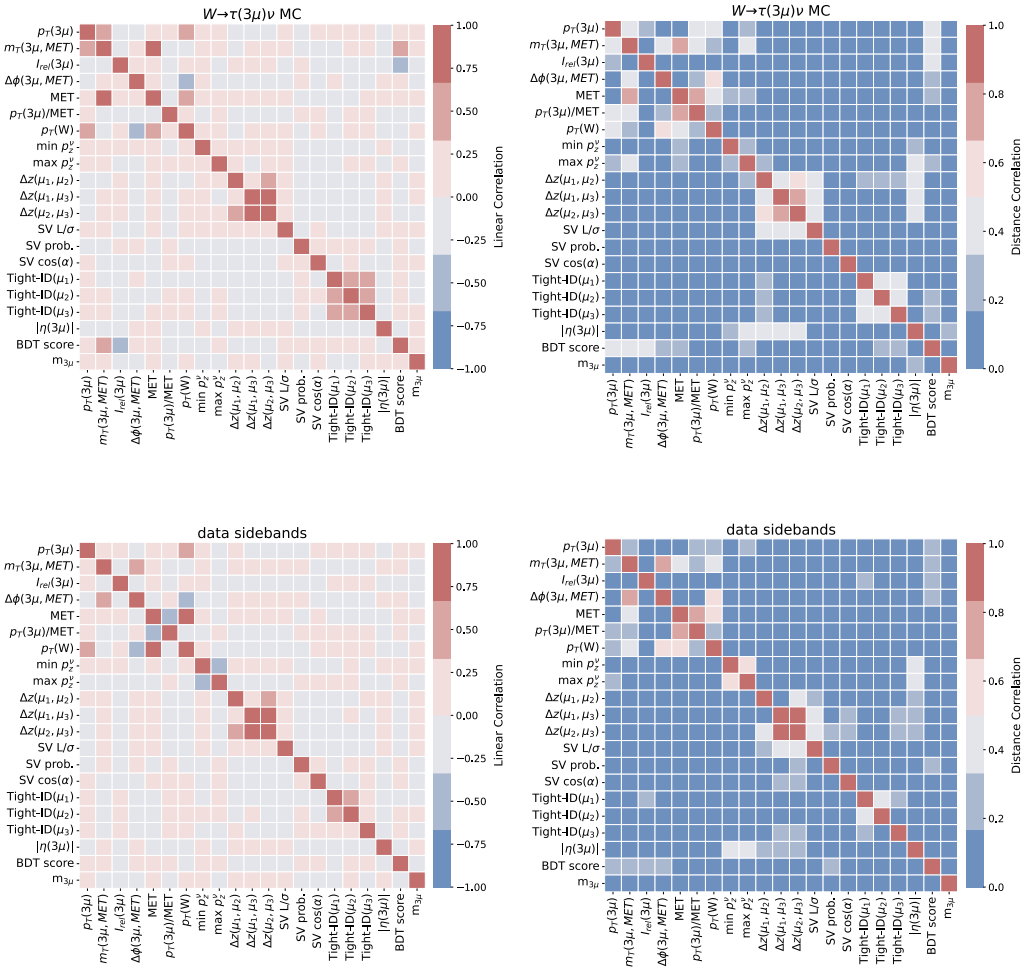


Figure 4.6 Linear (left) and distance (right) correlation between the BDT input features, its score and the muon triplet reconstructed mass in the $W \to \tau(3\mu)\nu$ MC (top) and in data sidebands (bottom).

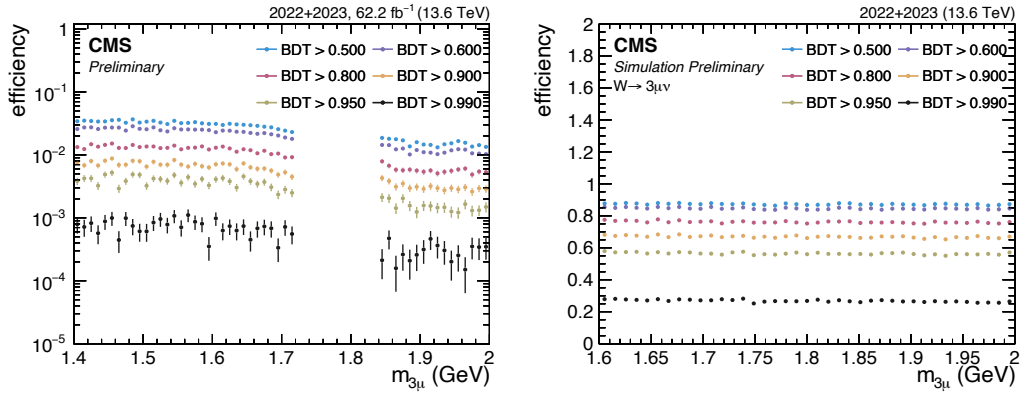


Figure 4.7 BDT selection efficiency for several values of the threshold on the BDT score, on data sidebands (left) and $W \to 3\mu\nu$ MC over the full mass range (right), in the latter sample the cut on L_{xy}/σ has been removed.

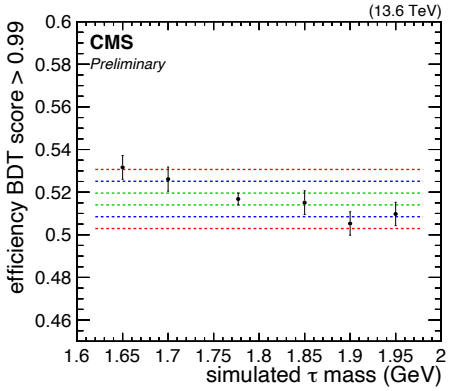


Figure 4.8 BDT selection efficiency requiring BDT score > 0.990, evaluated on simulated $\tau \to 3\mu$ samples with different τ lepton mass. The green, blue and red lines respectively indicate the 1σ, 2σ and 3σ interval around the efficiency value at the nominal τ mass.

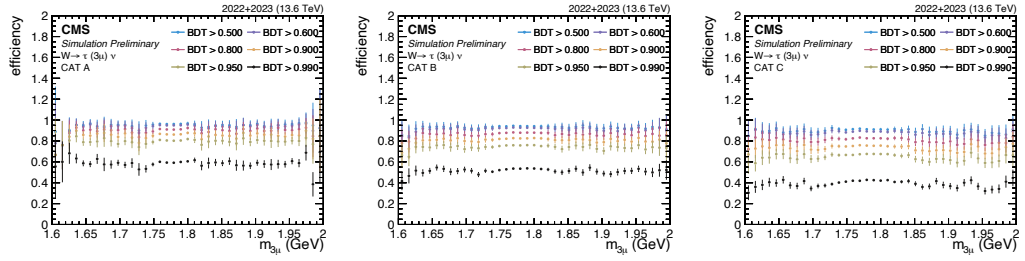


Figure 4.9 BDT selection efficiency for several thresholds on the BDT score, on simulated samples for $\tau \to 3\mu$ with τ lepton mass varying between 1.65 GeV and 1.95 GeV. To disentangle resolution effects the muon triplet mass spectrum is showed separately in the analysis category A (left), B (middle) and C (right).

4.1.5 Working points definition

In the final step of the analysis, $\tau \to 3\mu$ signal candidates are selected as potential signal events or rejected as background based on the BDT score and only $\tau \to 3\mu$ candidates with score higher than a given threshold are used to compute the final result. A different BDT selection working point is defined for each of the six analysis categories, optimizing the expected sensitivity of the given category. The expected signal yield is taken from simulation, while the expected background in signal region is estimated from a fit to the three muons invariant mass in data sidebands, as described in Sec. 5.2. It is found that achieving the best sensitivity requires a very tight BDT cut, typically above 0.990. After this selection, only order of ten events survive in the data sidebands per category. In this regime, the background estimate relies on very limited statistics, and under-fluctuations in the sidebands can bias the optimization of the BDT working points. To mitigate this potential bias, an orthogonal background sample is extracted from data by inverting the Medium-ID requirement on one of the two sub-leading muons (either the sub-leading or trailing muon) in the τ candidate. This inverted-ID sample provides 10 times larger statistics and serves as a proxy for the analysis background. It can thus be used for a more robust optimization of the BDT selection working points, provided that the BDT score distribution is consistent between the data sidebands and the inverted-ID sample. Since a large difference is observed between the two samples, a multivariate morphing is designed to re-weight the inverted-ID sample to match the data sidebands. A binary BDT, hereafter referred to as the morphing BDT, is trained using the $\tau \to 3\mu$ candidates from the data sidebands and those from the inverted-ID sample over the full mass spectrum. The algorithm is trained to minimize the binary cross-entropy (Eq. 4.5) in the binary classification problem and using the analysis BDT input features listed in Tab. 4.1 and score as training features. The muons Tight-ID flags, which are given in input to the analysis BDT, are strongly correlated with their Medium-ID ones. Therefore, a fictitious Tight-ID is assigned to the three muons in the inverted-ID sample which exactly reproduces the distribution in the data sidebands.
The inverted-ID sample is normalized to the yield in data and for each of the τ candidates the score $p(\mathbf{x})$ is interpreted as the probability for a $\tau \to 3\mu$ candidate to be a combinatorial background event of the τ mass sidebands, where $\mathbf{x}$ indicates the set of input observables the morphing BDT is trained on. Consequently, a per-event weight is defined as

$$w(\mathbf{x}) = \frac{p(\mathbf{x})}{1 - p(\mathbf{x})}. \tag{4.9}$$

The successful results of the event-by-event morphing on the BDT score distribution and $\tau \to 3\mu$ candidate mass are reported in Fig 4.10 and 4.11.

The optimal BDT working point is found studying the expected upper limit on the $\tau \to 3\mu$ branching ratio, $\mathcal{B}r(\tau \to 3\mu)$, as a function different lower thresholds on the BDT selection, ranging between 0.985 and 0.999. The technical details for the measurements of the exclusion limits are described in Section 6.1. For the current discussion it is enough to specify the exclusion limit value depends on the expected number of signal and background events in the signal region. The expected signal is

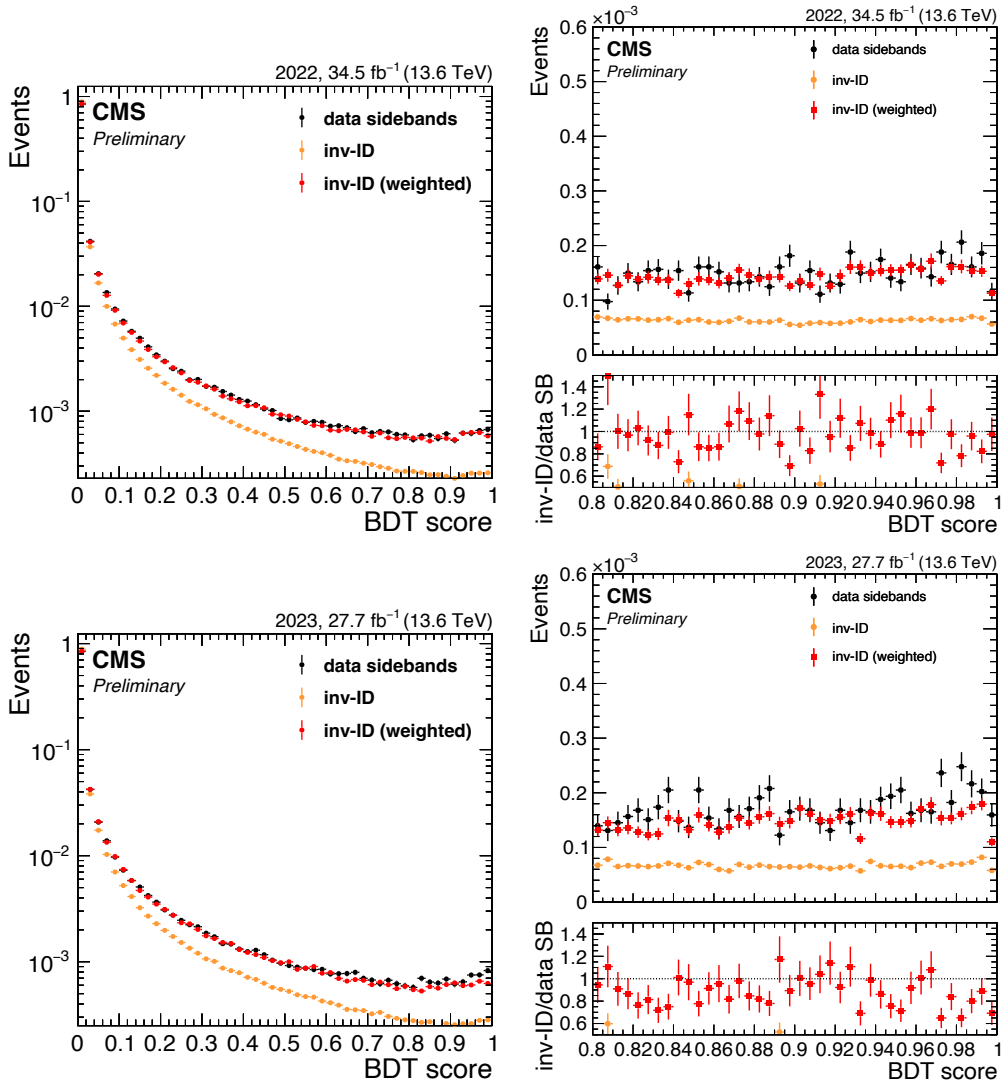


Figure 4.10 BDT score distribution in data sidebands (black), in the inverted-ID sample (yellow) and in the re-weighted inverted-ID sample (red), the 2022 and 2023 data are reported in the top and bottom row, respectively. On the left the BDT score distribution is represented between 0.8 and 1.0 for a better visualization of the interval where the BDT working points are optimized.

estimated from MC, while the background yield in the signal region is extrapolated from an exponential fit to the three muon mass sidebands using the inverted-ID sample, as illustrated in Fig. 4.12. As a comparison, also the signal significance estimated using the "Punzi" figure of merit [86] is studied as a function of the working points which, for a given BDT threshold $\bar{t}$, is defined as

$$\mathcal{S} = \frac{\varepsilon(\bar{t})}{1/2 + \sqrt{B(\bar{t})}}, \tag{4.10}$$

where $\varepsilon(\bar{t})$ indicates the signal efficiency for the given threshold and $B(\bar{t})$ indicates the expected number of background events in the signal region. One of the advantages of

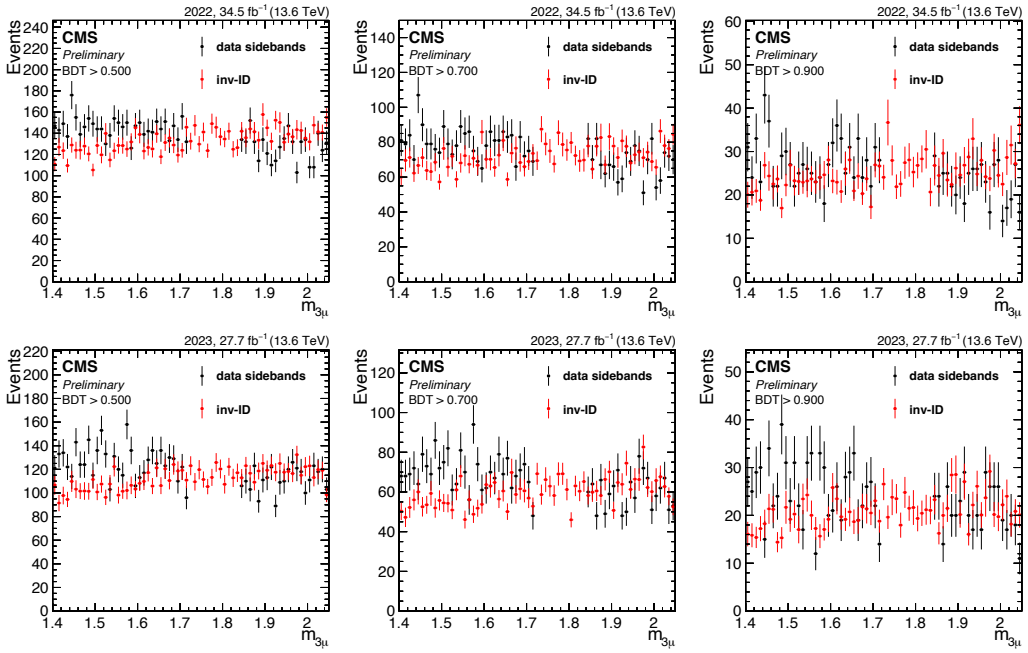


Figure 4.11 Distribution of the $\tau \to 3\mu$ candidates invariant mass in the data sidebands (black) and in the re-weighted inverted-ID sample (red) normalized to yield in data for tighter cut on the BDT score from left to right. The 2022 and 2023 data are displayed in the top and bottom row, respectively.

using this figure of merit is that it does not depend on the signal normalization. For each of the six analysis categories, the working point is chosen as the one yielding the best expected upper limit. The results of the working point optimization are summarized in Tab. 4.3. The expected upper limit (UL) as a function of the BDT threshold, shown in Fig. 4.13, indicates that the optimal working point is achieved for BDT scores above 0.990 across all categories, with a nearly flat dependence of the sensitivity on the BDT selection around this value of the score. For thresholds above approximately 0.996, the selection becomes too tight, and the background modeling uncertainties start to degrade the signal sensitivity.

Table 4.3 Results of the BDT working point optimization for the three analysis categories and the two years. The expected signal and background yields over the full mass spectrum are also reported.

Category		BDT score threshold	exp. $\tau \to 3\mu$ signal	exp. background (inv-ID sample)
2022	A	0.991	2.6	31.3
	B	0.994	2.2	37.4
	C	0.995	0.8	13.1
2023	A	0.993	2.0	27.0
	B	0.993	1.9	32.2
	C	0.988	1.0	38.9

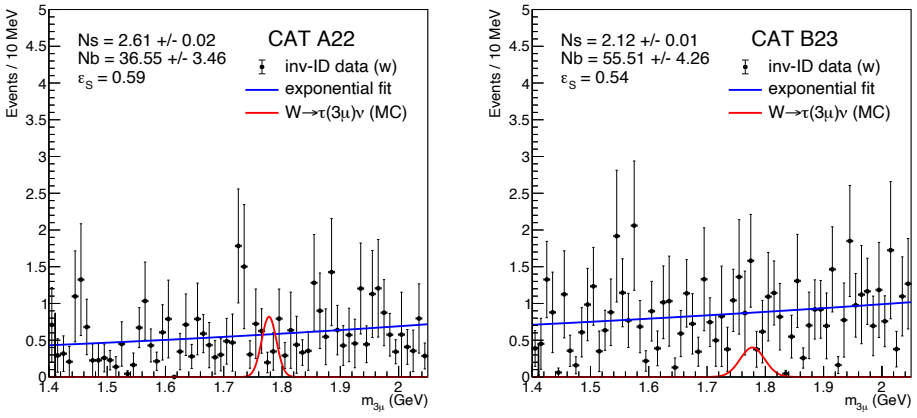


Figure 4.12 Invariant mass distribution of the $\tau \to 3\mu$ candidates in the reweighted inverted-ID sample (black points) and in the simulated $W \to \tau(3\mu)\nu$ sample (red histogram) for category A of 2022 (left) and category B of 2023 (right), after requiring a BDT score greater than 0.990. The background yield is estimated from an exponential fit (blue line) to the τ mass sidebands and extrapolated to the full mass range, while the signal contribution corresponds to the $\tau \to 3\mu$ yield in the normalized MC sample.

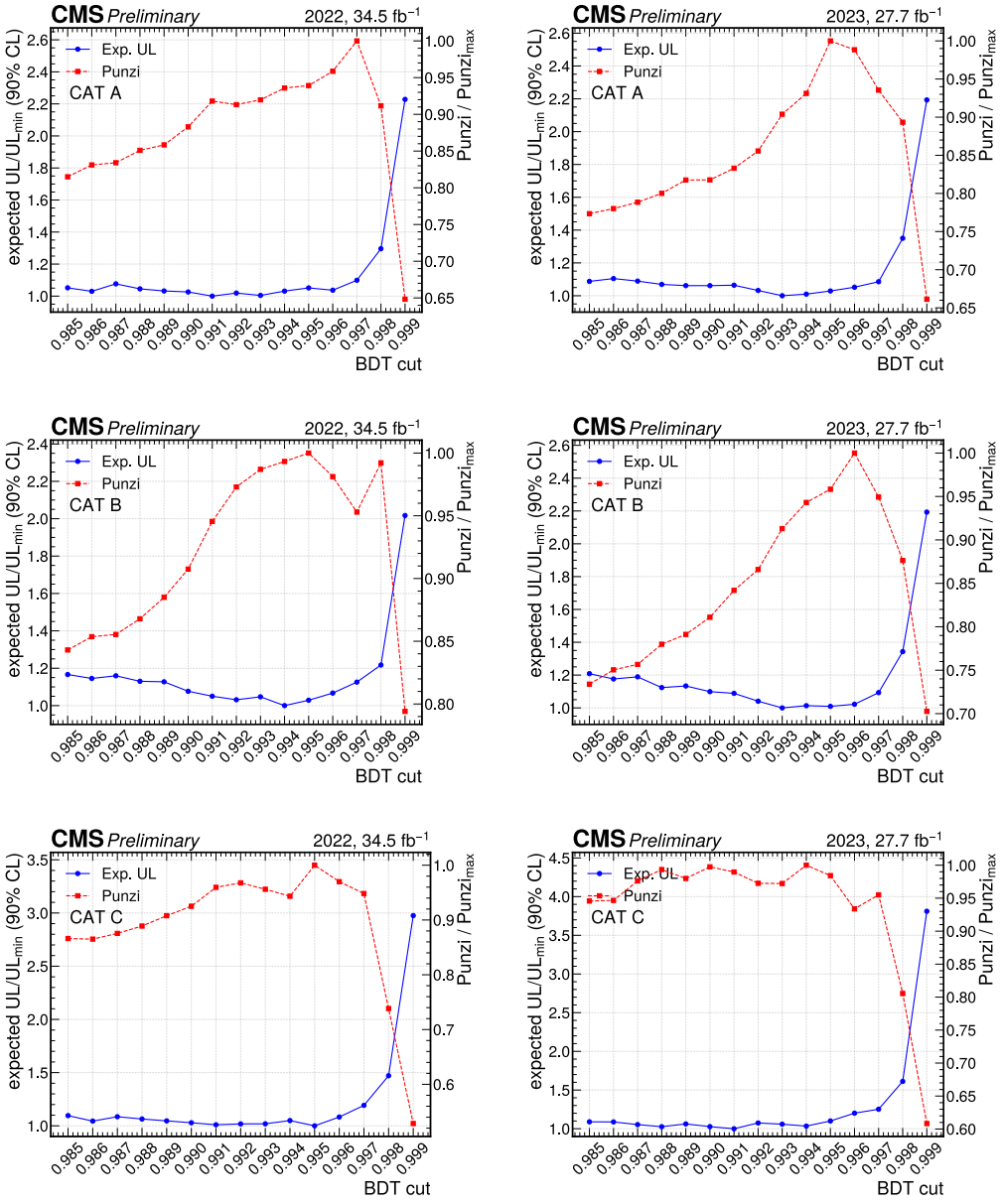


Figure 4.13 Expected 90% CL UL (blue) and Punzi significance (red), both normalized to their best value, as a function of the BDT working point for the six analysis categories. The working point scans are reported for the category A, B and C in the top, middle and bottom row, respectively for 2022 (left column) and 2023 (right column) datasets.

4.2 $W \to 3\mu\nu$ background

A significant fraction of the background contaminating the $\tau \to 3\mu$ signal region originates from the non-resonant SM decay $W \to 3\mu\nu$, which proceeds via the emission of a virtual photon either from initial or final state radiation. The two process, illustrated in the Feynman diagrams of Fig. 4.14, have identical experimental signature.

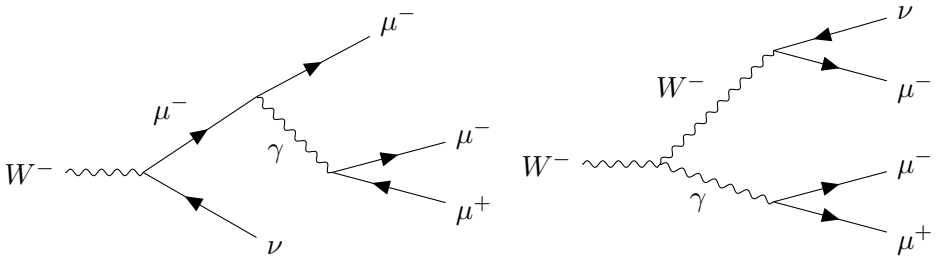


Figure 4.14 The main Feynman diagrams contributing to $W \to 3\mu\nu$ process.

The muons from this process do not form a resonant structure at the value of the τ mass, and their common vertex has a small displacement from the BS, as they originate come from the prompt W boson decay. Apart from these features, the overall kinematics and topology are very similar to those of the $W \to \tau(3\mu)\nu$ process, as shown in Fig. 4.15, where simulated samples for the two processes are compared. Even though the $W \to 3\mu\nu$ mass spectrum is flat, the analysis sensitivity benefits from applying a dedicated selection aimed at rejecting most of these events. A BDT classifier, trained with the architecture and features described in Sec 4.1.1, is found to be ineffective in removing the $W \to 3\mu\nu$ background, due to its strong similarity to the signal. Indeed, after applying a tight selection on the BDT score to retain only signal-like events, a large fraction of the remaining data is observed to cluster in the low SV displacement region, which is well described by $W \to 3\mu\nu$ MC, see Fig. 4.16, thus indicating that the differences between the $W \to \tau(3\mu)\nu$ and $W \to 3\mu\nu$ features are insufficient to drive the BDT into removing this prompt component of the background.

Several different approaches were evaluated and the combination of cuts on the L_{xy}/σ variable and on the output of the BDT discussed in the previous paragraph resulted the most effective one to reduce the $W \to 3\mu\nu$ background contribution. A lower threshold on the muon-triplet L_{xy}/σ is defined to remove the most prompt candidates. Its value is optimized minimizing the expected UL on $\mathcal{B}r(\tau \to 3\mu)$ for 2022. The set of values tested for the L_{xy}/σ cut are: 1.4, 1.5, 1.7, 1.9, 2.0, 2.1 and 3.0. For each value of the displacement cut the BDT was reoptimized and the best working points chosen. Figure 4.17 shows the results of this study: the expected sensitivity slightly improves for L_{xy}/σ thresholds between 1.9 and 2.1, reaching its best value for $L_{xy}/\sigma > 2.1$, then the cut becomes too tight and the sensitivity worsens. Based on these results it was chosen to define the selection on the $\tau \to 3\mu$ secondary vertex displacement significance as $L_{xy}/\sigma > 2.0$ in continuity with the Run2 analysis strategy. All the points of the scan are anyway compatible within the

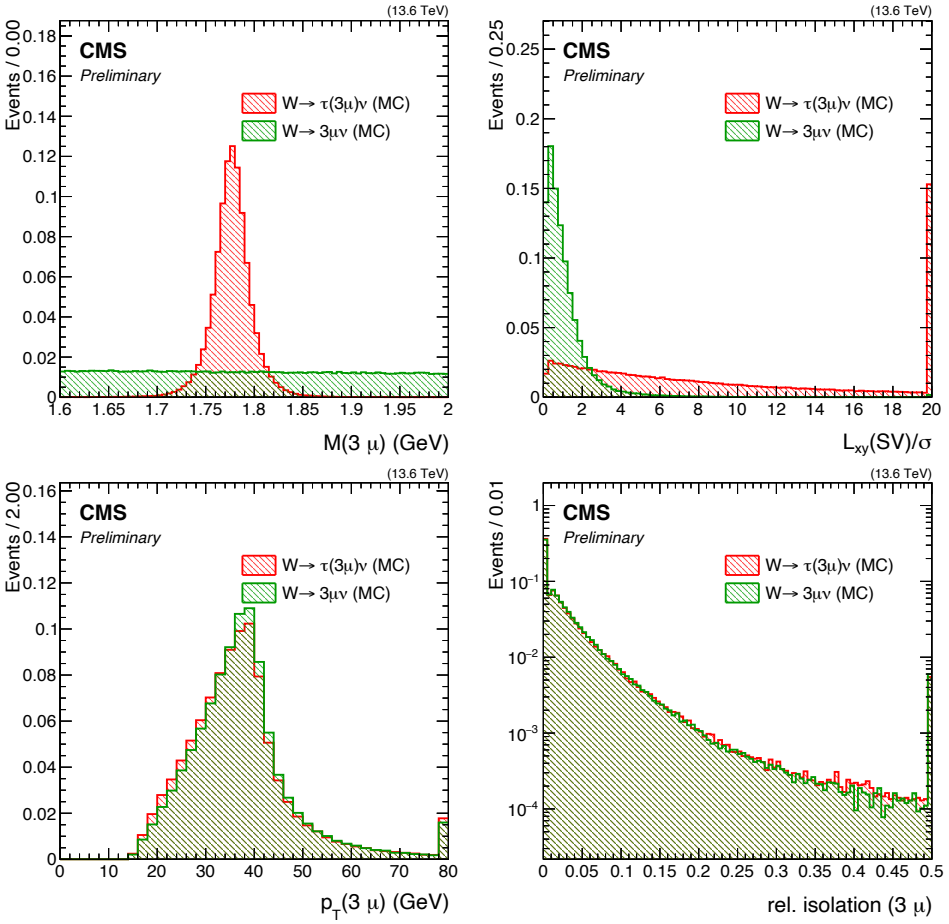


Figure 4.15 Comparison of the $W \to \tau(3\mu)\nu$ and $W \to 3\mu\nu$ kinematics in simulation at preselection level: 3μ invariant mass (top-left), SV displacement significance (top-right), 3μ transverse momentum (bottom-right) and relative isolation, defined as in Eq. 3.3 (bottom-left). First and last bins contain the underflow and overflow events.

errors, hence it was not considered necessary to perform a finer or continuous scan between $L_{xy}/\sigma > 1.4$ and $L_{xy}/\sigma > 3.0$.

After the preselection is applied, the $L_{xy}/\sigma > 2.0$ cut rejects 84% (89%) of the $W \to 3\mu\nu$ events and keeps 82% (79%) of the $W \to \tau(3\mu)\nu$ signal events in 2022 (2023) samples (see Table 5.1). The combined expected UL on $\mathcal{B}r(\tau \to 3\mu)$ improves by 10% with respect its value calculated without any selection on the candidates' displacement.

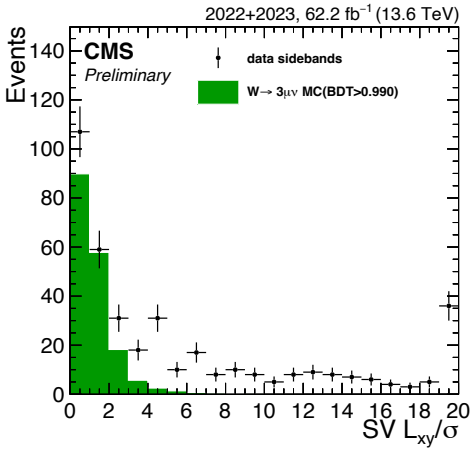


Figure 4.16 Events selected requiring score > 0.990 for a BDT that is trained without cutting on the $\tau \to 3\mu$ SV displacement. Data in the τ mass sidebands (black dots) and the $W \to 3\mu\nu$ simulated sample (green) are compatible in the $L_{xy}/\sigma < 2$ region.

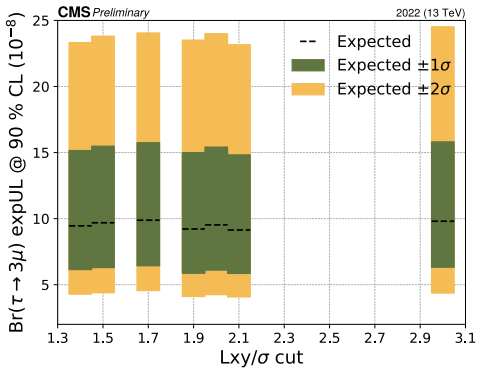


Figure 4.17 Expected upper limit on $Br(\tau \to 3\mu)$ for various cuts on the significance of the $\tau \to 3\mu$ secondary vertex displacement.

4.3 Dimuon resonances contamination

The contamination from ω, ρ, and ϕ mesons decaying into two muons is investigated by examining the invariant mass of the oppositely charged muon pairs within each $\tau \to 3\mu$ candidate, as shown in Fig. 4.18. The individual resonance contributions are evaluated from a fit to data in the mass sidebands, using a Gaussian function to model the signal and a second-order polynomial for the background. From the fit to the opposite-sign dimuon mass distribution reported in Fig. 4.19, it is observed that the combined ω and ρ signal, peaking around 780 MeV, has negligible significance after applying a medium-tight BDT selection (e.g., requiring a BDT score above 0.980). In this context, the significance is defined as the ratio between the number of signal events and the square root of the number of background events ($S/\sqrt{B}$) within a region of 3σ around the resonance peak obtained from the fit over a wider mass range, where σ denotes the fitted resonance width. In contrast, the significance of the $\phi \to \mu\mu$ resonance remains large, with $S/\sqrt{B} > 20$. Consequently, a dedicated

veto is applied to remove only the $\phi \to \mu\mu$ contribution in the following stages of the analysis. This veto excludes $\tau \to 3\mu$ candidates containing two oppositely charged muons with an invariant mass between 0.980 and 1.058 GeV, corresponding to a 3σ window around the measured resonance peak at 1.019 GeV. The value of σ is taken as the largest measured width among the three dimuon combinations $M(\mu_1, \mu_2)$, $M(\mu_1, \mu_3)$ and $M(\mu_2, \mu_3)$, as shown in Fig. 4.20.

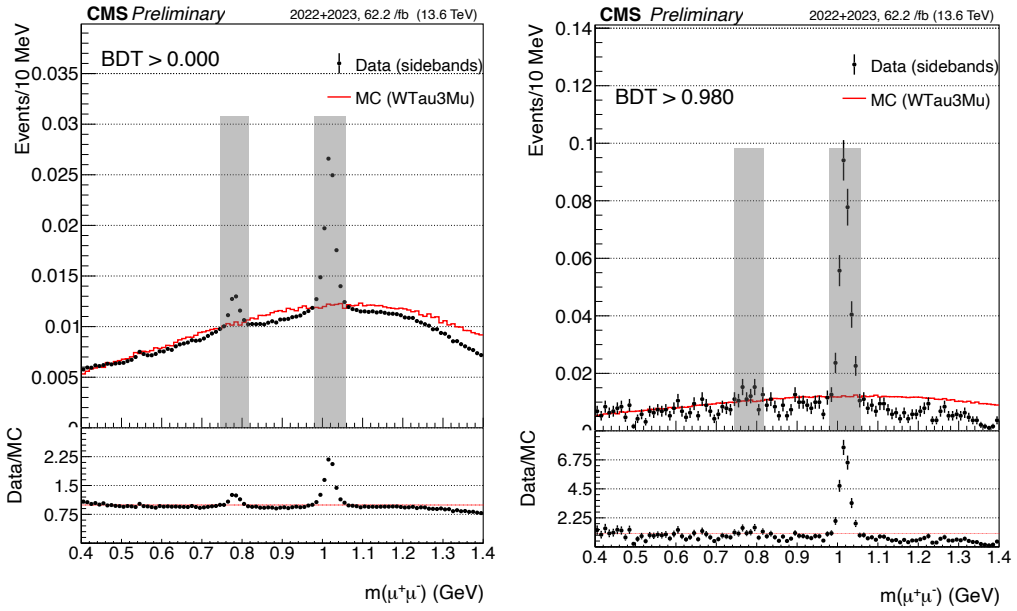


Figure 4.18 Distribution of the mass of two oppositely charged muons within the $\tau \to 3\mu$ candidate for the signal MC (red histogram) and for data in the mass sidebands (black points) applied no selection (left) and medium-tight cut (right) on the BDT score. The gray boxes indicates the 3σ region around the resonance center, as obtained fitting the data with a Gaussian for the resonance signal and a polynomial function for the combinatorics.

4.4 Peaking backgrounds from heavy flavor

The multivariate selection described in the previous section targets the combinatorial background which can be safely modeled from the data sidebands. The data-driven modeling does not cover potentially resonant background sources. In this section the contribution of peaking backgrounds is evaluated with D_s decays, because D_s mass is 1.968 GeV, close enough to the τ lepton [21]. D_s decays where one or more light hadrons are mis-reconstructed as muons represent a potential source of resonant background, since the wrong mass assignment shifts the reconstructed meson mass towards lower values, possibly in the $\tau \to 3\mu$ signal region. For these reasons this background cannot be modeled from the data sidebands.

The expected contamination from $D_s^{\pm} \to \phi(\mu\mu)\pi^{\pm}$ events, where one pion is misidentified as a muon, is estimated in data and used as a benchmark to evaluate the expected contamination from other D_s decays, according to the ratios between the

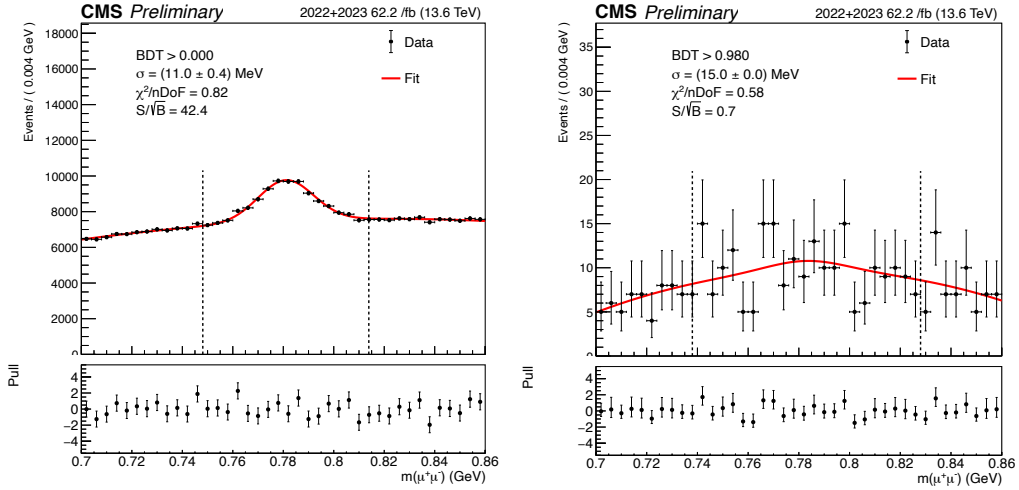


Figure 4.19 Fit of two oppositely-charged muons belonging to the $\tau \to 3\mu$ candidates, within the $\omega \to \mu\mu$ mass window. No selection and a medium-tight cut on the BDT score is applied in the left and right plot, respectively.

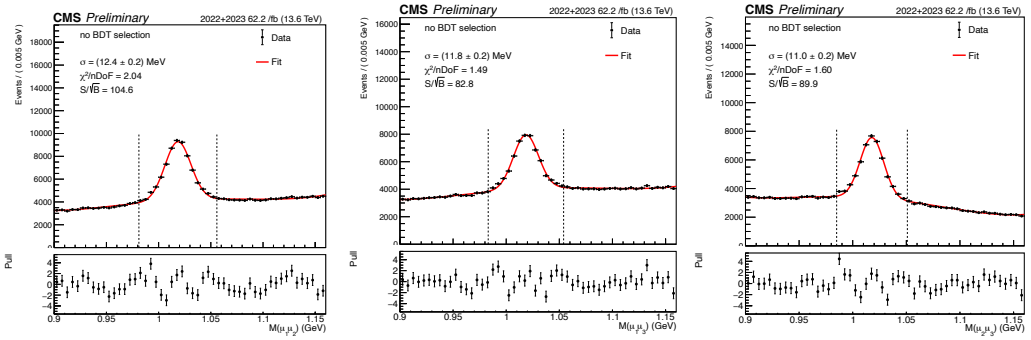


Figure 4.20 Fit of two oppositely-charged muons belonging to the $\tau \to 3\mu$ candidates, within the $\phi \to \mu\mu$ mass window with no selection on the BDT score. From left to right the invariant mass combining the leading and sub-leading, leading and trailing and sub-leading and trailing muons are represented.

branching fraction.

The analysis preselection, described in Sec. 3.4, is applied to reconstruct $D_s \to \phi\pi$ events in data and in simulated samples. Within the muon triplet the two oppositely charged muons whose refitted combined mass is closest to the nominal $\phi(1020)$ mass are assigned to the $\phi \to \mu\mu$ decay and the third muon is identified as the mis-reconstructed pion. To improve the D_s mass peak resolution, and to correct its position, the pion mass is assigned to the latter muon. In order to enhance the D_s resonance in data, the preselection is enforced adding the following requirements:

- $m_{\mu\mu} \in [1.980, 1.058]$ GeV,
- $L_{xy}/\sigma > 5$,

- SV probability $> 10\%$.

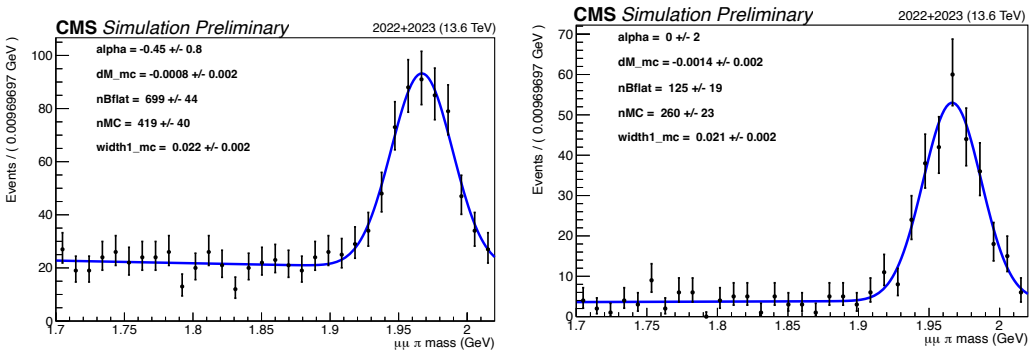


Figure 4.21 Fit to the three-muon invariant mass distribution after the preselection only (left) and with the additional $D_s^+ \to \phi(1020)(\mu^-\mu^+)\pi^+$ selection discussed in the text (right), in simulated $D_s^+ \to \phi(1020)(\mu^-\mu^+)\pi^+$ events. The efficiency of the additional selection is extracted from the ratio of the two yields.

The preselection reinforcement efficiency is measured on the $D_s^\pm \to \phi(\mu\mu)\pi^\pm$ MC samples, resulting in 62.0±2.4% as obtained comparing the D_s yield with and without the preselection reinforcement, shown in Fig. 4.21. The $D_s \to \phi\pi$ yield in data is 490, measured fitting the $\mu\mu\pi$ mass spectrum after the preselection-reinforcement, as represented in the left plot of Fig. 4.22. This yield is then divided by the reinforcement efficiency, obtaining that the expected number of $D_s \to \phi\pi$ events at the preselection level is 790. The contribution from other D_s decay is then obtained scaling the $D_s \to \phi\pi$ yield by the relative branching ratio and the correct power of the $K \to \mu$ and $\pi \to \mu$ misidentification rate, depending on the number of misidentified hadrons in the channel. A conservative value of 5×10^{-3} is used for both fake rates, as estimated in a different CMS analysis [87]. The expected contribution the considered D_s channels is summarized in Tab. 4.4 and the greatest contamination is expected from $D_s^\pm \to \phi(\mu\mu)\pi^\pm$ events. Applying the requirement that the BDT score > 0.990 only two events compatible with having two muons from a $\phi \to \mu\mu$ decay, are observed in data around the nominal D_s mass. These are highlighted in red in the left plot of Fig. 4.22. Consequently, it is possible to conclude that the BDT selection already removes more than 99% of $D_s^\pm \to \phi(\mu\mu)\pi^\pm$ events. Assuming the BDT has the same efficiency for all the considered resonant decays, it is concluded that the combination of a tight BDT selection and the exclusion of $\tau \to 3\mu$ candidates with two muons compatible with $\phi \to \mu\mu$ decays, as described in Sec. 4.3, is enough to suppress the peaking background contamination.

4.5 Control channel $D_s \to \phi\pi$

The selection requirements outlined so far fully define the analysis strategy for isolating potential $\tau \to 3\mu$ signal events. To ensure robustness of the strategy, particularly of the BDT selection, it is necessary to verify that no bias is introduced by the selection itself and, eventually, to account for possible sources of mismodeling. For this purpose, $D_s^\pm \to \phi(\mu\mu)\pi^\pm$ events are used: D_s candidates are reconstructed

Table 4.4 For each three-body final state decay channel of the D_s it is reported its branching ratio scaled by the misidentification probability and the expected yield in data after the signal preselection, extrapolated from the observed $D_s \to \phi\pi$ yield in data.

Channel	$\mathcal{B}r\times$ mis-id rate ($\times 10^{-8}$)	exp. yield
$D_s^+ \to \phi(1020)(\mu^-\mu^+)\pi^+$	6.5	790 (obs. in data)
$D_s^+ \to K^+K^-\pi^+$	0.67	82
$D_s^+ \to 3\pi$	0.14	17
$D_s^+ \to \omega(783)(\mu^-\mu^+)\pi^\pm$	0.071	9
$D_s^+ \to \eta(\mu^-\mu^+)\pi^\pm$	0.053	6

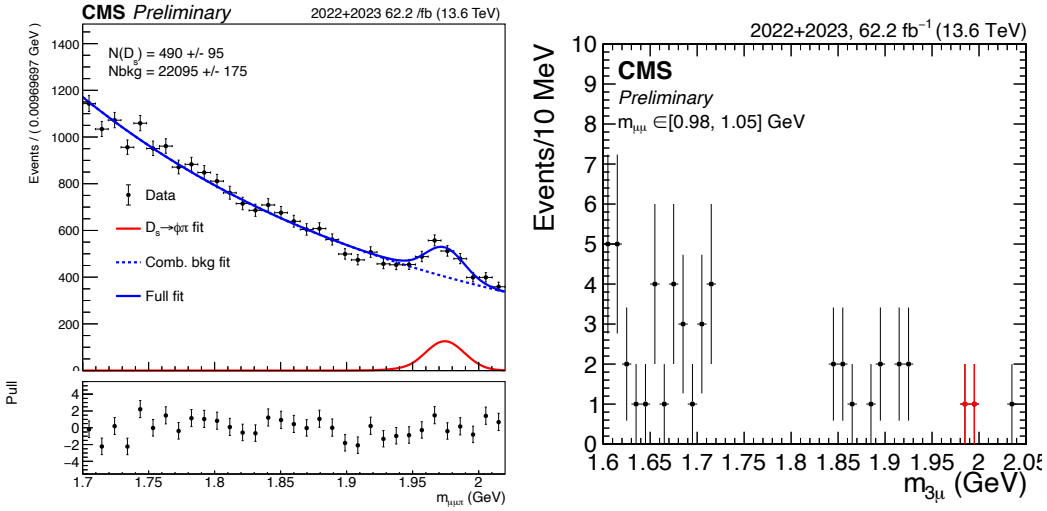


Figure 4.22 Right: fit to the three muons invariant mass distribution in data after the signal preselection and the additional $D_s^\pm \to \phi(\mu\mu)\pi^\pm$ selection discussed in the text. Left: the same events requiring BDT score greater than 0.990, the events in the D_s mass region are highlighted in red.

in both data and MC simulated samples, allowing for direct comparison and identification of potential sources of mismodeling. This process is therefore referred to as a control channel. It is an ideal choice for this purpose, as it shares the same phase space as the signal, being the nominal D_s mass 1.968 GeV, around 190 MeV larger than the τ lepton mass [21]. Being a hadronic process, it exhibits characteristics more similar to the combinatorial background rather than to the signal.

At trigger level, $D_s^\pm \to \phi(\mu\mu)\pi^\pm$ candidates are selected requiring the logical `OR` of the analysis HLT triggers, described in Sec. 3.3. The D_s candidates are built offline combining two reconstructed muons and one track. If the di-muon trigger is fired, the two muons are required to match the trigger objects, otherwise the two muons are required to match the trigger muons of the triple-muon path and the tracks are matched to the tracker muon, which is therefore misidentified.
To guarantee the same phase space region as for the signal, the signal preselection described in Sec. 3.4 is applied. Muons must satisfy the Medium-ID criteria and

have $p_T > 3.5$ GeV (2.0 GeV) for absolute pseudorapidity $|\eta| < 1.2$ ($1.2 < |\eta| < 2.4$). The two muons are associated with a track with $p_T > 1.0$ GeV and $|\eta| < 3.0$. The two muons and the additional track are required to have total electric charge +1 or -1 and invariant mass between 1.0 GeV and 3.0 GeV.
The $\phi \to \mu\mu$ decay in D_s candidates is built fitting the two muon tracks to a common vertex and requiring the fitted invariant mass to be between 0.980 GeV and 1.058 GeV, i.e. compatible with the $\phi(1020)$ mass within 3σ, as measured in Sec. 4.3. If these requirements are satisfied the three objects are fit together to a common secondary vertex and, in order to enhance the D_s resonance in data, the secondary vertex is requested to have a fit probability larger than 0.01%. Finally, all the corrections described in Section 3.6 are applied to the D_s candidates as well, except for the NLO re-weight which is specific of the vector boson signals.

4.5.1 BDT validation

$D_s^{\pm} \to \phi(\mu\mu)\pi^{\pm}$ events are used as a pure background sample to test the agreement between data and MC for the BDT output.
In order to evaluate the BDT on the control channel the track associated to the pion is treated as the trailing muon of the $\tau \to 3\mu$ candidates and it is associated with a fictitious muon tight-ID, which is satisfied the 83% of the times, corresponding to the fraction of trailing muons passing the identification requirements in the signal simulated samples. The BDT input features associated to the W and the MET are defined combining the D_s candidate with the PUPPI MET of the event, treating the D_s as it was a τ lepton candidate.

The comparison between data and MC is performed exploiting the ${}_s\mathcal{P}lots$ method, a statistical technique able to disentangle the contribution of different processes, given the data sample distribution in a given variable [88]. Specifically for this application, the ${}_s\mathcal{P}lots$ are used to isolate the D_s signal using its resonant structure in the distribution of the reconstructed invariant mass $m_{\mu\mu\pi}$. Based on the signal-plus-background likelihood fit performed on data, each D_s candidate is associated with a weight, namely an ${}_s\mathcal{W}eight$, that quantifies the probability of being a signal event. These weights are used to isolate the $D_s^{\pm} \to \phi(\mu\mu)\pi^{\pm}$ signal component in data and thus to fairly compare re-weighted data with the signal simulated sample. The ${}_s\mathcal{P}lots$ tool requires that the variable used to extract the weights, $m_{\mu\mu\pi}$ in this case, and the variables for the data-MC comparison, i.e. the BDT score and the BDT input-features, are uncorrelated, which is demonstrated by the correlation studies in Fig. 4.6.
In the likelihood fit to data, the signal from $D_s^{\pm} \to \phi(\mu\mu)\pi^{\pm}$ events in the $m_{\mu\mu\pi}$ distribution is modeled with a Crystal Ball [89] function while an exponential is used to model the combinatorial background. The Crystal Ball parameters are fixed to the values obtained in the fit to simulated events, except for the width and mean of the Gaussian core which are left free. The fit results for 2022 and 2023 datasets are reported in Figure 4.23 and 4.24, respectively. In data the contribution of $D^{\pm} \to \phi(\mu\mu)\pi^{\pm}$ events is also clearly visible. The $D^{\pm}$ resonance is modeled using a Gaussian whose mean value is fixed to the nominal $D^{\pm}$ mass, 1.869 GeV, all other parameters are free in the fit [21]. Figures 4.25 and 4.26 report the comparison

between data and MC for the BDT output, after the re-weight of data with the ${}_s\mathcal{W}eights$ obtained from the previous fit. Only the D_s component is considered. The two distributions agree well and this is considered a conclusive test for the good validation of the BDT algorithm robustness.

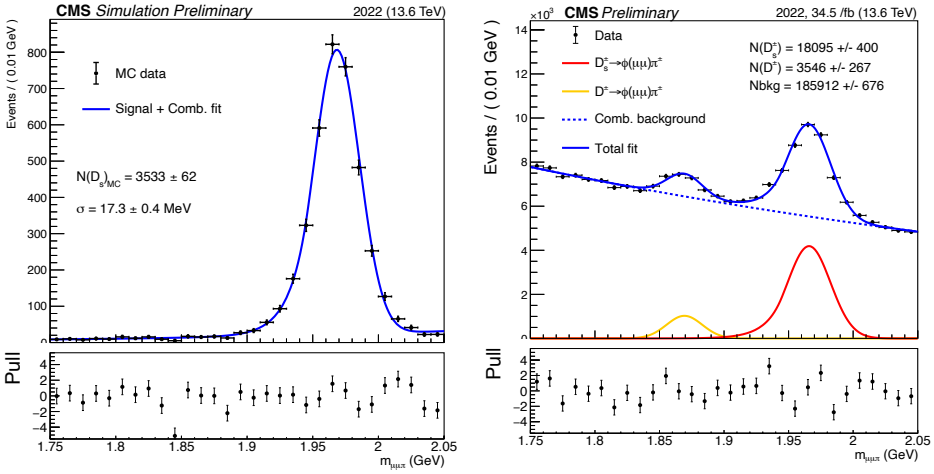


Figure 4.23 Fit to the mass distribution of two muons and one tracks in the control channel $D_s^{\pm} \to \phi(\mu\mu)\pi^{\pm}$ in and MC (left) and in data (right) for 2022 datasets. In data both the D_s (red line) and $D^{\pm}$ (yellow line) peaks are visible over the combinatorial background (blue dashed line), only the former is used for the data and MC comparison showed in Fig. 4.25.

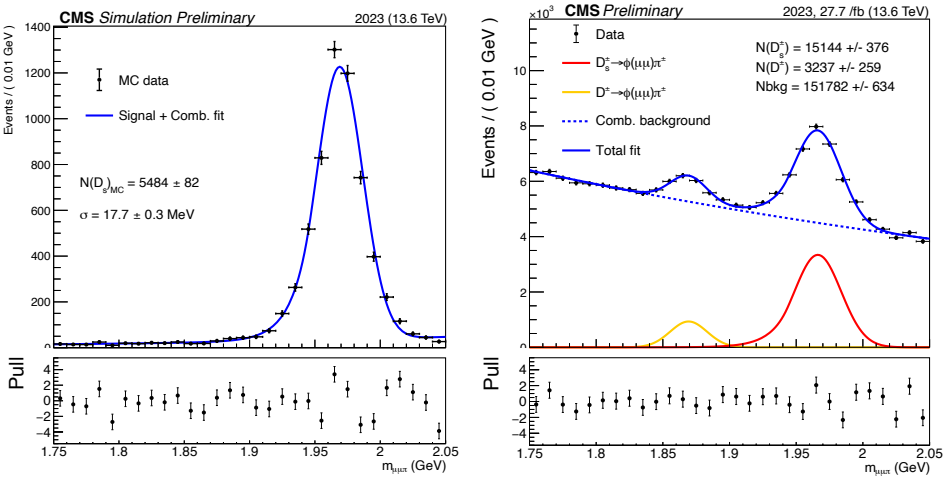


Figure 4.24 Fit to the mass distribution of two muons and one tracks in the control channel $D_s^{\pm} \to \phi(\mu\mu)\pi^{\pm}$ in and MC (left) and in data (right) for 2023 datasets. In data both the D_s (red line) and $D^{\pm}$ (yellow line) peaks are visible over the combinatorial background (blue dashed line), only the former is used for the data and MC comparison showed in Fig. 4.26.

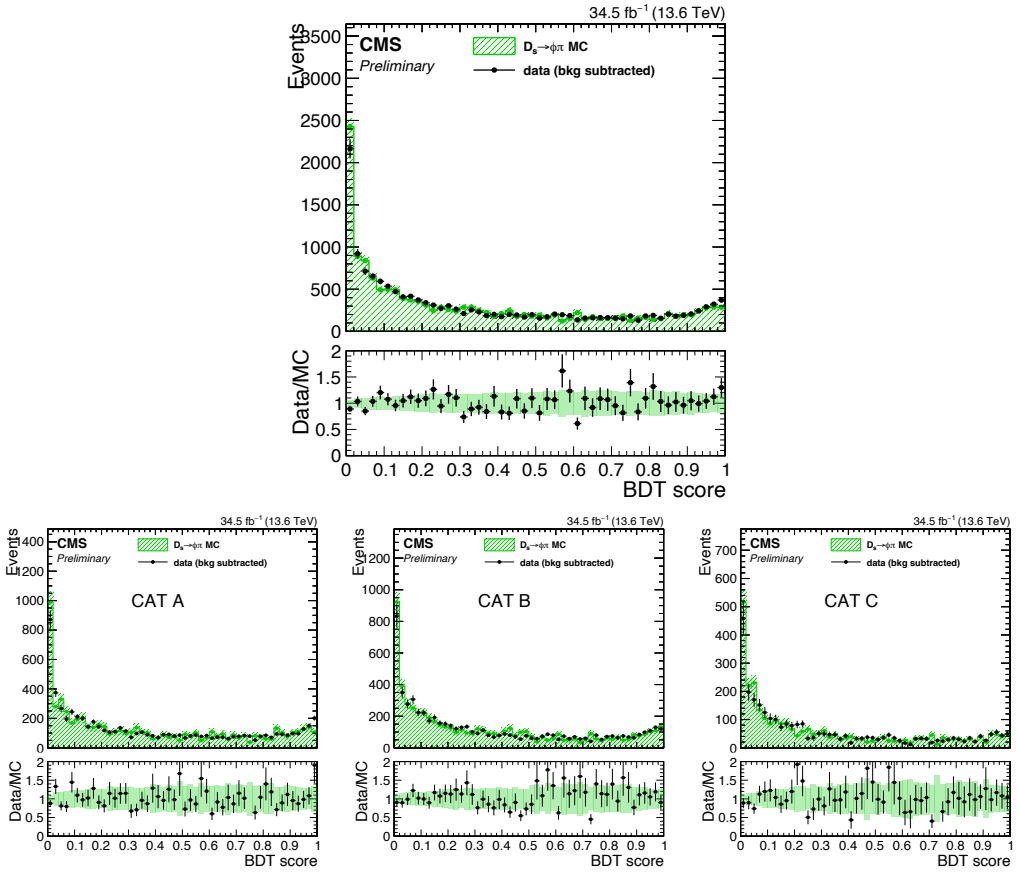


Figure 4.25 Comparison between the BDT output distribution in data with ${}_s\mathcal{W}eights$ applied (black dots) and signal MC for 2022 datasets (green histogram). In the lower panel the data the ratio between data and MC (black points) and the bib-by-bin MC uncertainties are reported. All events are reported inclusively with respect to the event categorization, in the top plot, and category-by-category in the bottom plots.

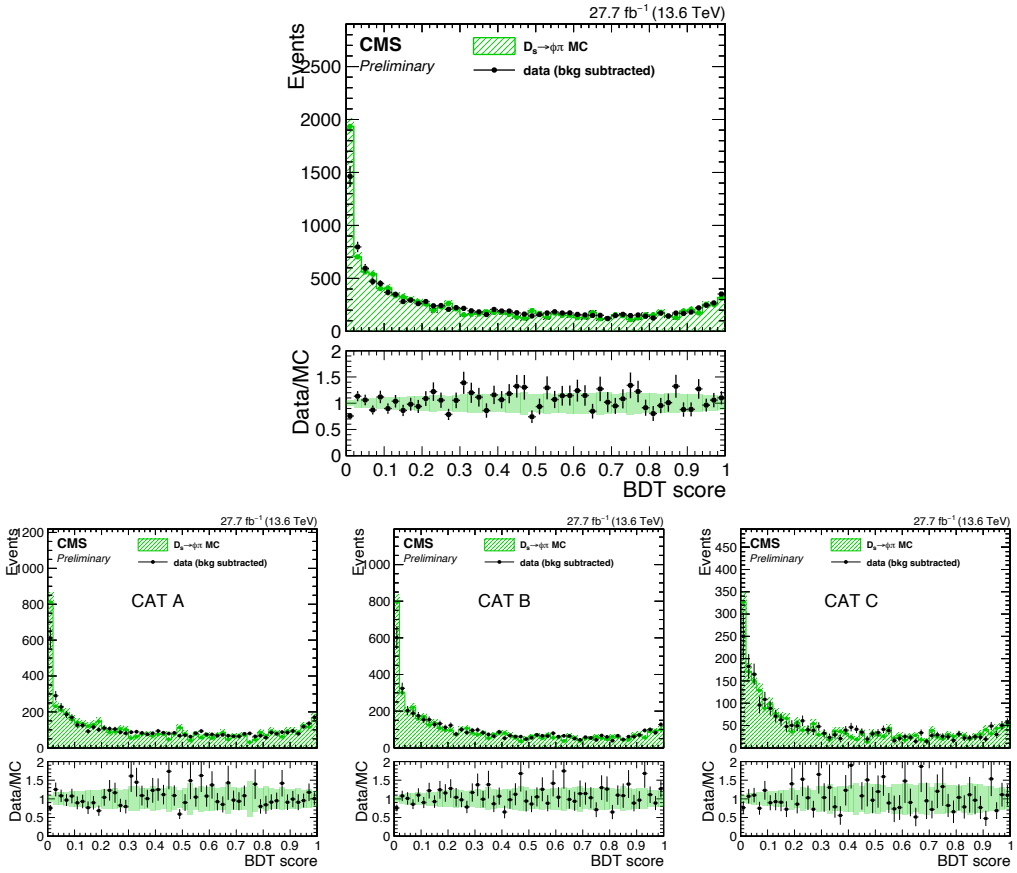


Figure 4.26 Comparison between the BDT output distribution in data with ${}_s\mathcal{W}eights$ applied (black dots) and signal MC for 2023 datasets (green histogram). In the lower panel the data the ratio between data and MC (black points) and the bib-by-bin MC uncertainties are reported. All events are reported inclusively with respect to the event categorization, in the top plot, and category-by-category in the bottom plots.

Chapter 5

Signal and background modeling

The selection criteria outlined in the previous chapter define the overall analysis strategy, which is specifically designed to target $W \to \tau(3\mu)\nu$ events. Nevertheless, LFV in $\tau \to 3\mu$ decays is expected to occur independently of the particular production mechanism of the τ lepton. As demonstrated in Section 5.1, the current selection strategy also exhibits a significant acceptance for $Z \to \tau(3\mu)\tau$ events, even though it was not explicitly optimized for this production mode. Therefore, the expected signal yield from Z boson decays is included for the first time in the search for $\tau \to 3\mu$ decays.
This work adopts a shape analysis approach to extract the signal. The $\tau \to 3\mu$ yield is determined from an unbinned maximum likelihood fit to the reconstructed invariant mass distribution of the three muons ($m_{3\mu}$). The observed spectrum is modeled as the sum of two probability density functions: one describing the expected signal and the other accounting for the background contribution, each normalized to their respective yields, taken from the MC simulations and from a fit to the data-sidebands respectively.
The primary goal of the analysis is to measure the number of potential $\tau \to 3\mu$ signal candidates by fitting the $m_{3\mu}$ distribution in data with this signal-plus-background model. The signal normalization and shape parametrization are derived from simulated $W \to \tau(3\mu)\nu$ and $Z \to \tau(3\mu)\tau$ samples, as detailed in Sec.5.2. The residual background after the BDT selection is modeled using data from the invariant mass sidebands. The background contribution is extrapolated into the signal region, as described in Sec.5.3. Events in the signal region remain excluded from the fit to prevent potential bias in the selection strategy, an approach commonly referred to as a blind analysis. Finally, the sources of systematic uncertainty that may affect the expected signal yield are discussed in Sec.5.4.

5.1 The $Z \to \tau(3\mu)\tau$ production channel

The LFV $\tau \to 3\mu$ decay does not depend on the τ lepton specific production channel, therefore alternative mechanisms can be taken into account to estimate the total sensitivity on $\tau \to 3\mu$ LFV process. The HF production mode is investigated by an orthogonal analysis, therefore the discussion will concentrate on the τ production

via vector boson decays. As illustrated in Sec. 3.1, τ leptons can be also produced via $Z \to \tau\tau$ decays in pp collisions, although with a cross section around 10 times smaller than the one of the W boson. In this section it is shown that the analysis has a significant efficiency also on the expected signal from $Z \to \tau(3\mu)\tau$ process, even though it is optimized on the $W \to \tau(3\mu)\nu$ signatures.

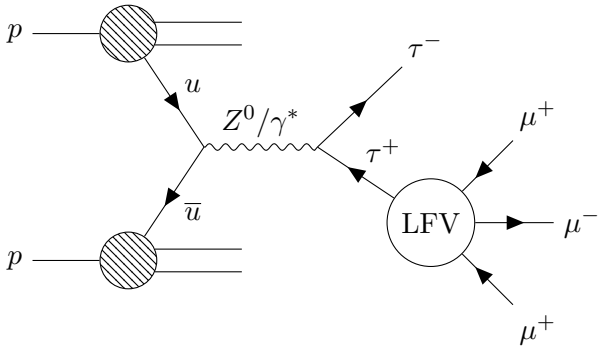


Figure 5.1 Feynman diagram of the $Z \to \tau(3\mu)\tau$ decay.

The same analysis strategy outlined so far is applied to MC samples simulating $Z \to \tau(3\mu)\tau$ events, whose Feynman diagram is shown in Fig. 5.1. The $\tau \to 3\mu$ kinematics are similar in the W and Z production channels, although the muons tend to have a harder momentum spectrum in the latter, as illustrated in Fig. 5.2. The main differences between the two production mechanisms arise from the particle produced in association with the LFV τ lepton, namely, a neutrino escaping detection in the W channel or an additional τ lepton, in the Z channel. The latter is not explicitly considered in the definition of the selection, as the $\tau \to 3\mu$ candidate is combined with the MET of the event regardless of the visible decay products of the accompanying τ lepton in the Z channel. As a result, the same observable and features used for the $W \to \tau(3\mu)\nu$ channel are reconstructed even in absence of a W boson in the event. The same BDT discriminator described in Sec. 4.1.2 is then employed for the final candidates selection. This approach represents the first step toward the inclusion of the $Z \to \tau\tau$ production mechanism in the $\tau \to 3\mu$ search, as it demonstrates that a significant fraction of $Z \to \tau(3\mu)\tau$ events already satisfies the current selection criteria. Therefore, future analyses should aim to optimize and design the selections for $W \to \tau(3\mu)\nu$ and $Z \to \tau(3\mu)\tau$ events simultaneously, in order to maximize both their orthogonality and the overall sensitivity.
The SM decay modes of the additional τ lepton, that is also referred to as the opposite-side τ, can be categorized as either leptonic, yielding an electron or a muon and two neutrinos in the final state ($\tau \to e\nu_e\nu_\tau$ or $\tau \to \mu\nu_\mu\nu_\tau$), or hadronic, producing hadrons and one tauonic neutrino. Figure 5.3 compares several BDT inputs involving the MET between $W \to \tau(3\mu)\nu$ and $Z \to \tau(3\mu)\tau$ simulated samples. The Z boson decays are characterized by lower transverse mass and a higher $p_T(3\mu)$/MET ratio with respect to the $W \to \tau(3\mu)\nu$ events, due to the fraction of p_T carried by the visible decay products of the additional τ lepton. In the Z channel, a large fraction of the $\tau \to 3\mu$ candidates are produced back-to-back with respect to the event MET, as most of the additional τ lepton momentum is carried by neutrinos in the final state. These kinematic features make the $\tau \to 3\mu$ candidates from Z boson decays

very signal like, consequently the same BDT algorithm described in Sec. 4.1.2 is able to effectively discriminate $Z \to \tau(3\mu)\tau$ events from the combinatorial background, as demonstrated by the plot in Fig. 5.4. More quantitatively, the preselection and muon identification requirements on final state muons have about 10% better efficiency for the Z channel compared to the W channel. The only step of the analysis that more specifically targets the W channel signature is the BDT selection, which has 45% (49%) efficiency on W channel and about 33% (37%) efficiency on Z channel using the 2022 (2023) simulated samples. These aspects are summarized in the comparison between the two channels efficiencies proposed in Fig. 5.5 and Tab. 5.1. Focusing on the $Z \to \tau(3\mu)\tau$ channel, the preselection exhibits similar efficiency across all decay modes of the opposite-side τ lepton. In contrast, the BDT selection is more efficient for events in which the additional τ lepton decays leptonically, where the event MET is typically larger and more aligned with the LFV τ lepton momentum, than for hadronic decays (see Fig. 5.6).
In conclusion given the significant acceptance of the selection on $Z \to \tau(3\mu)\tau$ events the expected signal for this process is also considered in the analysis. In addition, the potential LFV signal originating from W bosons produced in $pp \to t\bar{t}$ events, where one top quark decays into a bottom quark and a W boson, is also considered. However, since the τ lepton production cross section in $t\bar{t}$ events is expected to be roughly twenty times smaller than that from $Z \to \tau\tau$ decays, this contribution is considered negligible.

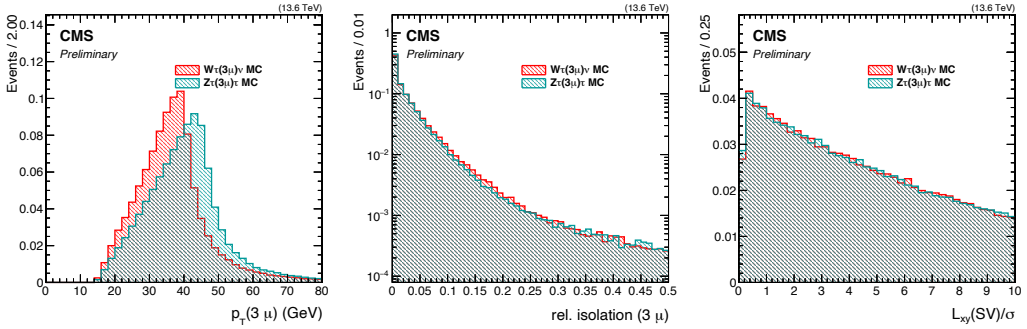


Figure 5.2 Comparison of some discriminating variables between the $W \to \tau(3\mu)\nu$ (red) and $Z \to \tau(3\mu)\tau$ (green) MC samples. Quantities not involving MET are shown.

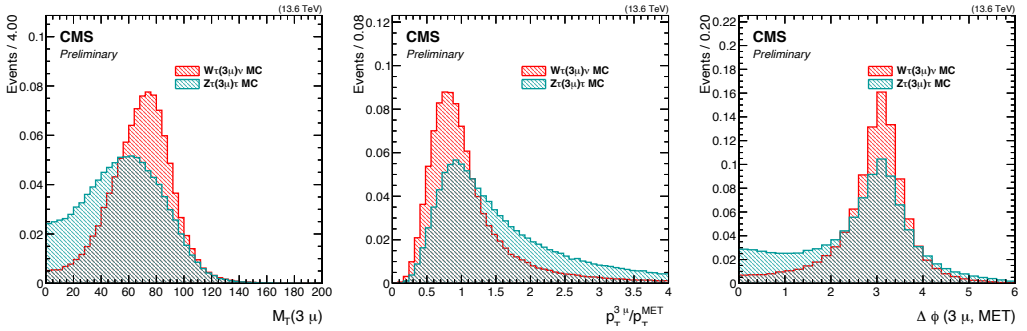


Figure 5.3 Comparison of some discriminating variables between the $W \to \tau(3\mu)\nu$ (red) and $Z \to \tau(3\mu)\tau$ (green) MC samples. Quantities involving MET are shown.

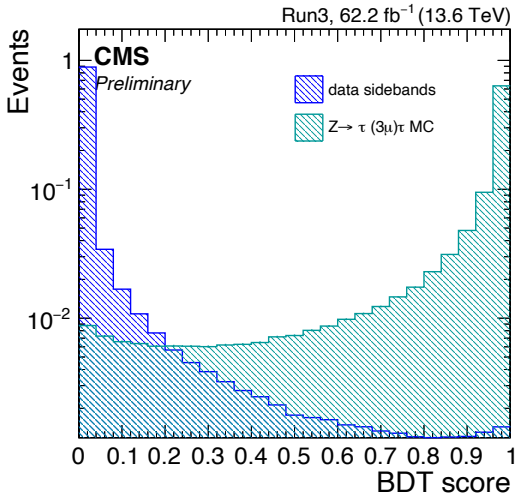


Figure 5.4 Comparison of the BDT score for $Z \to \tau(3\mu)\tau$ (green) and background (blue) events inclusively for the 2022 and 2023.

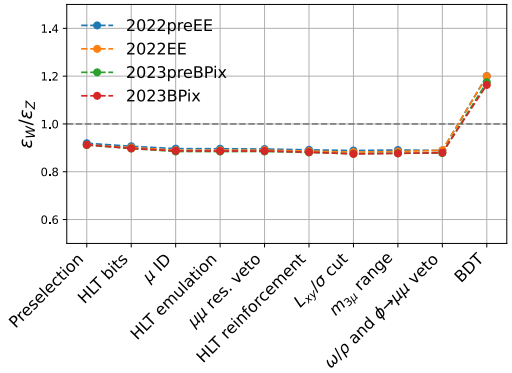


Figure 5.5 Ratio of the selection efficiency on $W \to \tau(3\mu)\nu$ (ε_W) and $Z \to \tau(3\mu)\tau$ (ε_Z) MC events, as a function of the selection step.

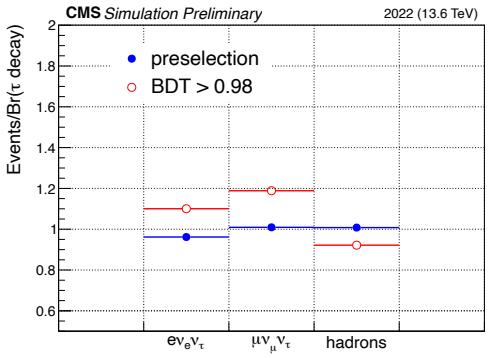


Figure 5.6 Ratio between the number of selected $Z \to \tau(3\mu)\tau$ events and the expected branching ratio for the different decay modes of the second τ in the event. The efficiencies are computed after the preselection and the full selection.

5.2 Signal model

The expected signal from $W \to \tau(3\mu)\nu$ and $Z \to \tau(3\mu)\tau$ events is modeled from their MC simulations, properly normalized to the luminosity in data. The muon triplet mass distribution is fit with a CrystalBall p.d.f. within the mass range $1.40 < m_{3\mu} < 2.05$ GeV, corresponding to the region selected at the HLT level by the tri-muon trigger.

5.2.1 Signal normalization

The $\tau \to 3\mu$ signal yields in the MC samples are normalized to the expected number of $\tau \to 3\mu$ events in data for the W and Z production modes, and thei are given by

$$N_{\tau 3\mu(W)} = \varepsilon_W \cdot \mathcal{L} \cdot \sigma(pp \to W + \mathrm{X}) \cdot \mathcal{B}r(W \to \mu\nu) \cdot \frac{\mathcal{B}r(W \to \tau\nu)}{\mathcal{B}r(W \to \mu\nu)} \cdot r \cdot \mathcal{B}r(\tau \to 3\mu), \quad (5.1)$$

and

$$N_{\tau 3\mu(Z)} = \varepsilon_Z \cdot \mathcal{L} \cdot \sigma(pp \to Z + \mathrm{X}) \cdot \mathcal{B}r(Z \to \mu\mu) \cdot \frac{\mathcal{B}r(Z \to \tau\tau)}{\mathcal{B}r(Z \to \mu\mu)} \cdot r \cdot 2\mathcal{B}r(\tau \to 3\mu), \quad (5.2)$$

where ε_W and ε_Z denote the selection efficiencies, evaluated on $W \to \tau(3\mu)\nu$ and $Z \to \tau(3\mu)\tau$ MC samples respectively. The selection efficiency at each step of the analysis is reported in Tab. 5.1. The integrated luminosity of data is indicated by $\mathcal{L}$. The CMS measurements of the cross sections of vector bosons inclusive production in proton-proton collisions at 13.6 TeV [75] are used. Their values are

$$\sigma(pp \to W + \mathrm{X}) \cdot \mathcal{B}r(W \to \mu\nu) = 20.79 \pm 0.14(\mathrm{sys}) \pm 0.29(\mathrm{lum})^{+0.12}_{-0.13}(\mathrm{acc})\ \mathrm{nb}, \quad (5.3)$$

and

$$\sigma(pp \to Z + \mathrm{X}) \cdot \mathcal{B}r(Z \to \mu\mu) = 2.021 \pm 0.009(\mathrm{sys}) \pm 0.028(\mathrm{lum})^{+0.011}_{-0.013}(\mathrm{acc})\ \mathrm{nb} \quad (5.4)$$

where the total uncertainty is split in the contributions from the systematic uncertainties (sys), the uncertainty on the dataset integrated luminosity and the uncertainty

associated to the acceptance (acc) [75]. Since the measurements are performed exploiting W and Z decays into muons, the cross sections are multiplied by the measured values of the ratios $\mathcal{B}r(W \to \tau\nu)/\mathcal{B}r(W \to \mu\nu)$ and $\mathcal{B}r(Z \to \tau\tau)/\mathcal{B}r(Z \to \mu\mu)$ resulting (1.002 ± 0.020) and (1.0010 ± 0.0026), respectively [21]. The factor 2 in Eq. 5.2 takes into account that both τs can undergo the LFV decay. The $\mathcal{B}r(\tau \to 3\mu)$ branching ratio is arbitrarily assumed to be 10^{-7} and normalized to the signal strength r, which is the analysis parameter of interest (POI).

Table 5.1 Selection efficiency on the $W \to \tau(3\mu)\nu$ and $Z \to \tau(3\mu)\tau$ signals at different stages of the analysis, evaluated on the simulated samples.

Step	**Efficiency [%]**			
	$W \to \tau(3\mu)\nu$		$Z \to \tau(3\mu)\tau$	
	2022	2023	2022	2023
HLT bits and τ candidate	29.0	28.9	32.2	32.2
Medium-ID	24.7	23.9	27.7	27.0
Full preselection	23.5	22.7	26.4	25.8
$L_{xy}/\sigma > 2.0$	19.2	17.9	21.8	20.4
$\phi \to \mu\mu$ veto	16.3	15.1	18.4	17.3
BDT selection	7.4	7.5	6.2	6.4

5.2.2 Shape parametrization

The invariant mass distributions of the three muons ($m_{3\mu}$) in the $W \to \tau(3\mu)\nu$ and $Z \to \tau(3\mu)\tau$ MC are fit with a CrystalBall function. The mean and the width of the fits to the two samples are compared in Fig. 5.7 for each analysis category separately, inclusively with respect to the year of data-taking. The absolute differences between the shape parameters are below 1 MeV in all categories.

To quantify how well the signal shapes in MC describe those in data, the $D_s^{\pm} \to \phi(\mu\mu)\pi^{\pm}$ resonance reconstructed in data and simulation is used. The event selection and fitting strategy are the same as those described in Sec. 4.5.1. The data–MC comparison reported in Fig. 5.8 indicates that the resonance mean is systematically overestimated in simulation, while the width is underestimated. Consequently, in the final fit used to extract the $\tau \to 3\mu$ yield, the signal mean and width are fixed to the values obtained from the fit to simulated samples and corrected by the relative differences observed between data and MC in the D_s signal shape. The Crystal Ball tail parameters are then fixed to their values extracted from the MC sample, in order to limit the number of free parameters in the maximum likelihood fit.

The Data-MC comparison of the $D_s \to \phi\pi$ signal shape parameters provides an estimation of the uncertainty associated with the signal modeling for each category. The magnitude of this uncertainty is used as a quantitative reference to assess the compatibility of the signal shapes between the W and Z channels. The observed difference between the reconstructed D_s mass in data and simulation is of the order of 0.1% (except for category C in 2023), which is larger than the maximum

0.05% difference observed between the reconstructed τ mass in $W \to \tau(3\mu)\nu$ and $Z \to \tau(3\mu)\tau$ MC events. Similarly, the difference between the D_s width measured in data and simulation ranges between 3% and 12%, which is compatible with the relative difference observed between the signal widths in the W and Z channels. In conclusion, the differences between the signal shapes in the two vector boson production channels are smaller than, or comparable to, the uncertainties on the shape parameters derived from the data–MC comparison in the $D_s^{\pm} \to \phi(\mu\mu)\pi^{\pm}$ channel. Therefore, the $W \to \tau(3\mu)\nu$ and $Z \to \tau(3\mu)\tau$ simulated samples are merged, each normalized to its corresponding yield, and fitted with a single Crystal Ball probability density function (p.d.f.), resulting in the signal shape shown in Fig. 5.9.

5.3 Residual background modeling

The background after the BDT cut is modeled directly from data. Analytic functions are fit to the observed three muons mass distribution in the mass-sidebands, i.e. in the disjoint intervals [1.40, 1.74] GeV and [1.82, 2.05] GeV, separately for each analysis category. Choosing an appropriate parametrization of the mass spectrum is important to ensure accuracy of the prediction, while guaranteeing a good analysis sensitivity. Nevertheless, the underlying model which should describe the background is not known apriori therefore the systematic uncertainties associated with the choice of a specific background p.d.f. are included in the final statistical model, described in Sec. 6.1, by letting the maximum likelihood fit to choose different background shapes. The choice of the function is included in the likelihood function as a discrete nuisance parameter indexing the set of shapes provided in input to the fit. In each analysis category the envelope of background p.d.f. is built considering four families of functions: exponential, power-law, Bernstein and Chebychev polynomials [90; 91]. Data sidebands are fit to initialize the p.d.f. parameters and select the most

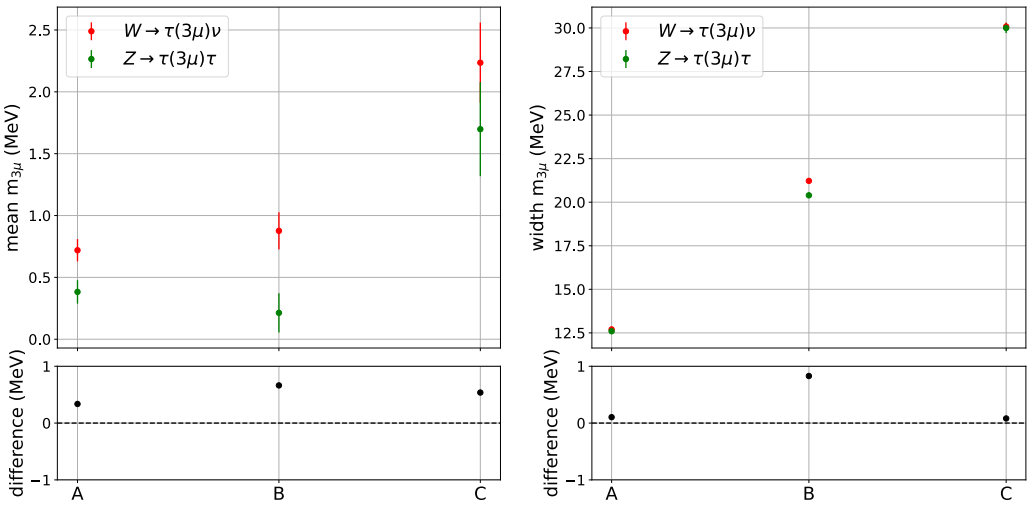


Figure 5.7 Comparison between the signal shape in the $W \to \tau(3\mu)\nu$ (red) and $Z \to \tau(3\mu)\tau$ (green) simulated samples. The mass shift with respect to the nominal τ lepton mass and the signal width are compared in the left and right plot, respectively.

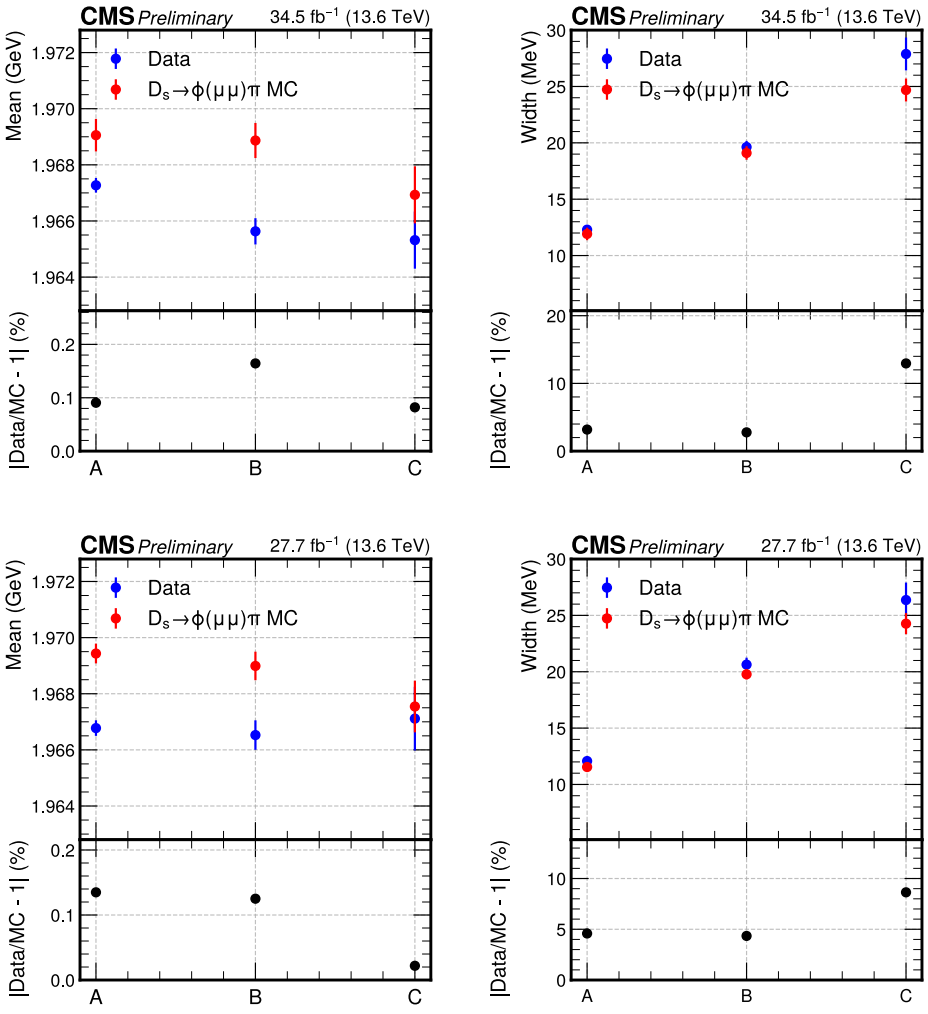


Figure 5.8 Signal shape comparison between data and MC for the $D_s^{\pm} \to \phi(\mu\mu)\pi^{\pm}$ decay in the different categories for 2022 (top) and 2023 (bottom). The mean value and the sigma of the core of the Crystal Ball p.d.f. used to fit the signal component are shown.

convenient set of function. Within the polynomial families the maximum order allowed in the envelope must be limited to control the number of free parameters in the fit, thus avoiding to overfit the data. Therefore, for the Bernstein and Chebychev polynomial families, a given order in that family (starting from the lowest) is first fit to data. Then, the order is increased and the p.d.f. is fit to data as well. To evaluate whether data are better described by the increased order an F-test is performed using the difference between the minimum negative log-likelihood resulting from the two fits, i.e. $2\Delta\text{NLL}_{N+1} = 2 \cdot (\text{NLL}_N - \text{NLL}_{N+1})$, so that the value is positive in case of improvement. As a penalization for the reduction of the degrees of freedom, the value of 2NNL resulting from the fit is increased by a numeric value corresponding to half of the polynomial order. Given the $2\Delta\text{NLL}_{N+1}$ should be distributed as a normalized χ^2, the p-value is calculated as the probability of 2ΔNLL of being

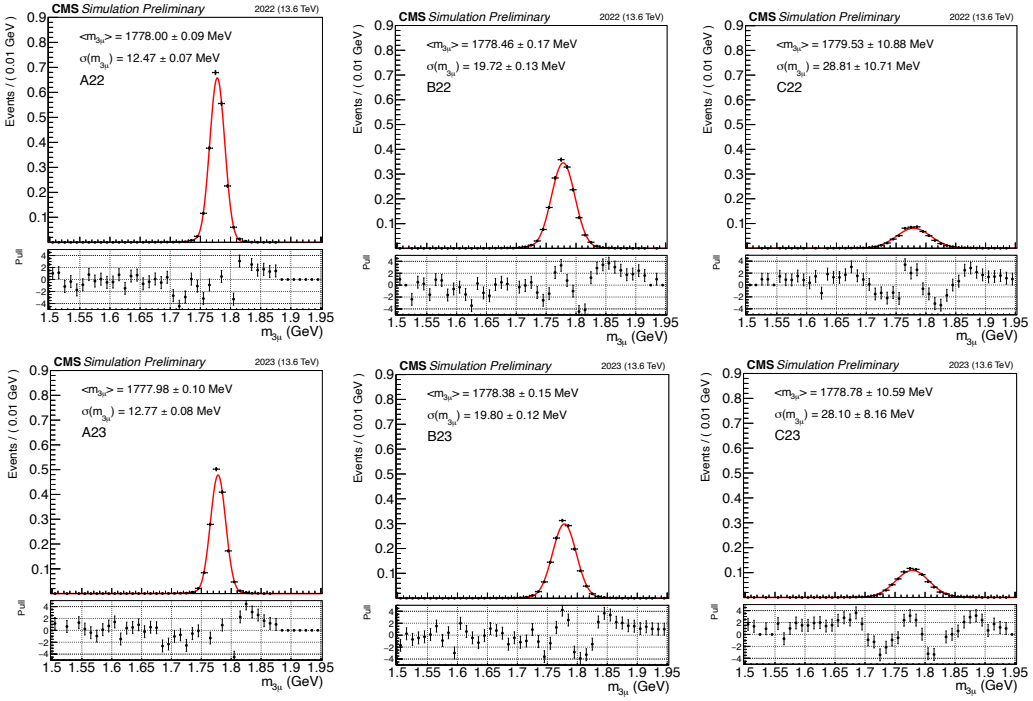


Figure 5.9 Fit of a Crystal Ball function to the $\tau \to 3\mu$ invariant mass distribution in the $W \to \tau(3\mu)\nu$ and $Z \to \tau(3\mu)\tau$ simulated events, merged with the proper normalization and passing the BDT selection. The three analysis categories A, B and C are shown from left to right for (top) 2022 and (bottom) 2023 samples.

greater than the observed value:

$$\text{p-value}(2\Delta\text{NLL}_{N+1}) = P(2\Delta\text{NLL} > 2\Delta\text{NLL}_{N+1}|\tilde{\chi}^2). \tag{5.5}$$

To include the higher order in the envelope the $2\Delta\text{NLL}_{N+1}$ is required to be positive with an associated p-value below 10%.
Before any function is included in the envelope, a goodness of fit test is performed using the p-value associated to the fit reduced-χ^2, evaluated as in the formula below and required to be larger than 1%:

$$\text{p-value}(\tilde{\chi}^2_{\text{obs}}) = P(\tilde{\chi}^2 > \tilde{\chi}^2_{\text{obs}}). \tag{5.6}$$

Given the limited statistics in data-sidebands the p-value is evaluated using pseudo-experiments. The p.d.f. envelopes used in the final likelihood fit for each analysis category are shown in Fig. 5.10 and listed in Tab. 5.2. Given the limited statistics, the F-score criterion selects the lowest order polynomial in each family, except for category B in 2023 where the first order Bernstein polynomial is included. In each envelope the function that better represents the data, according to the p-value of the fit χ^2, will be referred to as the *best-fit*.

The normalization of the background is left completely unbound, so that the fit can freely adjust it at no cost for the likelihood.

Table 5.2 The background p.d.f. included in the envelope for each of the six analysis categories, with the best-fit models highlighted in bold. The p-value associated to the χ^2 of the fit to the mass sidebands (Eq. 5.6) is reported in the third column. In the fourth column the p-value associate to the 2NLL improvement when a polynomial of higher order is included in the envelope.

category	background p.d.f.	p-value ($\tilde{\chi}^2_{\text{obs}}$)	p-value ($2\Delta\text{NLL}_{N+1}$)
A22	exponential	0.40	-
	power-law	0.40	-
	Bernstein 0	0.31	-
	Chebychev 1	0.38	-
B22	exponential	0.37	-
	power-law	0.38	-
	Bernstein 0	0.35	-
	Chebychev 1	0.37	-
C22	exponential	0.63	-
	power-law	0.67	-
	Bernstein 0	0.64	-
	Chebychev 1	0.58	-
A23	exponential	0.35	-
	power-law	0.35	-
	Bernstein 0	0.36	-
	Chebychev 1	0.35	-
B23	**exponential**	0.27	-
	power-law	0.30	-
	Bernstein 1	0.10	0.02
	Chebychev 1	0.16	-
C23	**exponential**	0.45	-
	power-law	0.43	-
	Bernstein 0	0.32	-
	Chebychev 1	0.39	-

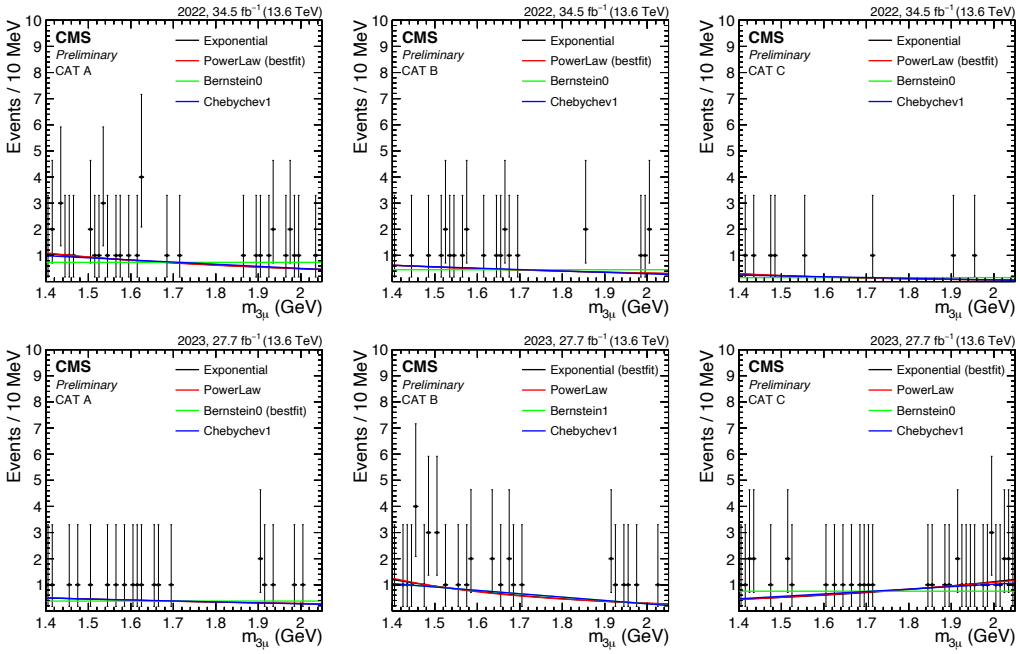


Figure 5.10 Background p.d.f. envelope obtained for all analysis categories applying the BDT selection working points. Plots are blind and data in the mass sidebands are represented by the black dots.

5.4 Systematic uncertainties

Several sources of systematic uncertainties on the expected $\tau \to 3\mu$ signal yield are considered and incorporated as nuisance parameters in the final likelihood fit, described in Sec. 6.3. Tables 5.3 and 5.4 summarize the possible sources of systematic uncertainty and their relative value. Depending on their definition, the uncertainties might be correlated between the analysis categories.
As the background yields are estimated from a fit to data, no additional uncertainties are considered apart from those introduced through the discrete profiling method described in Sec. 5.3.

Table 5.3 Summary of the relative systematic uncertainties on the expected $\tau \to 3\mu$ signal for 2022 dataset. The correlation degree in the last column indicates whether the values are correlated among the years and categories (YC), among the categories of the same year (C) or uncorrelated between the categories.

Source	**Uncertainty (%)**			**correlation**
	A	B	C	
luminosity	1.4	1.4	1.4	C
$\sigma(pp \to V + \mathrm{x})$	1.4	1.4	1.5	YC
$Br(V \to \tau\mathrm{x})/Br(V \to \mu\mathrm{x})$	1.7	1.7	1.7	YC
finite MC statistics	0.5	0.5	0.9	-
HLT trimuon	1.2	2.9	8.4	YC
HLT dimuon	0.4	2.3	3.4	C
muon Medium-ID	1.4	2.0	3.1	C
L_{xy}/σ cut	3.5	3.7	2.7	-
NLO	4.0	3.4	3.3	C
$\mathrm{p}_T(V)$ mismodeling	0.2	0.2	0.3	YC
pileup	1.1	2.7	4.0	C

5.4.1 Signal normalization

The simulated samples for $W \to \tau(3\mu)\nu$ and $Z \to \tau(3\mu)\tau$ are normalized to the integrated luminosity of the data, as described in Sec. 5.2.1.

The uncertainties on the measured production cross sections and branching ratios are propagated as systematic uncertainties on the signal normalization. For each observable the total uncertainty on the expected signal yield is computed as the quadrature sum of the uncertainties from the W and Z production channels, each weighted by the corresponding expected event fraction in each category. The value of $\sigma(pp \to W + \mathrm{X}) \cdot \mathcal{B}r(W \to \mu\nu)$ and of $\sigma(pp \to Z + \mathrm{X}) \cdot \mathcal{B}r(Z \to \mu\mu)$ are associated with a total uncertainty of 1.7% and 1.6%, respectively. The uncertainties on $\mathcal{B}r(W \to \tau\nu)/\mathcal{B}r(W \to \mu\nu)$ and $\mathcal{B}r(Z \to \tau\tau)/\mathcal{B}r(Z \to \mu\mu)$ are 2.0% and 0.26%, respectively.

Table 5.4 Summary of the relative systematic uncertainties on the expected $\tau \to 3\mu$ signal for 2023 dataset. The correlation degree in the last column indicates whether the values are correlated among the years and categories (YC), among the categories of the same year (Y) or uncorrelated between the categories.

Source	**Uncertainty (%)**			**correlation**
	A	B	C	
luminosity	1.3	1.3	1.3	C
$\sigma(pp \to V + \text{x})$	1.4	1.4	1.4	YC
$Br(V \to \tau\text{x})/Br(V \to \mu\text{x})$	1.7	1.7	1.7	YC
finite MC statistics	0.4	0.5	0.7	-
HLT trimuon	1.1	2.7	7.8	YC
HLT dimuon	0.3	1.6	2.3	C
muon Medium ID	0.8	1.5	3.0	C
L_{xy}/σ cut	4.3	4.3	4.5	-
NLO	3.9	3.3	3.1	C
$\text{p}_T(V)$ mismodeling	0.3	0.2	0.3	YC
pileup	1.4	2.1	2.3	C

The cross section and branching ratio uncertainties are correlated between all the categories and years.

The uncertainty on the luminosity is measured by a working group in the CMS Collaboration with a dedicated study for each year of data taking [92; 93]. The relative uncertainty is 1.4% and 1.3% for 2022 and 2023 respectively, therefore it is considered correlated among categories of the same year and uncorrelated among different years.

Finally, the uncertainty due to the finite statistics in the signal MC samples is evaluated as the relative Poisson uncertainty on the number of simulated events, separately for each of the six categories. Thus, it is assumed uncorrelated between all the categories.

5.4.2 Online selection efficiency

The systematics uncertainties associated to the `HLT_DoubleMu_4_3_LowMass` efficiency are obtained shifting the scale factor value described in Sec. 3.6.3 up and down by the value of the total uncertainty on the scale factor itself. The relative impact of such a shift on the expected signal yield is taken as systematic uncertainty. The uncertainties are assumed to be correlated across the categories but uncorrelated among different years.

For the events fired only by `HLT_Tau3Mu_Mu7_Mu1_TkMu1_IsoTau15` * path no scale factor is applied and we assume a 30% systematic uncertainty weighted by the fraction of events which do not pass the `HLT_DoubleMu_4_3_LowMass` requirements; this uncertainty is correlated among all the categories and years. The value is chosen

as a conservative estimation based on the previous CMS analysis, where the total systematic uncertainty is estimated to be below 20% [45]. Given the small fraction of events triggered by these paths this uncertainty has a negligible impact on the final result. Indeed, the overall fraction of simulated events to which the large systematics is applied is around 10% in all years.

5.4.3 Muon identification algorithm efficiency

In each of the six analysis categories, the systematic uncertainty associated with the muon identification efficiency is computed by shifting the scale factors discussed in Sec. 3.6.2 up and down by 1σ and evaluating the impact on the expected signal yield, the maximum deviation of the signal yield in MC is taken as the systematic uncertainty. The associated nuisances are assumed to be correlated among the categories of the same year and uncorrelated among different years.

5.4.4 Displacement significance selection

Discrepancies between the SV displacement significance distributions in simulation and data may introduce biases in the expected signal yield, as a consequence of the selection on L_{xy}/σ described in Sec. 4.2. To investigate for possible discrepancies, the L_{xy}/σ distribution and the $L_{xy}/\sigma > 2.0$ cut efficiency are compared between data and MC using $D_s^{\pm} \to \phi(\mu\mu)\pi^{\pm}$ events and the ${}_s\mathcal{P}lot$ technique. The D_s candidates are reconstructed as detailed in Sec. 4.5, removing any requirement on the secondary vertex displacement. The data-MC comparison for the displacement significance distribution is reported in Fig. 5.11 and the efficiency for the $L_{xy}/\sigma > 2.0$ requirement in the two samples is summarized in Fig. 5.12. To account for the small discrepancy in the efficiency, a systematic uncertainty is associated to each category as the ratio between the cut efficiency in data and in MC, i.e. the values in the lower pad of the plots of Fig. 5.12.

Since they are derived separately in each analysis category, the systematic uncertainties associated to the displacement significance mismodeling in simulation are assumed uncorrelated between all the categories and years.

5.4.5 Vector bosons kinematics

The systematic uncertainty associated with the NLO re-weighting of the W boson kinematics is computed by shifting the scale factors discussed in Sec. 3.6.4 up and down by 1σ and evaluating the impact on the signal yield in MC. The maximum deviation of the expected signal yield is taken as the systematic uncertainty. The uncertainties are assumed to be correlated among all the categories and uncorrelated among different years.

An additional systematic uncertainty is assessed to take into account possible further mismodeling of the vector bosons kinematics in MC with respect to data due to the limited validity of perturbative QCD assumption. With the re-weight discussed in Sec. 3.6.4, the prediction of the transverse momentum distribution for the W and Z boson is obtained with the NLO accuracy. The bulk of the W and Z events have a transverse momentum smaller than 40 GeV which is a region where the logarithmic

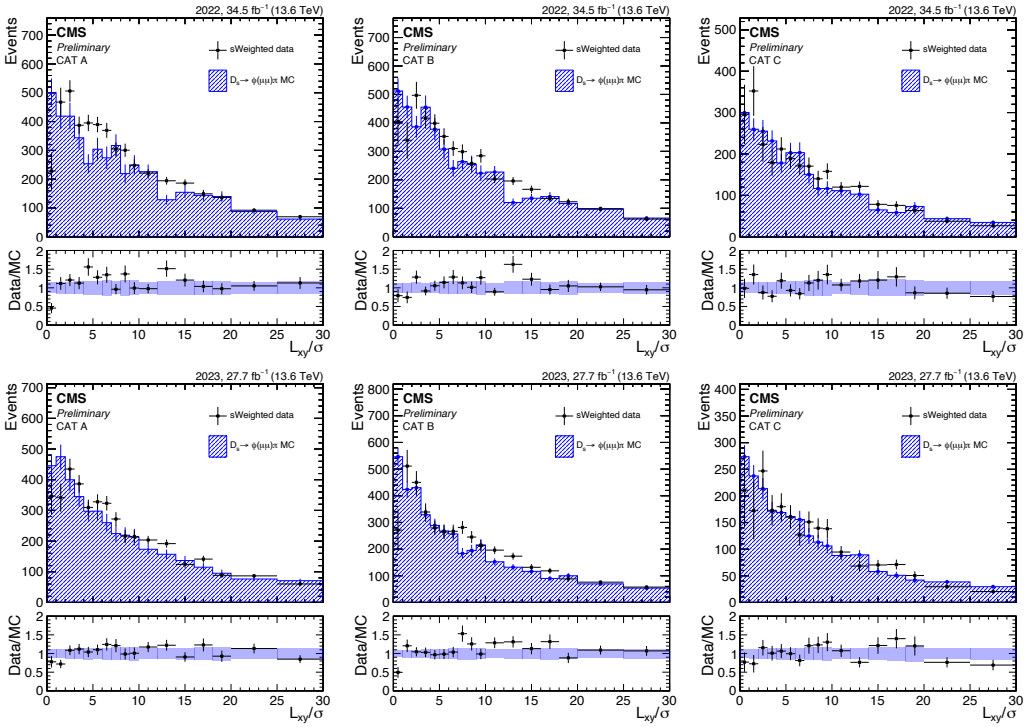


Figure 5.11 Comparison between the $D_s^{\pm} \to \phi(\mu\mu)\pi^{\pm}$ SV displacement significance distribution in *sWeighted* data and signal MC, split in category A (left), B (center) and C (right) and for 2022 (top) and 2023 (bottom) datasets.

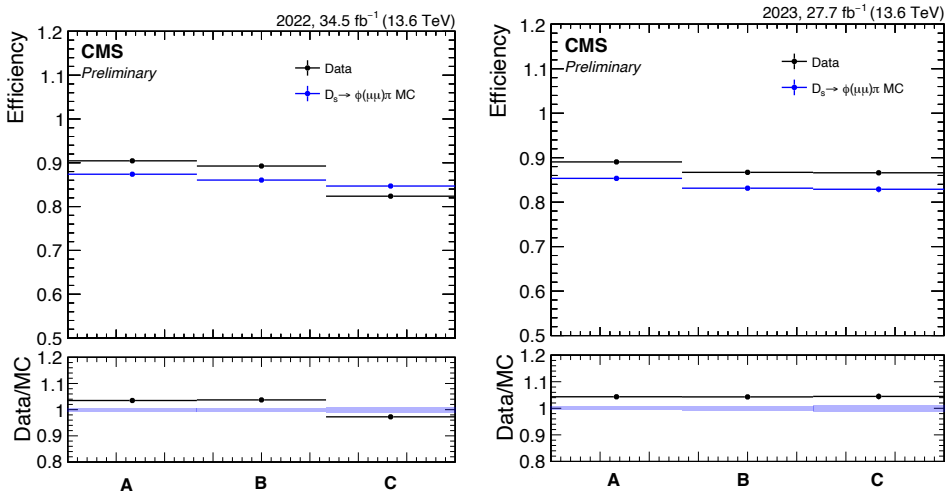


Figure 5.12 Comparison between efficiency of the $L_{xy}/\sigma > 2.0$ cut on ${}_s\mathcal{W}eight$ data and $D_s^{\pm} \to \phi(\mu\mu)\pi^{\pm}$ MC for 2022 (left) and 2023 (right) datasets. The data/MC ratio in the lower panel is included as a systematic uncertainty for the $\tau \to 3\mu$ signal extraction.

terms of the cross section need to be re-summed to get an accurate prediction.
The momentum of the W boson cannot be measured with great precision, given the presence of a neutrino, therefore it is usually assumed that the transverse momentum

distribution of Z and W bosons are similar enough to use Z events to correct both Z and W. In this analysis the vector bosons kinematic is not corrected and only a systematic on the expected signal is assessed based on the correction computed for the study presented in Ref. [75]. The uncertainties are assumed to be correlated among all the categories and years

5.4.6 Pileup corrections

The pileup profile in simulated events is weighted to match the profile in data, as discussed in Sec. 3.6.1, using a minimum bias cross section of 69.2±2.8 mb. The pileup corrections and their associated systematic uncertainties are used and the maximum variation of the expected signal yield due to the up and down shift of the pileup weights is assumed as systematic uncertainty. These uncertainties are assumed to be correlated among the categories of the same year and uncorrelated among different years.

Chapter 6

Results and statistical interpretation

This chapter presents the statistical methods used to determine the exclusion limit on $\mathcal{B}r(\tau \to 3\mu)$, based on the mass distribution of $\tau \to 3\mu$ candidates under the hypothesis that LFV occurs via the decay of the τ lepton into three muons. The expected upper limits are obtained by performing a profile likelihood fit simultaneously across the six analysis categories. Owing to the limited statistics expected in the signal region, the limit extraction relies on the "Hybrid-Bayesian" method to derive the distribution of the test statistic specific to the $\tau \to 3\mu$ search described here. The details of this method are provided in Sec. 6.1. Before evaluating the analysis sensitivity, the fit stability is validated using Asimov datasets with an injected signal, as described in Sec. 6.2. The analysis is currently blind, and data in the signal region are not used, therefore only the expected sensitivity is presented in Sec. 6.3.

6.1 Statistical analysis

The exclusion upper limit (UL) on $\mathcal{B}r(\tau \to 3\mu)$ is computed using the confidence level (CLs) technique [94], based on an unbinned profile likelihood ratio fit performed simultaneously across the six analysis categories, using the COMBINE statistical framework [95].
As described earlier, both the signal and background components are modeled with parametric p.d.f.s. The signal shape is obtained from a fit to the $\tau \to 3\mu$ invariant mass distribution in simulated events, separately for each category. In the fit to data the signal shape parameters are fixed to their fitted values with the corrections detailed in Sec.5.2.2, and the only free parameter is the signal normalization. As explained in Sec.5.2.1, the simulated signal samples are normalized assuming $\mathcal{B}r(\tau \to 3\mu) = 10^{-7}$; therefore, the parameter of interest (POI) in the fit is the signal strength modifier r defined in Eq.5.1. The background parametrization is determined from the data sidebands regions and is not fixed a priori. The likelihood maximization procedure includes the selection of the optimal background model among a set of possible parametrizations, as described in Sec.5.3. The shape

parameters and the background normalizations are free and floating in the likelihood fit.

Fits to the data sidebands using the best-fit background shape from the p.d.f. envelope, overlaid with the expected signal contribution, are shown in Fig. 6.1 for all analysis categories.
The signal region is defined as the $m_{3\mu}$ mass interval between 1.74 GeV and 1.82 GeV, i.e. the region which is currently blind. Such a region contains the 99%, 94% and 84% of the signal in categories A, B and C, respectively, for both years. The expected background yield in the signal region is calculated as the integral of the best background model, and it is reported in Tab. 6.1.

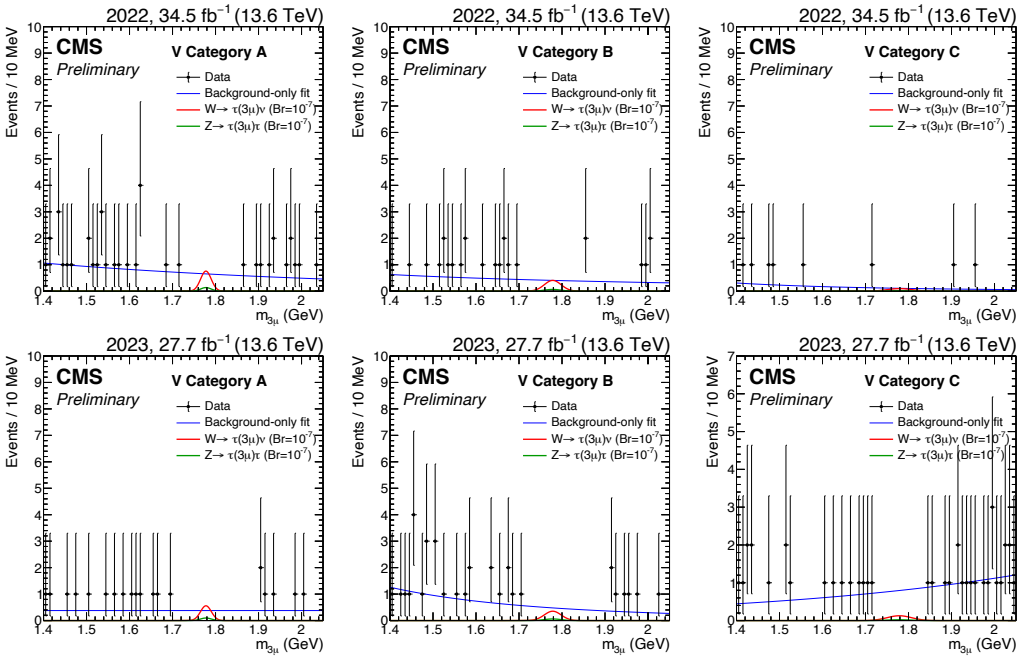


Figure 6.1 Best background model in the multi-pdf envelope, fit to the $\tau \to 3\mu$ data mass sidebands (blue line), and expected signal from $W \to \tau(3\mu)\nu$ (red) and $Z \to \tau(3\mu)\tau$ (green) processes for (top) 2022 and (bottom) 2023 data in the six analysis categories. Data in signal region are not displayed.

6.1.1 Statistical model

The starting point of the profiled-likelihood fit is the definition of the data statistical model, $p(\text{data}; \vec{\Omega})$, generically defined as the probability density of the observed data that depends on a set of parameters $\vec{\Omega}$. In this specific case the observed data are the τ candidate unbinned mass distributions, reconstructed in each category. The parameters of the model consist in the POI, the signal strength modifier r, and the nuisance parameters ($\vec{\nu}$) that are used to model the impact of the systematic uncertainties on the signal sensitivity. Under a computational point of view, it is convenient to factorize the statistical model into a primary component, depending on the $m_{3\mu}$ observable, the POI and the nuisance parameters, and into an auxiliary

Table 6.1 Expected signal and background yields in the signal region in all analysis categories. The expected background contribution in the signal region is extrapolated from the fit to the data sidebands, while the expected signal yield is derived from MC. The data in the signal region are blind, therefore the observed values are still unknown.

Category		Expected yield	
		Signal	Background
2022	A	2.1	5.9 ± 2.4
	B	1.7	3.2 ± 1.8
	C	0.6	1.3 ± 1.2
2023	A	1.5	2.9 ± 1.7
	B	1.5	3.4 ± 1.8
	C	0.8	6.9 ± 2.6

component that only depends on the nuisances, and it is defined as

$$p(m_{3\mu}; r, \vec{y}, \vec{\nu}) = p(m_{3\mu}; r, \vec{\nu}) \cdot \prod_k p_k(y_k; \nu_k), \tag{6.1}$$

where $p(m_{3\mu}; r, \vec{\nu})$ term indicates the probability distribution of the $\tau \to 3\mu$ mass and depends on the signal strength and the nuisance parameters. The $p_k(y_k; \nu_k)$ indicates the probability distribution of the auxiliary observable y_k related to the k-th systematic uncertainty. Considering the six analysis categories as statistically independent, the total likelihood function is expressed as

$$\mathcal{L}(r, \vec{\nu}) = \prod_c p_c(m_{3\mu}; r, \vec{\nu}) \cdot \prod_k p_k(y_k; \nu_k), \tag{6.2}$$

where c runs over the analysis categories and the product over the auxiliary observable takes into account the correlation between the systematic uncertainties across the analysis categories. In this work a parametric shape analysis is employed as the $\tau \to 3\mu$ signal and the combinatorial background model are defined by parametric p.d.f.s, as described Sec. 5.2 and 5.3. Considering only one category for the sake of simplicity, the first term in Eq. 6.2 gets the specific expression as follows:

$$\begin{aligned} p(m_{3\mu}; r, \vec{\nu}) =& \mathcal{P}(n; \lambda_{\mathrm{LFV}} + \lambda_{\mathrm{B}}) \cdot \\ & \frac{\lambda_{\mathrm{LFV}}(r, \vec{\nu}) \cdot f_{\mathrm{LFV}}(m_{3\mu}; r, \vec{\nu}) + \lambda_{\mathrm{B}}(r, \vec{\nu}) \cdot f_{\mathrm{B}}(m_{3\mu}; r, \vec{\nu})}{\lambda_{\mathrm{LFV}}(r, \vec{\nu}) + \lambda_{\mathrm{B}}(r, \vec{\nu})}, \end{aligned} \tag{6.3}$$

where the first term is the Poisson probability for n observed independent events in the mass spectrum. The f_{LFV} and f_{B} term indicate the signal and background p.d.f.s respectively and the λ_{LFV} and λ_{B} terms their expected normalizations, which generally depend on the signal strength and on the nuisances.

The expected yield for the signal or background process can be estimated using external inputs from independent measurements, such as cross-sections or integrated luminosity. In this work a data-driven background modeling is employed, therefore

systematic uncertainties related to external measurements affect only the signal rate estimation. All the considered sources of systematic uncertainty are associated with a log-normal distribution, see Tab. 5.3 and Tab. 5.4, as the signal rate has a proportionality dependence on them. If the expected signal rate λ_{LFV} is proportional to the central value of measured quantity, e.g. the luminosity, the impact of the uncertainty associated to the measured quantity is modeled via a multiplicative factor κ^{ν}, where $1-\kappa$ is the external measurement relative uncertainty and the nuisance ν is normally distributed with mean 0 and unitary width. These definitions correctly model the dependence of the signal sensitivity to the finite precision of an external measurement. The probability density function associated that enters the statistical model is the normal distribution $\mathcal{N}(y;\nu,1)$ and, for a single category, generalizing to a set log-normal systematics, the likelihood is defined as

$$\begin{aligned}\mathcal{L}(r,\vec{\nu}) &= \mathcal{P}(n_{\mathrm{obs}};\lambda_{\mathrm{LFV}}+\lambda_{\mathrm{B}}) \\ &\times \sum_{i=\mathrm{LFV,B}} \frac{\lambda_{\mathrm{i}}(r,\vec{\nu})\cdot f_{\mathrm{i}}(m_{3\mu};r,\vec{\nu})}{\lambda_{\mathrm{LFV}}(r,\vec{\nu})+\lambda_{\mathrm{B}}(r,\vec{\nu})} \\ &\times \prod_{k} \frac{1}{\sqrt{2\pi}} e^{\frac{(\nu_k-y_k)^2}{2}},\end{aligned} \tag{6.4}$$

where k indexes the systematic uncertainties affecting the single category, and the expected signal depends on the nuisance as:

$$\lambda_{\mathrm{LFV}} = r\cdot N_{\tau 3\mu}\cdot\prod_{j}(\kappa_j)^{\nu_j}. \tag{6.5}$$

The final goal is to define the value of the signal strength and the nuisances that maximize the likelihood, given the observed data. The maximum likelihood problem in the multidimensional $(r,\vec{\nu})$ space is formalized using the profile negative-log-likelihood function defined as

$$-2\ln\left(\mathcal{L}(r,\hat{\hat{\vec{\nu}}}(r))\right), \tag{6.6}$$

where the nuisance parameters are profiled so that $\hat{\hat{\vec{\nu}}}(r)$ represents the maximum likelihood estimator for a specific value r. In this way the profiled likelihood depends explicitly on the signal strength only.

For the upper limit extraction the test statistic $\tilde{q}_{\mathrm{LHC}}(r)$ is defined from profiled likelihood ratio as follows:

$$\tilde{q}(r) = \begin{cases} -2\ln\left(\frac{\mathcal{L}(r,\hat{\hat{\vec{\nu}}}(r))}{\mathcal{L}(\hat{r},\hat{\vec{\nu}})}\right) & \text{if } 0\le\hat{r}\le r, \\ -2\ln\left(\frac{\mathcal{L}(r,\hat{\hat{\vec{\nu}}}(r))}{\mathcal{L}(0,\hat{\hat{\vec{\nu}}}(0))}\right) & \text{if } \hat{r}<0, \\ 0 & \text{if } \hat{r}>r, \end{cases} \tag{6.7}$$

where $\hat{r}$ is the maximum likelihood estimate for the signal strength and $\hat{\vec{\nu}}$ indicates the maximum likelihood estimate of $\vec{\nu}$ for $r=\hat{r}$. The test statistic is a real-valued number

defined in the multidimensional space $(r, \vec{\nu})$ and itself associated to a distribution, $f(\tilde{q}; r, \vec{\nu})$, that depends on the likelihood.
In the CLs criterion exclusion limits are calculated from the p-values of the test statistic distribution, calculated under the signal-plus-background (p_r) and under the background-only (p_b) hypotheses as

$$p_r = \int_{q(r)}^{\infty} f(\tilde{q}(r); r) d\tilde{q}, \tag{6.8}$$

$$1 - p_b = \int_{q(r)}^{\infty} f(\tilde{q}(0); r = 0) d\tilde{q}, \tag{6.9}$$

$$CL_s(r) = \frac{p_r}{1 - p_b}. \tag{6.10}$$

The upper limit on the signal strength is derived at (1-α) CLs imposing CL=α. In this work exclusion limits are calculate at 90% CLs (α=0.10), to be consistent with previous results on this topic and make the comparison easier. The above discussion generally defines the approach of the profiled likelihood fit. The specific derivation of the test statistic distribution must be defined though.

6.1.2 Extracting limits with pseudo-data

For sufficiently large data samples, Wilks' theorem can be applied to obtain the analytical distribution of the test statistic $f(\tilde{q}; r, \vec{\nu})$, up to corrections of order $1/\sqrt{N}$ [96; 97].
However, as consequence of the selection strategy adopted in this analysis, all categories feature a small number of expected events in the signal region (see Table 6.1). In this regime, the assumptions underlying Wilks' theorem are not valid, and the exclusion limits cannot be derived from the asymptotic distribution of the test statistic. Instead, the latter is obtained empirically by generating pseudo-datasets for different assumed values of the signal strength r. To this end, the "Hybrid-Bayesian" method is employed. In this approach, toy datasets of the $m_{3\mu}$ distribution are generated from the sampling of the likelihood described in Sec. 6.1.1. Therefore, the toy generation relies on the signal and background parametric models, while the nuisance parameters are randomized according to their probability density functions defined before. It is particularly important that the nuisance parameters are not fixed to the post-fit values obtained from data, since the limited number of events prevents the data from constraining these parameters in a statistically meaningful way. Figure 6.2 shows the test statistic distribution in the signal-plus-background and background-only hypothesis obtained for the combination of the six categories for $r = 0.72$.

The background modeling is also constrained by few events in the data sidebands. Therefore, uncertainties associated with both its parametrization and normalization must be incorporated into the statistical model. Regarding the background normalization, a nuisance parameter is introduced that scales the background rate according to $\lambda_{\mathrm{B}} = r_B \cdot N_B$, in direct analogy to the signal strength modifier r. The initial value N_B, referred to as the pre-fit value, corresponds to the normalization extracted from

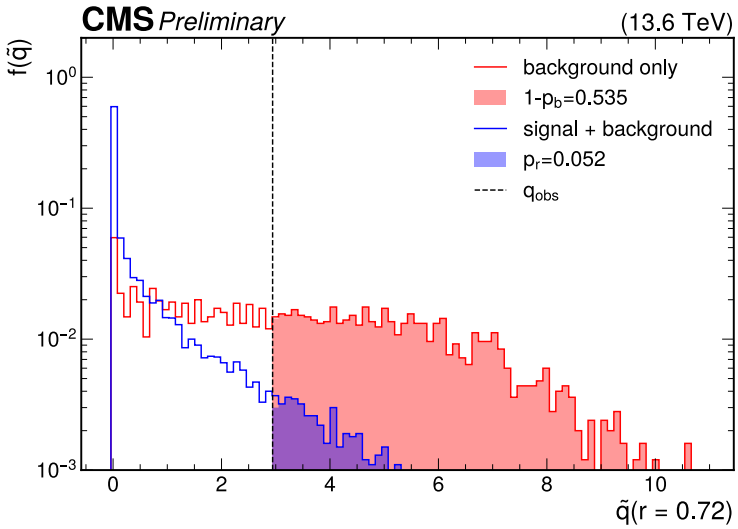


Figure 6.2 Distribution of the test statistic $\tilde{q}(r)$ for $r = 0.72$ using 10k pseudo datasets under the background-only (red histogram) and the signal-plus-background (blue histogram) hypotheses. The black dashed line indicates the observed value of the test statistic and filled red and blue areas represent the p_r and $1 - p_b$ p-values respectively used to calculate the associated CLs.

a fit to the data sidebands using the best-fit background p.d.f. from the envelope described in Sec. 5.3. In the toy generation the background rate is randomized using a flat prior within an interval corresponding to the 99% coverage of a binomial distribution centered on the pre-fit background yield N_B. This interval is computed using the Clopper–Pearson method for binomial confidence intervals [98], where the number of background events expected in the full mass range before and after the BDT selection are interpreted as the number of trials and successes, respectively. The resulting Clopper–Pearson intervals for each category are reported in Table 6.2. This randomization is essential to avoid biasing the pseudo-datasets due to under- or over-fluctuations in the data sidebands, which would otherwise translate into a systematic bias in the expected sensitivity.
As discussed in Sec.5.3, the background modeling is based on a set of candidate p.d.f.s, all evaluated in the likelihood fit to the observed data, where the selection of the most appropriate functional form is part of the minimization problem itself. This is incorporated in the statistical model by introducing the functional choice as a discrete nuisance parameter, which is profiled analogously to continuous nuisance parameters. In the toy generation the discrete index is also randomized generating toys with different background shapes. This approach is commonly referred to as discrete profiling [99].

6.2 Signal injection test

The stability of the profiled likelihood fit is tested using an Asimov dataset generated under the signal-plus-background hypothesis with an injected signal strength of

Table 6.2 For each category: pre-fit values of the background yields (N_B) obtained extrapolating the data sidebands, the 99% binomial confidence interval for the rate modifier r_B and for the total expected rate $\lambda_B = r_B \cdot N_B$.

Category		N_B	99% CL interval r_B	99% CL interval λ_B
2022	A	46.8	[0.66, 1.44]	[31.3, 67.4]
	B	28.9	[0.59, 1.59]	[17.0, 45.9]
	C	9.3	[0.36, 2.20]	[3.3, 20.4]
2023	A	24.5	[0.56, 1.65]	[13.7, 40.4]
	B	39.7	[0.64, 1.49]	[25.4, 59.1]
	C	49.6	[0.67, 1.43]	[33.2, 70.9]

$r = 1.0$. The Asimov dataset is a representative dataset in which the observed data are replaced by their expected values from the total signal-plus-background model.

The scan of the double negative log-likelihood as a function of the POI is shown in Fig. 6.3. As expected, the minimum is centered at the injected value of r, and the 1σ (68%) and 2σ (95%) confidence intervals correspond to the values of r where the curve crosses 1 and 4, respectively. The total uncertainty is decomposed into its statistical and systematic components. The statistical-only contribution is obtained by freezing all nuisance parameters to their post-fit values. The precision on r is limited by the finite size of the dataset, with relative statistical and systematic uncertainties of approximately 50% and 13%, respectively.

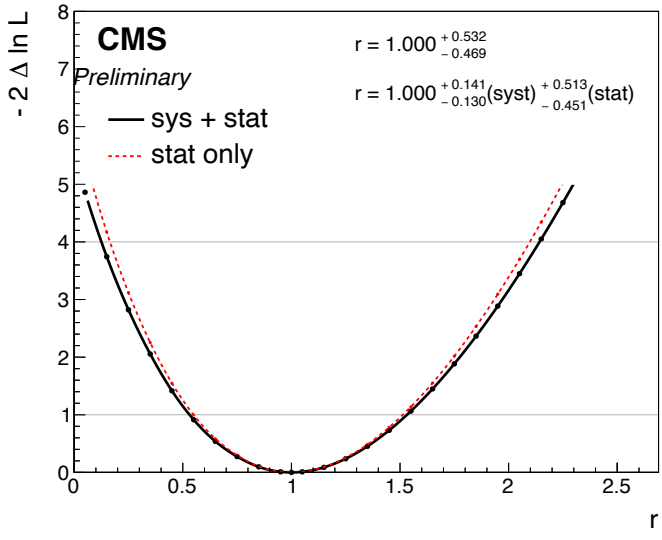


Figure 6.3 Test statistic value for the combination of all the analysis categories, as function of the POI r. The curve is obtained using an Asimov-dataset injecting $\tau \to 3\mu$ signal corresponding to $r = 1$. The horizontal gray lines indicate the test statistic values used to determine the 68%(1σ) and 95%(2σ) CL intervals.

A further quality test of the likelihood fit is the evaluation of the impact of the nuisance parameters. In Fig. 6.4, the pre- and post-fit values and uncertainties of all nuisance parameters are compared. The middle panel shows the pull of each nuisance parameter with respect to its pre-fit uncertainty: the central value of the pull indicates how the global likelihood fit shifts the nuisance from its initial value, normalized to its pre-fit uncertainty. The error bar on the pull represents the ratio between the post- and pre-fit uncertainties, thereby quantifying the degree to which the parameter is constrained by the fit. The right panel shows the corresponding change in the signal strength modifier r when each nuisance parameter is varied by $\pm 1\sigma$.

As expected, the nuisance pulls obtained from the Asimov dataset are centered around zero, and the corresponding impacts on r are symmetric, an indication that no systematic uncertainty introduces a directional bias in the signal sensitivity. It is also observed that the dominant contributions to the uncertainty in r arise from the background rate parameters. These parameters are classified as unconstrained, and thus only their post-fit central values and uncertainties are reported.

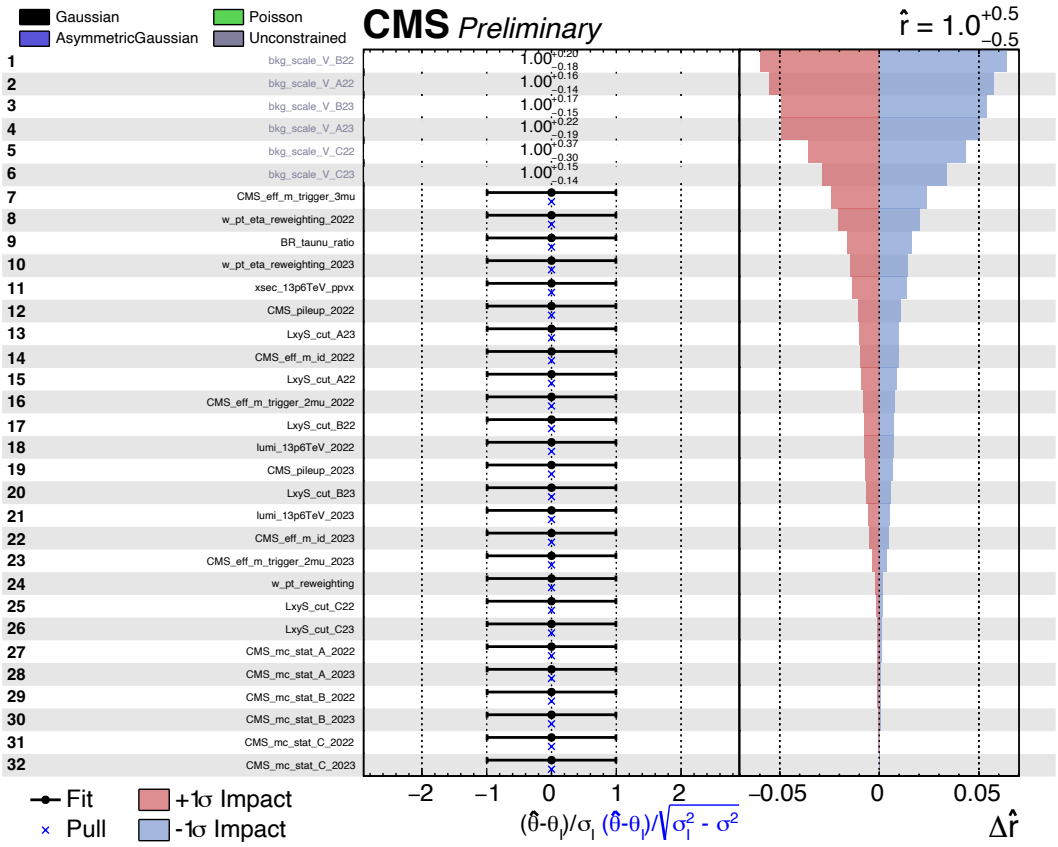


Figure 6.4 Pulls and impacts of the analysis nuisance parameters evaluated on an Asimov dataset generated under the hypothesis $r = 1$. The nuisance parameters are listed in the second column and their pulls and impacts are reported in the third and fourth column, respectively. The fitted value of the signal strength is shown in the upper-right corner.

6.3 Results

The analysis is currently blind, therefore only the expected exclusion limits are derived. The observed data are not used for the maximum likelihood fit and replaced, in each category, by an Asimov sample derived in the background-only hypothesis from the best-fit background shape of the envelope.

The analysis presented in this thesis, performed on the the 2022 and 2023 Run 3 datasets, yields an expected upper limit at 90% CL on $\mathcal{B}r(\tau \to 3\mu)$ of 7.2×10^{-8}, obtained by combining all analysis categories. The corresponding 68% and 95% confidence interval for the expected upper limit computed at 90% CL on $\mathcal{B}r(\tau \to 3\mu)$ are [4.9, 10.3] $\times 10^{-8}$ and [3.5, 14.8] $\times 10^{-8}$, respectively. The expected exclusion limits for each individual category are summarized in Tab. 6.3, while Fig. 6.5 compares the sensitivity achieved in the single data-taking years with that of the combined dataset.

Table 6.3 Expected 90 %CL upper limits on $\mathcal{B}r(\tau \to 3\mu)$, in units of 10^{-8}, computed under the hypothesis of absence of signal as the analysis is blind.

Category		**Expected**				
		-2σ	-1σ	median	+1σ	+2σ
2022	A	8.2	10.5	14.7	21.4	30.9
	B	10.2	12.6	17.9	26.3	37.6
	C	30.8	31.7	40.8	60.0	89.8
	comb.	5.0	6.9	9.6	14.3	20.6
2023	A	10.6	12.3	17.0	24.4	35.5
	B	11.8	15.5	21.3	31.8	46.1
	C	30.9	40.9	58.4	88.4	128.2
	comb.	6.1	8.0	11.5	17.2	24.6
2022+2023	**comb.**	3.5	4.9	7.2	10.3	14.8

The total expected sensitivity obtained in this work is compatible within 7% with the combined 2017 and 2018 expected upper limit of 5.6×10^{-8}, scaled by luminosity [37]. Figure 6.6 puts in comparison this work's results with the expected and observed exclusion limits on $\mathcal{B}r(\tau \to 3\mu)$ obtained in past CMS analyses, as a function of the integrated luminosity. The figure highlights the significant improvement in sensitivity achieved between the 2016 and the 2017-2018 analyses, driven by the development of dedicated trigger streams and the adoption of more sophisticated analysis techniques. The expected signal sensitivity obtained with the 2022 and 2023 dataset scales approximately with the integrated luminosity, despite the extended acceptance for $Z \to \tau(3\mu)\tau$ events. This behavior can be attributed to a more conservative approach to background modeling, including the use of discrete profiling, and may also be associated with efficiency losses due to evolving detector conditions as well as with changes in the background composition resulting from the different trigger strategy employed in the new analysis. Nevertheless, the overall CMS performance remain

strong and competitive.

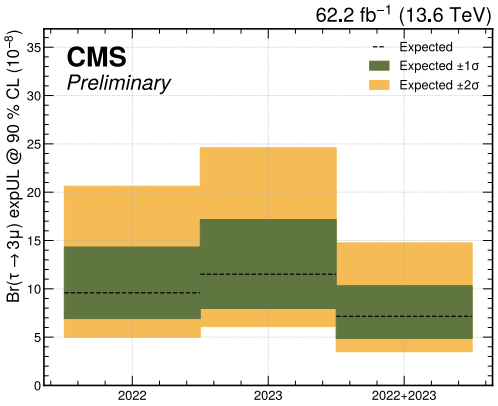


Figure 6.5 Expected upper limits at 90% CL on $\mathcal{B}r(\tau \to 3\mu)$, obtained from the combination of all analysis categories in the 2022 and 2023 datasets separately, and from their combined fit. The associated 68% and 95% uncertainty on the expected exclusion limits is represented by the green and yellow band, respectively.

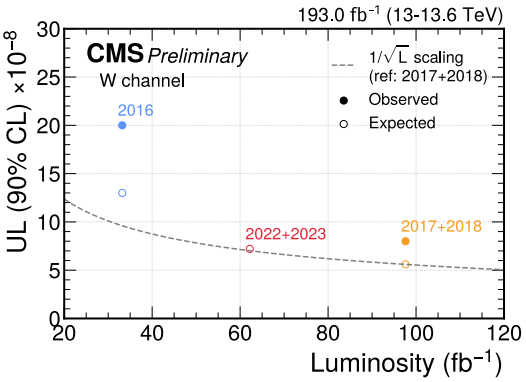


Figure 6.6 Expected and observed -when available- upper limit on $\mathcal{B}r(\tau \to 3\mu)$ at 90% CL as function of the integrated luminosity obtained from the CMS 2016 analysis [100], this work and the 2017+2018 CMS analysis [45], in the W channel. The results are shown in blue, red, and orange, respectively, with empty markers indicating expected limits and filled markers indicating observed limits.

6.3.1 Preliminary combined CMS sensitivity

The CMS experiment physics program for the search for LFV in $\tau \to 3\mu$ decays consists of two orthogonal analyses which focus on τ leptons produced in heavy-flavor decays of D and B mesons, and in $W \to \tau\nu$ decays. The overall CMS sensitivity is obtained by combining the two analyses through a simultaneous unbinned maximum likelihood fit to the muon triplet mass distribution.

At the time of writing this thesis, the HF channel analysis using the 2022 and 2023 datasets is undergoing the internal review within the CMS Collaboration and is being carried out by a different research group. Although the HF analysis results are preliminary and may become more conservative as the review progresses, a preliminary combined expected sensitivity is presented. Both analyses use data collected by the CMS experiment in 2022 and 2023 through the double-muon parking data stream, corresponding to an integrated luminosity of 62.2 fb^{-1}. The expected exclusion limits obtained from a preliminary combination of the HF and vector boson analyses at 90% CL is

$$Br(\tau \to 3\mu)_{\mathrm{HF+W}} < 5.0 \times 10^{-8} \text{ (2022+2023 combined).} \tag{6.11}$$

The combination assumes that systematic uncertainties in the two channels are uncorrelated, as the analyses are fully independent. At the moment of the final combination, signal contamination from the vector boson analysis in the HF signal region will be also taken into account: events selected by both analyses are removed from the HF-channel fit to ensure strict orthogonality.

The final CMS results will include the combination with the previous analyses performed using the full Run 2 dataset, corresponding to an integrated luminosity of 131 fb^{-1} [45]. The Run 2 combination yields a total observed (expected) upper limit at 90% CL on $\mathcal{B}r(\tau \to 3\mu)$ of 2.9 (2.4)$\times 10^{-8}$. A preliminary combination of the Run 2 results with the expected sensitivity from the 2022 and 2023 datasets leads to an expected exclusion limit at 90% CL of

$$Br(\tau \to 3\mu)_{\mathrm{HF+W}} < 2.1 \times 10^{-8} \text{ (Run 2+2022+2023 combined).} \tag{6.12}$$

Although these results are preliminary and may evolve into weaker limits, they already indicates that the sensitivity achieved in this work is competitive with the current experimental scenario. In particular, the CMS experiment sensitivity to LFV in $\tau \to 3\mu$ results in the best one within the LHC experiments and is comparable to the current world-best limit set by Belle II experiment. This result, among others, highlights the strong performance of the CMS experiment in probing regions of phase space beyond its original design goals.

6.3.2 Future prospects

The preliminary expected sensitivity discussed in the previous section directly motivates further investigations of LFV in $\tau \to 3\mu$ decays using future data. The full CMS Run 3 dataset is foreseen to exceed 300 fb^{-1}, opening the possibility of improving the exclusion limit down to the level of $\mathcal{B}r(\tau \to 3\mu) < 1.3\times 10^{-8}$, at 90% CL.

Beyond the increase in available statistics, future CMS searches for $\tau \to 3\mu$ decays could benefit from more sophisticated analysis strategies. For instance, improvements in the treatment of the $W \to 3\mu\nu$ background could help reduce the impact of the current cut-based approach on signal efficiency. In addition, this analysis provides a fertile ground for advanced machine learning approaches, such as graph neural networks, which could naturally encode the $\tau \to 3\mu$ topology and isolation properties, enabling more effective background discrimination.

This work proves the potential of exploring τ lepton produced in Z bosons decays. One promising future direction would be a full rethinking of the analysis strategy to target vector bosons production modes in a more systematic manner. In particular, an event categorization scheme could be introduced to distinguish between τ leptons produced in W and Z boson decays. This could be achieved by analyzing the event content opposite to the $\tau \to 3\mu$ candidate and identifying an isolated muon, electron, or hadronically decaying τ lepton(τ_h), corresponding to $Z \to 3\mu + \tau_\mu$, $+ \tau_e$ or $+ \tau_h$ final states, respectively, as well as by selecting events with isolated MET to select $W \to \tau(3\mu)\nu$ candidates.
The unprecedented size of the Run 3 datasets will be essential to pursue innovative solutions, not only to provide sufficient statistics for training machine-learning models, but also to enable robust optimization of event selections and accurate modeling of category-specific backgrounds.

Looking further into the future, the LHC at CERN will be upgraded to the High-Luminosity LHC by around 2030, delivering an unprecedented instantaneous luminosity. During this phase, the CMS detector will be facing pileup levels of 140-200 collisions per bunch crossing. The expected CMS sensitivity to LFV in $\tau \to 3\mu$ decays is projected to reach an upper limit on $\mathcal{B}r(\tau \to 3\mu)$ as low as $3.9{\times}10^{-9}$ for a target integrated luminosity of 3 ab^{-1} over the HL-LHC period. This corresponds to an improvement in sensitivity by approximately a factor of five with respect to the results presented in this work. Such a projection is included among the latest inputs to the European Strategy for Particle Physics, jointly provided by the ATLAS, Belle II, CMS and LHCb Collaborations [101].
It should be noted that this projection does not yet account for the significant hardware and software upgrades that CMS is undertaking in preparation for the HL-LHC, aimed at preserving detector performance and fully exploiting the vastly increased data volume. Among these upgrades is the muon system, which includes the introduction of further GEM detectors in the forward region. The upgrades in the muon and in the L1 trigger system will extend the geometrical acceptance for muons from the current limit of $|\eta| < 2.4$ to approximately $|\eta| < 2.8$, and will enable the reconstruction of standalone muons at the Level-1 trigger stage up to $|\eta| = 2.4$, compared to the currently limit of $|\eta| = 1.6$ [102; 103]. This will substantially increase the fiducial acceptance for $\tau \to 3\mu$ decays, particularly for τ leptons originating from HF decays, whose decay muons are typically forward and highly collimated. In this context, the LFV decay $\tau \to 3\mu$ represents a flagship analysis of the HL-LHC physics program, serving as a key benchmark for the performance and physics impact of some of the CMS detector upgrades.

Conclusions

The conservation of lepton flavor number is an accidental symmetry of the SM, not supported by any fundamental principle. Evidence of the decay $\tau \to 3\mu$ at current experiments would unambiguously demonstrate the existence of new phenomena, as predicted by several BSM theories.
This thesis presented a search for the lepton flavor violating decay $\tau \to 3\mu$, performed with proton-proton collision data collected by the CMS detector at a center-of-mass energy of 13.6 TeV during 2022 and 2023, and corresponding to an integrated luminosity of 62.2 fb^{-1}.

The analysis presented in this work focuses on τ leptons produced in W boson decays and searches for a localized excess over a smooth background in the invariant mass distribution of three muons originating from a common vertex. The event selection proceeds in several stages. An initial cut-based preselection is applied to define a sample of three-muon candidates, enforcing basic requirements on muon quality, vertex compatibility, and kinematic properties.
Subsequently, a multivariate selection is designed to suppress background events and maximize the sensitivity to potential signal candidates. Two complementary approaches are combined and jointly optimized: a cut-based selection, primarily targeting contamination from the SM $W \to 3\mu\nu$ process, and a ML–based selection, focused on reducing the combinatorial background arising from random associations of unrelated muons. A dedicated BDT is trained using the $W \to \tau(3\mu)\nu$ MC simulation as signal sample, while the combinatorics is modeled directly from data in the muon triplet mass sidebands.
Given the small signal yield expected in the signal region, the BDT working points definition is designed to maximize sensitivity while maintaining robust control of background modeling and statistical fluctuations. An orthogonal, and ten time larger, background-like control sample is obtained by muon triplets having two muons failing the muon identification criteria required for the signal candidates. This control sample is re-weighted through another BDT algorithm to reproduce the kinematic properties of the data sidebands, enabling a reliable optimization of the selection working points and of the signal-versus-background BDT discriminator.

Although the analysis is optimized for τ leptons produced in W boson decays, the selection strategy retains significant acceptance for $\tau \to 3\mu$ decays in which the τ originates from Z boson decays. This thesis presents the first attempt to explicitly include the $Z \to \tau\tau$ production mechanism of τ leptons in the $\tau \to 3\mu$ search, thereby expanding the experimental sensitivity to a previously unexplored

production channel. This establishes the foundation for future analyses focusing on $Z \to \tau\tau$ events.
Finally, a shape-based statistical approach is employed for the signal extraction, where the invariant mass distribution of the three-muon system is modeled using parametric functions for both the signal and background components. The final results are extracted through a simultaneous profile likelihood fit across six analysis categories.
Currently, data in the signal region remain blind, and only the expected exclusion limits are reported. The analysis yields an expected upper limit at 90% confidence level of $\mathcal{B}r(\tau \to 3\mu) < 7.2\times10^{-8}$, combining the 2022 and 2023 datasets. This result is combined with the sensitivity achieved by the full Run 2 analysis, including also the current preliminary expected sensitivity in the HF channel. According to this preliminary combination, the overall CMS sensitivity can reach the current best exclusion limit, set by the Belle II experiment.

The total integrated luminosity expected for the full Run 3 period is anticipated to reach nearly 300 fb^{-1}, more than double the dataset used in this work. This thesis therefore represents the first step towards exploiting the full potential of the Run 3 dataset for the search for $\tau \to 3\mu$ decays in vector boson production channels. It also demonstrates, for the first time, the feasibility of extending the analysis sensitivity to $Z \to \tau(3\mu)\tau$ decays, an aspect that should be further explored with the complete Run 3 statistics.

Appendix A

Performance of the front-end electronics of the CMS electromagnetic calorimeter barrel for the High-Luminosity LHC

The CMS ECAL, presented in Sec. 2.2.3, was designed to achieve an excellent energy resolution up to an integrated luminosity of 500 fb^{-1} over ten years of data taking at a peak luminosity of 1 $\times 10^{34}$ $cm^{-2}s^{-1}$. The performance of the LHC during Run 2 and Run 3 has exceeded the peak luminosity by more than a factor of two, with the ECAL continuing to deliver excellent results. The High-Luminosity (HL) phase of the LHC, also called Phase-2 of the LHC and itroduced in Sec. 2.1, is expected to deliver an integrated luminosity of 3000 fb^{-1} with radiation levels in the detector six times higher than the nominal LHC design. To maintain the current performance in such a difficult environment the ECAL barrel must be upgraded [104]. The ECAL endcap, along with the entire CMS forward calorimetry, will be completely replaced by a silicon-based high-granularity calorimeter [105].

This section presents the results of the performance studies of the upgraded electronics for the ECAL, made with a 5×5 matrix of $PbWO_4$ crystals and high energy electron or pion beams. The data were collected in separate campaigns, in 2018 and 2021, in the H4 and H2 test beams at the CERN Super Proton Syncroton (SPS) [106]. In the following, the current electronics is referred to as the "legacy electronics" and the upgraded version as the "Phase-2 electronics". Section A.1 generally describes the ECAL barrel upgrade for HL-LHC, and details the components of the Phase-2 electronics. Section A.2 describes the experimental setup of the beam test campaigns. The strategy and the results for the signal reconstruction are detailed in Sec. A.3, A.4, and A.5. Finally, the results concerning time and energy resolution measurements are presented in Sec. A.6.

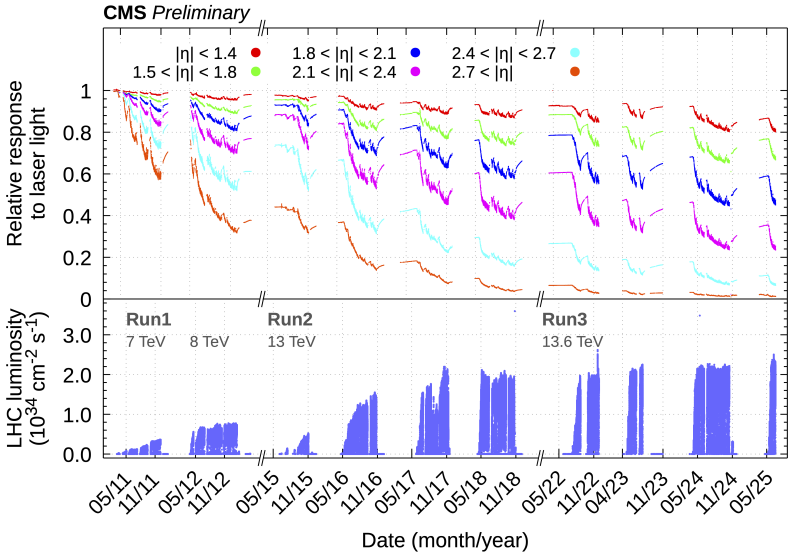


Figure A.1 Relative response to laser light injected into the ECAL crystals, measured by the laser monitoring system, averaged over all crystals in bins of $|\eta|$. The recovery of the crystal response during the periods without collisions is visible. The lower panel shows the LHC instantaneous luminosity as a function of time. [108]

A.1 The upgrade of the CMS ECAL

The primary predicted effects of operating at HL-LHC luminosities include a decrease in the light transmission of the $PbWO_4$ crystals, caused by irreversible radiation damage of the crystal lattice, and an increase of the APD leakage current, due to bulk damage in the silicon. These effects contribute to a reduction in the signal to noise ratio. The crystal transparency evolution since the CMS Run 1 is illustrated in Fig. A.1. The crystal scintillation light is monitored with a laser system that scans the crystals and reference photodiodes to measure relative response variations. At present the transparency loss in the barrel is around 20% and above 40% in the endcap.

Studies have been conducted to assess the impact of the radiation on crystals and photo-detectors, demonstrating that both components can be successfully operated up to the end of HL-LHC [107]. To mitigate the increase in the bulk current of the APDs, the operating temperature will be lowered from the current 18 °C to 9 °C. This will not require modifications to the existing on-detector cooling services. The lower temperature impact on the APD dark current evolution, as function of the integrated luminosity, is shown in Fig. A.2, separately for the central ($|\eta|$ =0) and forward region ($|\eta|$ =1.45) of the ECAL barrel.

To cope with higher trigger rates, improve the rejection of signals from direct ionization of the APDs (spikes), and mitigate the effect of additional interactions within the same or neighboring bunch crossings (in-time and out-of-time pileup, respectively), the front-end and back-end ECAL barrel readout electronics will be replaced during the Long Shutdown 3 of the LHC, which is scheduled from 2026 to 2029.

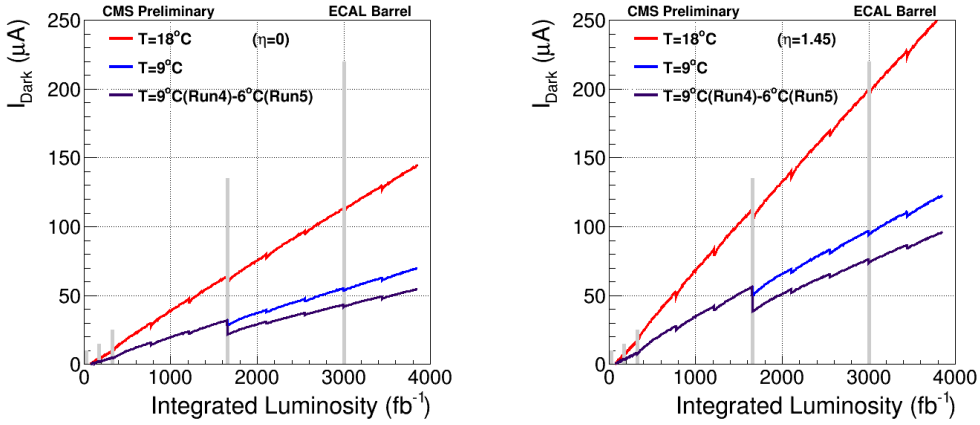


Figure A.2 The ECAL dark current as function of the total integrated luminosity of APDs in the central region ($|\eta|$ =0), on the left, and in the forwars region ($|\eta|$ =1.45), on the right. Three scenarios are shown: for APDs operated at 18 °C (red curve), at 9 °C (blue curve) and at 9 °C and then at a lower temperature of 6 °C afterwards (purple curve). The vertical shaded lines indicate long shutdowns. [104]

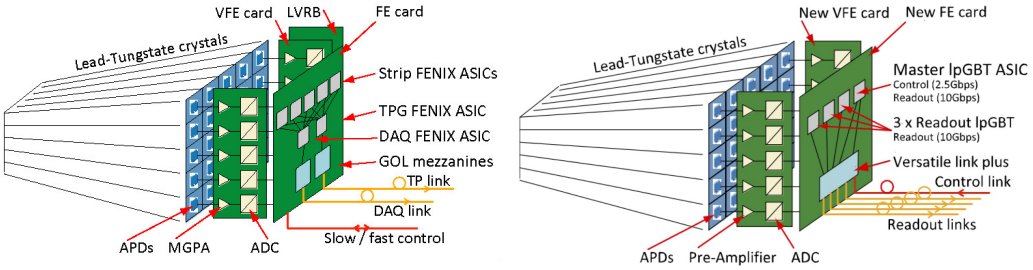


Figure A.3 Layout of the main components of a readout unit of the legacy electronics (left) and of the Phase-2 electronics (right). The motherboards–not shown in here–are located parallel to the FE boards, between the crystals and the 5 VFEs. A low-voltage board is present in both systems, although shown only for the legacy one.

A.1.1 Readout electronics

A schematic representation of both the legacy and Phase-2 electronics is shown in Fig. A.3. The arrangement of an array of 5×5 crystals forming a single readout unit is matched to the motherboards that distribute bias voltage to the APDs and low voltage to the very front-end (VFE) and front-end (FE) cards. The motherboards contain only passive components and have been tested, demonstrating that they can operate throughout the entire HL-LHC, thus requiring no refurbishment. Consequently, the geometrical layout of the legacy electronics has been retained in the upgraded design.

Front-end electronics

One VFE card reads out five channels; for each channel, the Phase-2 VFE card will include an amplifier, a digitizer, and a data-transmission unit, as illustrated in Fig. A.4. The amplifier ASIC, CAlorimeter TIA (CATIA) [109], is based on

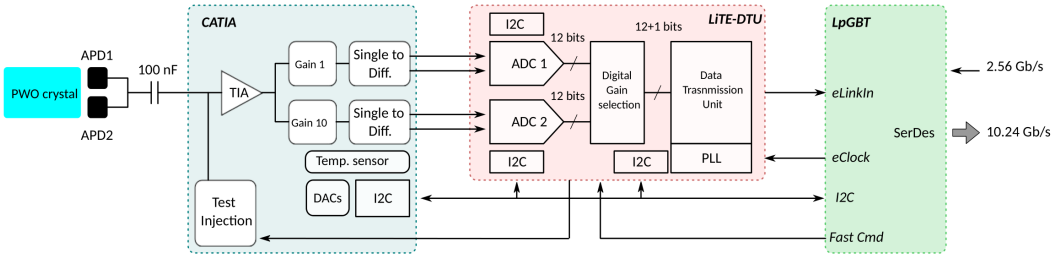


Figure A.4 Layout of the main components of the Phase-2 electronics along the signal path from the photodetectors to the back-end board. The CATIA and LiTE-DTU ASICs are located on the VFE board, while the green box represents the FE board.

a trans-impedance amplifier (TIA) that generates a voltage representation of the photocurrent produced by the APD. Its accuracy is primarily determined by the system bandwidth, limited by the kapton connection between the APDs and the motherboard, which is approximately 35 MHz. The bandwidth of the CATIA alone reaches 50 MHz. The output electronics pulses from the amplifier are approximately 40 ns long, compared to about 350 ns in the legacy system. The shorter pulse significantly reduces the overlap of signals from out-of-time pileup. Two different gains ($\times 1$ and $\times 10$) provide a dynamic range covering signals up to 2 TeV, with the gain switch occurring at around 200 GeV, with the precise value depending on the crystal light-yield.

The preamplifier is followed by a data conversion and transmission ASIC, the LiTE-DTU [110] (Lisboa-Torino ECAL data transmission unit). Each LiTE-DTU receives the analog signals from both gains of the CATIA, digitizes them at 160 MS/s, using 12-bits ADCs, and transmits the data to the FE board. To optimize the output data bandwidth, only the signal from the highest unsaturated gain is transmitted, and a lossless data compression scheme is applied. Whereby 6 bits are transmitted for baseline samples, and 13 bits for samples that exceed the 6-bit representation. The gain switch is handled by sending an extendable window of samples in gain 1, positioned around the first saturated sample, ensuring that a pulse that saturates in gain 10 is fully acquired in gain 1. The sampling frequency, 4 times higher than in the legacy system, has been chosen to achieve a time resolution of approximately 30 ps at energies higher than 50 GeV, which are in the relevant range for photons from Higgs boson to diphoton decays.

This sampling frequency provides performance that is independent of the sampling phase, and, along with the large bandwidth of the system, it also gives better discrimination between the signals from electromagnetic showers and the spikes. The latter can be suppressed almost completely by the L1T. Additionally, thanks to the improved time resolution, the precise knowledge of the electromagnetic particle arrival time will allow for a better mitigation of the in-time pileup [107]. The FE card utilizes recent developments in radiation-tolerant optical links, such as the Versatile Link plus [111] and the lpGBT [112], to serialize and transmit the digital data stream from each individual LiTE-DTU to the back-end electronic system for processing.

Back-end electronics

The back-end electronics transmits the data to the central data acquisition (DAQ) system of CMS and provides single-crystal trigger primitives to the L1T system. Hence, the ECAL information available at L1T will be 25 times greater than that provided by the legacy system, which is based on a 5×5 channel readout unit. The Barrel Calorimeter Processing boards [113] are designed with commercially available FPGAs and high speed optical links. The FPGAs are sufficiently powerful to provide a rejection of the spikes based both on the pulse shape and on topological variables.

A.2 Beam test campaigns

The first prototypes of the ECAL Phase-2 electronics have been tested at the CERN SPS North Area, on the H4 and H2 beam lines [114]. These tests concentrated primarily on the CATIA, LiTE-DTU, and low-voltage regulator boards. They were conducted in stages, starting with initial discrete-component prototypes and progressing to nearly final versions of the ASICs. In the absence of a complete set of the prototype components of the CMS DAQ chain, additional ad-hoc boards with similar functionalities have been produced to read out the data and to provide precise clock and trigger signals.

This section presents the results obtained from the 2018 and 2021 beam test campaigns, which featured the most recent detector prototypes available at the time. Greater emphasis is given to the results from the latter campaign, for which I developed the data analysis.

A.2.1 Beamline setup

The H4 beam is supplied with particles produced by 400 GeV protons extracted from the CERN SPS striking a 30 cm long beryllium target, and can be tuned to provide pions, electrons or photons. Secondary beams of charged pions are extracted directly after the primary target, achieving a purity greater than 90%, with the main contamination arising from the presence of electrons and muons. Tertiary beams of electrons or positrons are obtained by the conversion of photons from the decays of secondary neutral pions. The beam purity depends on the energy and is better than 98%. A specific set of beam optics provides a relative momentum spread $\Delta p/p$ for electrons and positrons as low as 0.5% [114].

Two sets of hodoscopes are located in the beamline upstream of the ECAL crystals, each consisting of two planes of orthogonal scintillating fibers. The hodoscopes provide a reference position in the transverse plane with a spatial resolution of approximately 150 μm. A reference for the arrival time of the beam particles is provided by two microchannel plate detectors (MCPs) [115], with a resolution of approximately 15 ps. A schematic representation of the experimental setup in 2018 and 2021 is shown in Fig. A.5.

The crystals are arranged in a 5×5 matrix, emulating an ECAL readout unit in the second module ($|\eta| \approx 1.09$), including the cooling system of the readout electronics. At variance with the arrangement in the ECAL, where all crystals at a given η are of

Figure A.5 Schematic representation of the experimental setup along the H4 beamline.

the same type, the matrix hosts crystals of type 5, 6, and 7. The matrix is installed inside a copper case, water-cooled with a temperature stability of 0.1 °C, in line with the nominal detector operations of CMS. The APDs are connected to the input of the VFE boards using Kapton-insulated cables of type and length identical to those used in CMS.

The copper case and the readout electronics are housed within a larger light-tight, thermally insulated aluminium box, which also contains an air convection system along with temperature and humidity sensors. A 20×20 cm^2 carbon fiber window in the front panel of the box reduces the amount of material encountered by the beam.

The temperature of the entire area around the box is kept constant at 18 °C. The temperature of the copper case can be adjusted to the operating conditions of the legacy system (18 °C) and of the Phase-2 system (9 °C). The dew point inside the box, regulated using a dry air flow, has been continuously monitored. The box is mounted on a remotely controllable table that can be moved in the transverse plane allowing the beam to impinge on each crystal of the matrix, in a way that replicates the geometry of the ECAL in CMS with respect to particles originating from the nominal interaction point. Pictures of the matrix housing are shown in Fig. A.6.

A.2.2 Front-end electronics prototypes

The 2018 campaign was conducted with VFEs incorporating a first prototype version (v0) of the CATIA amplifier. The objective of implementing this purely analogue device was to validate the TIA design, and it was paired to a 14-bit commercial ADC Analog-Device AD9642 digitizing signals at 160 MS/s. The VFEs were read out by an *ad-hoc* board designed to support a data stream from a single CATIA gain. The first complete prototype of a VFE, including the ASICs CATIA v1.2 and LiTE-DTUs v1.2, was tested for the first time during the 2021 campaign.

A.2.3 Trigger and data acquisition

The trigger to the DAQ readout is given by the coincidence of signals from at least two of three plastic scintillators measuring 3×3 cm^2, 1×1cm^{-2} and 2×2 cm^{-2}, as in Fig. A.5, and aligned along the beamline to select a portion of the beam that covers the desired surface. The DAQ software is based on a modular architecture that provides for the integration of several detector components and is capable of providing a run controller, an event builder, a fast data quality monitoring, and a system for data storage and bookkeeping. A more complete description of the DAQ can be found in [116].

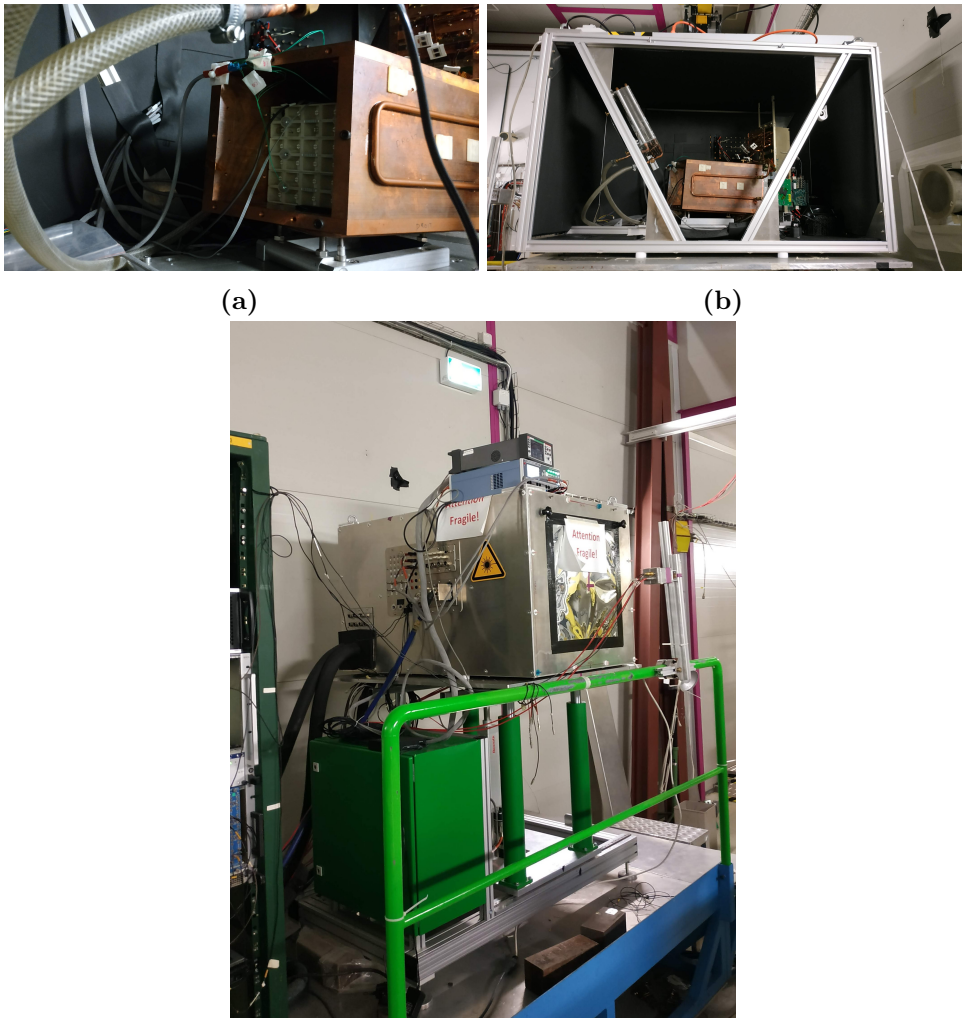


Figure A.6 The housing for the crystal array, visible within the copper case (a). The aluminum box containing the copper case (b). Also visible, above the case towards the rear, are unmounted APDs, used to measure with pions the signal produced by direct ionization. On the right side of the box, at the back, the patch panel for the VFE boards can be seen. The aluminum box is mounted on a remotely controlled table that allows the array to be moved in the transverse plane to center any selected crystal in the beam (c).

A.3 Time reference

The MCPs along the beamline provide a precise time reference for the arrival time of the particles. The signal from the MCPs is digitized at a sampling frequency of 5 GS/s by a CAEN V1742 board that processes the input signal through a switched-capacitor array of 1024 cells [117]. Each cell samples the signal with a delay of approximately 200 ps with respect to the previous one. This delay is fixed during the manufacture of the ASIC and has been calibrated using a precise sine wave generator. Approximately 10 000 input sinusoidal signals of the same frequency were

digitized, and a fit to these data was performed with the overall frequency and the delay of each cell treated as free parameters. This procedure provided an optimal synchronization. The time jitter between input channels within each of the two blocks of eight inputs of the CAEN V1742 board is negligible compared to the MCPs resolution.

The ECAL readout electronics and the CAEN V1742 board utilize two different clocks to digitize respectively the APD (160 MS/s) and the MCP signals (5 GS/s). In order to measure the phase between the two clocks, the one used by the first is digitized with the V1742 board, using a second channel. For each event, the relative phase $\bar{t}_{\mathrm{ECAL}} - \bar{t}_{\mathrm{MCP}}$ between the particle arrival time, measured in the ECAL and in the MCP, within one ADC clock period (T_{clk}), can be defined as

$$\Delta t \equiv \bar{t}_{\mathrm{MCP}} - t_{clk}, \tag{A.1}$$

$$\bar{t}_{\mathrm{ECAL}} - \bar{t}_{\mathrm{MCP}} \equiv (\bar{t}_{\mathrm{ECAL}} - \Delta t) \mod T_{\mathrm{clk}}. \tag{A.2}$$

Where the times $\bar{t}_{\mathrm{X}}$ and t_{clk} are all referred to the start of the acquisition windows in the corresponding boards. The time t_{clk} is referred to the rising edge of the clock signal digitized by the MCPs board, and $T_{\mathrm{clk}} = 6.238\,\mathrm{ns}$. The quantities $\bar{t}_{\mathrm{ECAL}}$ and $\bar{t}_{\mathrm{MCP}}$ are the reconstructed time of the signal maximum, measured separately in the ECAL and the MCPs.

A.4 Signal amplitude and time reconstruction

The amplitude $\mathcal{A}$ and the time $\bar{t}$ of the signals from the ECAL and the MCPs are extracted by fitting the measured waveforms with two template shapes. In the fit, a χ^2 is minimized, defined as

$$\chi^2 = \sum_i \frac{(A(t_i) - \mathcal{A} \cdot \mathcal{S}(t_i - \bar{t}))^2}{\sigma_{\mathrm{N}}^2}, \tag{A.3}$$

where $A(t_i)$ is the sample amplitude at the i-th sample time t_i, $\mathcal{S}$ is the normalized template shape evaluated at t_i, and $\mathcal{A}$ and $\bar{t}$ are the free parameters of the fit, representing the signal amplitude and arrival time of the pulse, respectively. The ECAL and MCPs signals are reconstructed fitting 9 and 40 points, respectively, around the maximum amplitude of the pulse.

By exploiting the time distribution of the particles inside a de-bunched extracted beam, which is random with respect to the DAQ clock phase, template shapes $\mathcal{S}$ for ECAL and MCPs pulses can be obtained by averaging normalized, phase-aligned signals. For the ECAL, only signals from crystals directly impacted by the beam are used. Events are selected within a 4×4 mm^2 window around the point of maximum shower containment, using the information from the hodoscopes, as shown in Fig. A.7. Within this window, the variation of the average signal amplitude along x and y is below 1%. Additionally, to minimize the contribution from the pion contamination in the electron/positron beam, events are required to be associated with a signal of amplitude larger than 200 ADC counts in the MCP located furthest upstream.

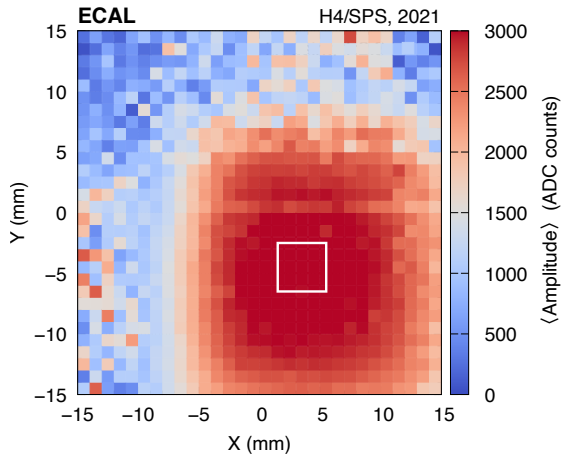


Figure A.7 Hodoscope heat map, where the x and y axis correspond to the coordinates on the hodoscope plates, and the z-axis represents the average signal amplitude measured in the crystal impacted by the electron beam. The white square contains the maximum shower energy at the position shown and indicates the region selected for the analysis.

In each event the phase and amplitude of the signal are initially reconstructed and then normalized and aligned. The first step is to increase the number of samples by a large amount, specifically by a factor of 50. This is performed in the frequency domain, using a Discrete Fourier Transformation (DFT) of the sampled signals, and by adding extra samples with values set at 0 in the frequency domain. An inverse DFT is then performed to obtain the oversampled signal. The oversampling provides a better resolution in the time domain, and gives a more precise alignment of the signals. The phase, $\bar{t}_{\text{o.p.}}$, and the amplitude of the oversampled pulses are obtained by fitting five samples centered around the maximum one with a second order polynomial. This procedure leaves a bias in the reconstruction of the phase as a function of the sampling phase, as can be seen in Fig. A.8 left. This bias is corrected by applying a sinusoidal curve fitted to the data, and the result is shown in Fig. A.8 right. For the 2018 campaign, signal templates were derived for ECAL channels using electrons with an energy of 150 GeV. The choice of this energy is based on the larger available data sample, as all crystals have data at this energy, obtained when intercalibrating their response. Additionally, measurements made with a charged-pion beam in the H2 beamline allowed the derivation of template shapes for signals resulting from the direct ionization of the APD sensitive volume. The comparison, presented in Fig. A.9, reveals non-negligible differences in the pulse shapes, that can be exploited for their discrimination. For the 2021 campaign, the dependency of the digitized waveforms on the particle energy and on the entire electronics readout chain, including the effects of light production and transmission in the crystals and the response of the APDs, is evaluated. To investigate the dependence on the particle energy, templates have been derived for a single channel of the ECAL matrix adjacent to the central one and read out by the same VFE card. Electron beams with energies ranging from 25 to 150 GeV are used and the comparison, presented in the left plot of Fig. A.10, shows that the signal shapes are independent of the particle energy up to 125 GeV, i.e. within the energy range

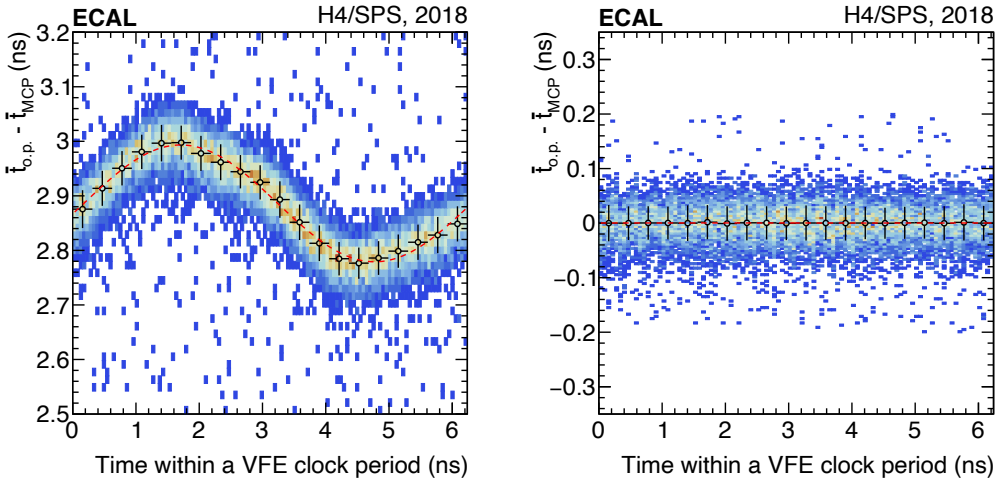


Figure A.8 Bias in alignment of ECAL pulses for the derivation of the template shape, $\bar{t}_{\text{o.p.}} - \bar{t}_{\text{MCP}}$, as a function of the ECAL sampling phase. Each point represents the mean of a Gaussian fit to the corresponding slice in x, while the error bars indicate the standard deviation of the Gaussian fit. A sinusoidal curve, shown in red, is fitted to the data and used to correct the bias. The right plot shows the residual bias after this correction has been applied.

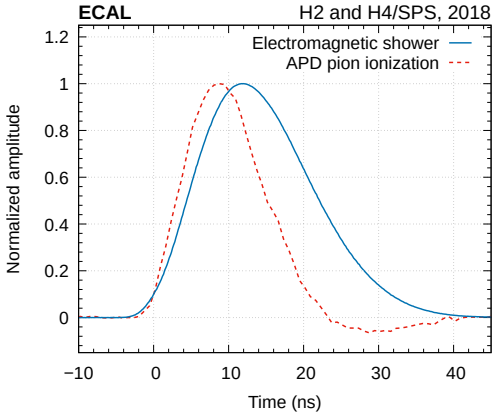
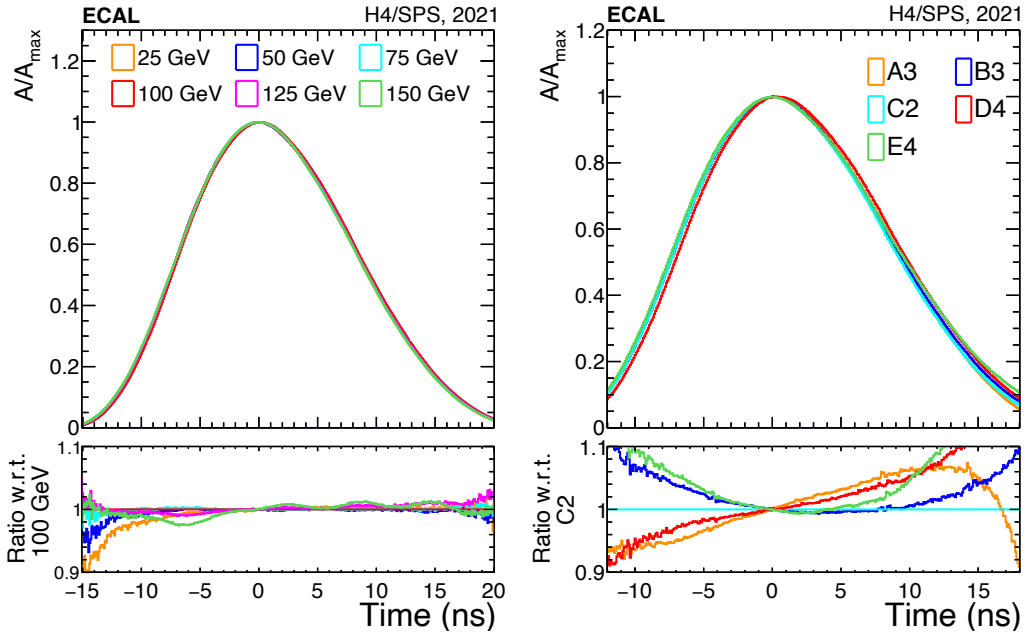


Figure A.9 Comparison of the signal shape for electromagnetic signal resulting from 150 GeV electrons (blue line) and for signals from direct ionization of the APDs measured with high energy pions (red line). The different smoothness of the curves is related to the different numbers of events available for deriving the shapes.



(a) (b)

Figure A.10 Comparison of the signal shapes measured for a selected channel at different beam energies (left) and for different channels on different VFE boards (right). The signal shape for a given channel is independent of the electron energy, while different channels have different signal shapes.

covered by gain 10. It was verified that, at a fixed beam energy, signal shapes for channels on the same VFE board are consistent. To examine the dependence on the crystal and electronics readout chain, templates were derived for five crystals on different VFE boards, using electrons at a fixed energy of 100 GeV. The comparison, presented in Fig. A.10 right, shows non-negligible differences in the pulse shapes, which need careful consideration when designing algorithms for amplitude and time reconstruction.

In light of these results, the signal shapes can be considered independent of particle energy up to 125 GeV, but they are specific to each individual channel. In the following analysis of 2021 data, amplitude and time are reconstructed using dedicated per-VFE templates derived at 100 GeV. For energies greater than 125 GeV, templates derived at 150 GeV are utilized.

A.5 Channel intercalibration

The energy intercalibration is obtained by comparing the response of individual channels to an electron beam directly impinging on each of them. Electrons of 150 GeV and of 100 GeV were used in the 2018 and 2021 beam tests,respectively. The intercalibration procedure is similar to that described in [118].

Events for intercalibration are selected by requiring the signal amplitude in the

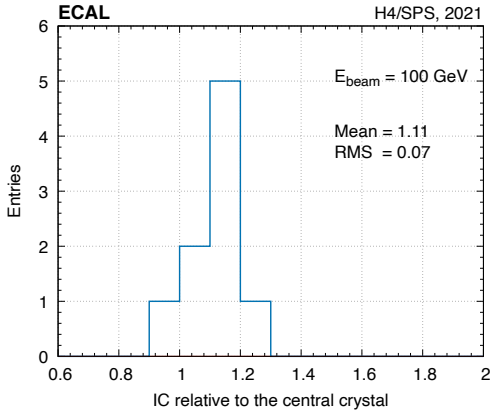


Figure A.11 Distribution of the values of the intercalibration coefficients relative to the central channel of the 3×3 crystal matrix for the 2021 beam test. The spread of the coefficients is compatible with the expected spread of the crystal light yield.

MCP located upstream to be greater than 200 ADC, to veto events from pions. Additionally, by using the hodoscope information the impinging particle is required to be within a window of $4{\times}4$ mm^2 centered on the position that maximizes the energy response of the crystal. Figure A.11 shows the distribution of the intercalibration coefficients relative to the central crystal obtained in the 2021 beam test for the central 3×3 matrix.

A.6 Results

During the 2018 campaign, measurements were made to compare the performances at operating temperatures of 18 °C and 9 °C. The pulse shapes are found to be identical, and the light yield increases by about 18% at the lower temperature, as expected from the known temperature dependence of $-2\%/\,°\text{C}$ [104]. All subsequent measurements have therefore been made at a single temperature, chosen to be 18 °C for ease of operation.

Although the energy resolution of the ECAL is driven by the physical design of the detector, which is unchanged, we have performed a cross-check measurement of the energy resolution using the energy deposited in a 3×3 matrix centered around the crystal hit by the beam.
Events are selected using the same criteria applied for the intercalibration described in Section A.5. Results from the 2021 campaign are derived only for the high gain mode. The limited amount of data available with low gain precludes a reliable measurement of the energy resolution in this mode.
The total amplitude in the 3×3 matrix is defined as the sum of the signal amplitudes of the nine intercalibrated channels. For each beam energy, the total amplitude distribution is fitted with a double-sided Crystal Ball function [89], as reported in Fig. A.12. Given the linear response of the ECAL to the energy of the impinging electromagnetic particles, the relative energy resolution is obtained from the ratio between the standard deviation and the mean of the Gaussian core of the Crystall

Ball.
The function modeling the relative energy resolution $\sigma(E)/E$ versus the beam energy, E, is represented by the quadratic sum of three terms:

$$\frac{\sigma(E)}{E} = \frac{N}{E} \oplus \frac{S}{\sqrt{E}} \oplus C, \tag{A.4}$$

where N, S and C are the noise, stochastic and constant terms respectively. The stochastic term S accounts for the statistical fluctuations of the electromagnetic shower in the crystal matrix and of the photons produced in each crystal. The constant term C includes effects from the longitudinal leakage of the electromagnetic showers in the crystal matrix and from intercalibration uncertainties. In the fit of the relative energy resolution as a function of the beam energy, S and C are the free parameters. The term N quantifies the average noise of the readout electronics of the 3×3 matrix. As this term can be measured independently, it is fixed in the fit of the energy resolution to reduce parameter correlations.
For a single channel, the impact of the noise on the amplitude reconstruction has been estimated using a toy simulation. The noise, defined as the sample RMS, is measured on dedicated runs without beam. The full sample-to-sample covariance matrix is also measured. The noise is then added to simulated waveforms of a fixed amplitude. The RMS of the reconstructed amplitude is found to be equal to that of the noise, increased by 2.5%, independently of the amplitude.
To estimate the term N in Eq. A.4 and to properly account for channel-to-channel correlations, the noise of the 3×3 array is estimated from the noise of the individual channels as:

$$N = 1.025 \cdot \sqrt{\mathbf{B^T C B}}, \tag{A.5}$$

where $\mathbf{B}$ is the 9-dimension vector containing the average of the single channel noise, $\mathbf{C}$ is the 9×9 correlation matrix of the noise between the different channels, and the factor 1.025 accounts for the 2.5% noise increase described above. This correlation matrix is reported in Fig. A.13.

The values of the beam energy are corrected for estimated synchrotron radiation losses along the beamline, as provided in the beamline documentation and derived from a simulation of the beamline optics and components. The Phase-2 ECAL electronics targets a value for the constant term C lower than 1% at energies higher than 50 GeV. The energy resolution as function of the beam energy, measured in the two beam test campaigns, along with the fit of the S and C terms are shown in Fig. A.14.

The stochastic term is measured to be (0.029 ± 0.006) $\text{GeV}^{1/2}$ and (0.020 ± 0.010) $\text{GeV}^{1/2}$ for the 2018 and 2021 campaigns, respectively, in agreement with beam test results of the legacy system [119]. The constant term is measured to be (0.37 ± 0.03)% and (0.58 ± 0.03)% for the 2018 and 2021 campaigns, respectively. The difference between these two values is attributed to a misalignment of the beam with respect to the crystal matrix, which did not occur in the 2018 campaign. The constant term measured at 2018 beam test is slightly higher than the one measured in beam tests with the legacy system. This is assumed to be a consequence of the test matrix using crystals originating from the commissioning production batches, which

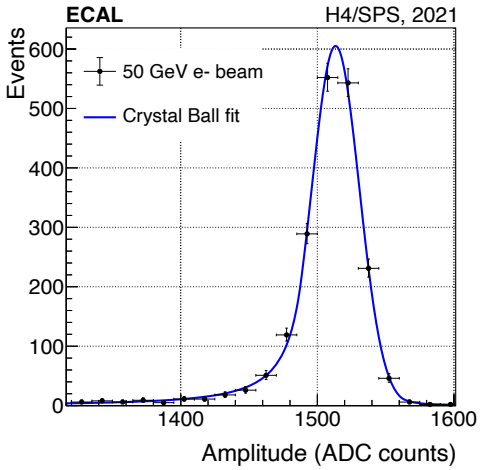


Figure A.12 Example of the fit to the total amplitude distribution of the 3×3 crystal matrix for a beam energy of 50 GeV. The blue line represents a Crystal Ball fit to the data, indicated by the black points.

have suboptimal characteristics compared to those used in ECAL. Additionally, the amount of material in front of the calorimeter was different in the two sets of measurements, and the 2018 and 2021 results have not been corrected for contributions from the energy spread of the beam. Overall, given these considerations, the energy resolution performance meets the requirements for the HL-LHC.

Using events where the beam hit the center of a crystal, the single ECAL channel time resolution is extracted from the width of the $\bar{t}_{\mathrm{ECAL}} - \bar{t}_{\mathrm{MCP}}$ distribution for electrons impinging within a window of $1{\times}1$ cm^2 around the position of the maximum shower containment. The time resolution for an individual channel is modeled as a function of the effective amplitude, i.e. the measured amplitude A normalized to the channel average noise RMS σ_n. For any given channel, the quantities A and σ_n are derived independently, with the amplitude A obtained from the template fit to the digitized signal and σ_n measured from the average of the pedestal RMS.
The analytic expression of the time resolution function, used to perform the fit to the data, is given by the quadratic sum of two terms:

$$\sigma_t = \frac{N}{A/\sigma_n} \oplus C, \tag{A.6}$$

where N is a noise term and C is the constant term. The contribution of the stochastic term is consistently found to be negligible in all beam test campaigns with legacy and Phase-2 electronics, and is therefore omitted from Eq. (A.6). The results are shown in Fig. A.15, where each point on the x-axis is obtained from the peak of a Crystal Ball function fitted to the distribution of A/σ_n.
The MCP time resolution, σ_t^{MCP}, is subtracted in quadrature from each of the measured $\sigma(\bar{t}_{\mathrm{ECAL}} - \bar{t}_{\mathrm{MCP}})$ values, before fitting the data. Assuming the MCPs are identical, σ_t^{MCP} is measured as the spread of the time difference between the two MCPs, divided by $\sqrt{2}$, considering all the events contributing to each $\sigma(\bar{t}_{\mathrm{ECAL}} - \bar{t}_{\mathrm{MCP}})$

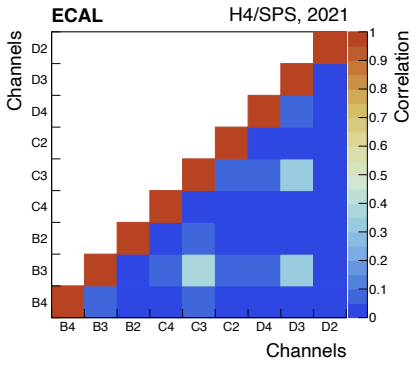
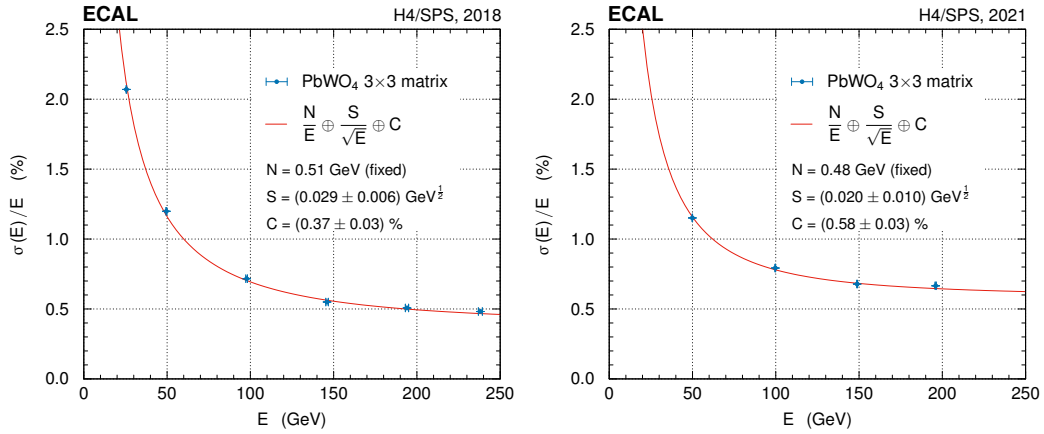


Figure A.13 Correlation matrix of the pedestal RMS for the 9 channels of the 3×3 matrix **C**.



(a) **(b)**

Figure A.14 Relative energy resolution as a function of the beam energy obtained for the 2018 (a) and 2021 (b) campaigns. The energy deposits are measured in a 3×3 matrix centered on the crystal impacted by the beam. Because of the significant correlation between the fit parameters, the noise term N is fixed to the value measured in dedicated runs without beam. The error bars represent the statistical uncertainties.

measurement, separately. The value of σ_t^{MCP} unfolded from each point is basically independent of the beam energy and amounts to approximately 13 ps.
During the 2021 beam test, the time resolution was measured both for low and high electronics gain. Only runs with beam energies exceeding 150 GeV have a sufficient amount of data in the low gain mode, with measurements at 200 and 250 GeV.

The results fulfill the design goals of having a resolution of 30 ps or better at the beginning of HL-LHC for energies greater than 50 GeV.

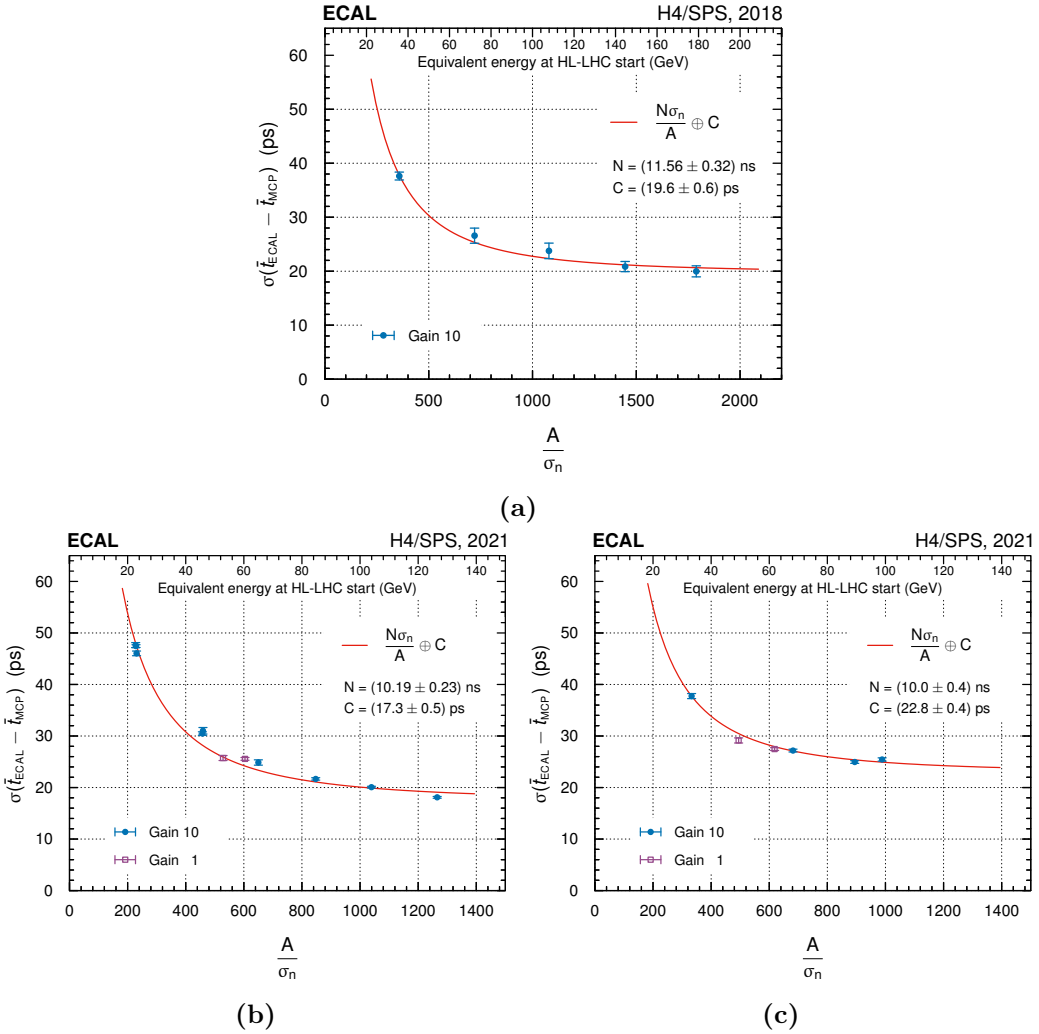


Figure A.15 Time resolution as a function of the average reconstructed amplitude normalized by the noise RMS, A/σ_n, for a single crystal with the beam impinging within a 1×1 cm^2 window, positioned to contain the maximum shower energy. Figures (a) and (b) show the resolution for the central channel of the ECAL matrix during the 2018 and 2021 beam tests, respectively, while (c) presents the resolution for a crystal adjacent to the central one. All plots present results after the subtraction of the MCP contribution to the resolution, as described in the text. The results for the 2021 beam test also include two points obtained at high energies, for which the electronics readout switches to gain 1. The error bars represent the statistical uncertainties. The red curve is a fit of Eq. A.6 to the points in gain 10.

A.7 Summary

Results have been presented from studies of the performance of first prototypes of the CMS ECAL barrel readout electronics that will replace the existing system during the future high luminosity operation of the LHC. The measurements were made with high energy electrons and pions in the H4 and H2 beam lines of the CERN SPS, in separate campaigns in 2018 and 2021, to whose data analysis I contributed. The

performance of the Phase-2 readout electronics has been characterized in terms of time and energy resolutions, measured on a 5×5 matrix of lead tungstate crystals. The results for the energy resolution, primarily driven by the design of the detector, in absence of photodetector aging and crystal irradiation, are compatible with previous publications with the legacy electronics. In both campaigns the constant term of the energy resolution is measured to be better than 0.6% and the time resolutions for electrons with energies above 50 GeV is measured to be better than 30 ps, fulfilling the design requirements.
Since 2021, additional test beam campaigns have been carried out annually with a spare supermodule equipped with newer readout prototypes. I participated to the data-taking, and the ongoing analysis of these datasets will provide a performance characterization of the electronics for the full readout chain.

Acronyms

ALICE A Large Ion Collider Experiment. 24

APD avalanche photodiodes. 33, 137, 138

ATLAS A Toroidal LHC ApparatuS. 24

AUC Receiver Operating Characteristic Area under the Curve. 81, 82, 84, 85

BDT Boosted Decision Tree. 20, 21, 77, 78, 81–85, 95, 106

BMTF barrel muon track finder. 43, 44

BMTF overlap muon track finder. 43

BS beam-spot. 60, 95

BSM Beyond the Standard Model. 15, 16, 26, 49, 134

CATIA CAlorimeter TIA. 138

CCI charged current interactions. 7

CERN Conseil Européen pour la Recherche Nucléaire. 24

CKM Cabibbo-Kobayashi-Maskawa. 11, 13, 14

CL confidence level. 17, 19–21, 126, 127

CLFV charged lepton flavor violation. 13, 15, 16

CMS Compact Muon Solenoid. 1, 2, 17, 24, 26, 37, 38, 40, 46, 50, 52, 53, 110, 118, 136

CSC cathode strip chamber. 36, 37, 43

DCA distance of closest approach. 54

DT drift tube. 36, 37, 43

ECAL elecromagnetic calorimeter. 32, 136

EMTF endcap muon track finder. 43

EW electroweak. 6

FE front-end. 138

FSR final-state radiation. 18

GEM gas electron multiplier. 36, 37

GT Global Trigger. 42–44

HCAL hadronic calorimeter. 32, 34, 35

HF heavy flavor. 17, 21, 42, 51, 106

HL-LHC High Luminosity LHC. 25, 26, 34, 37

HLT High Level Trigger. 42, 44–47, 53–57, 59, 60, 63, 72, 101, 110

IP interaction point. 29

ISR initial-state radiation. 18

L1T Level 1 Trigger. 42, 44, 45, 48, 54, 55, 72, 139, 140

LEP Large Electron Positron. 24

LFV lepton flavor violation. 1, 2, 14–18, 24, 50, 51, 57, 81, 106, 108, 122

LHC Large Hadron Collider. 1, 2, 24–26, 28, 29, 42, 48, 51, 55

LHCb Large Hadron Collider beauty. 25

LLP long-lived particle. 49

LO leading order. 52, 74, 75, 81

MC Monte Carlo. 52, 53, 58, 62, 63, 69, 71, 74, 75, 77, 81, 87, 91, 93, 95, 100, 101, 103–105

MET missing transverse energy. 51, 60–62, 107

ML machine learning. 78, 81, 134

NCI neutral current interactions. 7

NLO next-to-leading order. 75, 81, 102, 119

NP new Physics. 17

p.d.f. probability density function. 112, 122, 124

PACT pattern comparator trigger. 43

PDFs parton distribution functions. 27

PF Particle Flow. 37–41, 44, 45, 57, 60

PL physical limit. 14, 15

PMNS Pontecorvo-Maki-Nakagawa-Sakata. 14

POI parameter of interest. 111, 123, 128

PU pileup. 28, 48, 54, 60, 61, 69, 70

PUPPI Pileup Per Particle Identification. 60, 102

PV primary vertex. 60, 69, 82

QCD quantum chromo-dynamics. 6, 9, 17, 119

ROC Receiver Operating Characteristic. 85

RPC resistive plate chamber. 36, 37, 43

SM Standard Model. ii, 1, 3, 5, 6, 8, 9, 12, 13, 15–18, 26, 52, 77, 95, 134

SPS Super Proton Syncroton. 136

SSB spontaneous symmetry breaking. 9, 11, 17

SUSY supersymmetry. 16

SV secondary vertex. 55, 57, 58, 60, 65, 66, 73, 81, 82, 85, 95–97, 119

TP trigger primitive. 43

TPG Trigger Primitive Generator. 43

UL upper limit. 92, 95, 96

VBF vector-boson fusion. 49

VFE very front-end. 138

Bibliography

[1] S. L. Glashow. Partial-symmetries of weak interactions. *Nuclear Physics*, 22:579–588, 2 1961. doi:10.1016/0029-5582(61)90469-2.

[2] S. Weinberg. A Model of Leptons. *Physical Review Letters*, 19:1264, 1967. doi:10.1103/PhysRevLett.19.1264.

[3] A. Salam. Weak and electromagnetic interactions. *Selected Papers of Abdus Salam*, May 1994. doi:10.1142/9789812795915_0034.

[4] F. Englert and R. Brout. Broken Symmetry and the Mass of Gauge Vector Mesons. *Physical Review Letters*, 13:321–323, Aug 1964. doi:10.1103/PhysRevLett.13.321.

[5] P. W. Higgs. Broken Symmetries and the Masses of Gauge Bosons. *Physical Review Letters*, 13:508–509, Oct 1964. doi:10.1103/PhysRevLett.13.508.

[6] M. J. Herrero. The Standard Model. 1998. arXiv:hep-ph/9812242.

[7] J. C. Street and E. C. Stevenson. New Evidence for the Existence of a Particle of Mass Intermediate Between the Proton and Electron. *Phys. Rev.*, 52:1003–1004, Nov 1937. doi:10.1103/PhysRev.52.1003.

[8] M. L. Perl et al. Evidence for Anomalous Lepton Production in e^+e^- Annihilation. *Physical Review Letters*, 35:1489–1492, Dec 1975. doi:10.1103/PhysRevLett.35.1489.

[9] E. Noether. Invariante Variationsprobleme. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, 1918:235–257, 1918. URL: http://eudml.org/doc/59024.

[10] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson. Experimental Test of Parity Conservation in Beta Decay. *Phys. Rev.*, 105:1413–1415, Feb 1957. doi:10.1103/PhysRev.105.1413.

[11] A. Pich. The Standard Model of Electroweak Interactions, 2007. arXiv:0705.4264.

[12] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. Global Conservation Laws and Massless Particles. *Physical Review Letter*, 13:585–587, Nov 1964. doi:10.1103/PhysRevLett.13.585.

[13] PDG. CKM Quark-Mixing Matrix. 6 2020. URL: https://pdg.lbl.gov/2021/reviews/rpp2020-rev-ckm-matrix.pdf.

[14] L. Wolfenstein. Parametrization of the Kobayashi-Maskawa Matrix. *Physical Review Letters*, 51:1945–1947, Nov 1983. doi:10.1103/PhysRevLett.51.1945.

[15] J. Ellis. Outstanding questions: Physics beyond the Standard Model. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 370(1961):818–830, Feb 2012. doi:10.1098/rsta.2011.0452.

[16] G. Bertone, D. Hooper, and J. Silk. Particle dark matter: Evidence, candidates and constraints. *Physics Reports*, 405:279–390, 2005. doi:10.1016/j.physrep.2004.08.031.

[17] A. D. Sakharov. Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe. *Pisma Zh. Eksp. Teor. Fiz.*, 5:32–35, 1967. doi:10.1070/PU1991v034n05ABEH002497.

[18] D. Bödeker and W. Buchmüller. Baryogenesis from the weak scale to the grand unification scale. *Reviews of Modern Physics*, 93(3), Aug 2021. doi:10.1103/revmodphys.93.035004.

[19] Y. Ashie et al. Evidence for an oscillatory signature in atmospheric neutrino oscillation. *Physical Review Letters*, 93, 2004. doi:10.1103/PhysRevLett.93.101801.

[20] Super-Kamiokande Collaboration. Constraint on the matter–antimatter symmetry-violating phase in neutrino oscillations. *Nature*, 580(7803):339–344, 2020. doi:10.1038/s41586-020-2177-0.

[21] S. Navas et al. Review of particle physics. *Phys. Rev. D*, 110(3):030001, 2024. doi:10.1103/PhysRevD.110.030001.

[22] C. Giganti, S. Lavignac, and M. Zito. Neutrino oscillations: The rise of the PMNS paradigm. *Progress in Particle and Nuclear Physics*, 98:1–54, 2018. doi:10.1016/j.ppnp.2017.10.001.

[23] P. Blackstone, M. Fael, and E. Passemar. $\tau \to \mu\mu\mu$ at a rate of one out of 10^{14} tau decays? *The European Physical Journal C*, 80(6):506, 2020. doi:10.1140/epjc/s10052-020-8059-7.

[24] J. Heeck. Interpretation of lepton flavor violation. *Physiscs Review D*, 95:015022, Jan 2017. doi:10.1103/PhysRevD.95.015022.

[25] MEG Collaboration. Search for the Lepton Flavour Violating Decay $\mu^+ \to e^+\gamma$ with the Full Dataset of the MEG Experiment, 2016. arXiv:1605.05081.

[26] U. Bellgardt et al. Search for the Decay $\mu^+ \to e^+e^+e^-$. *Nucl. Phys. B*, 299:1–6, 1988. doi:10.1016/0550-3213(88)90462-2.

[27] Belle Collaboration. Search for lepton-flavor-violating τ decays into three leptons with 719 million produced $\tau^+\tau^-$ pairs. *Physics Letters B*, 687(2):139–143, 2010. doi:10.1016/j.physletb.2010.03.037.

[28] BABAR Collaboration. Searches for Lepton Flavor Violation in the Decays $\tau^\pm \to e^\pm\gamma$ and $\tau^\pm \to \mu^\pm\gamma$. *Physical Review Letters*, 104:021802, Jan 2010. doi:10.1103/PhysRevLett.104.021802.

[29] Belle2 Collaboration. Search for lepton-flavor-violating $\tau \to \mu\mu\mu$ decays at Belle2. *J. High Energ. Phys.*, (62), 2024. doi:10.1007/JHEP09(2024)062.

[30] K. J. Healey, A. A. Petrov, and D. Zhuridov. Nonstandard neutrino interactions and transition magnetic moments. *Phys. Rev. D*, 87(11):117301, 2013. [Erratum: Phys.Rev.D 89, 059904 (2014)]. arXiv:1305.0584, doi:10.1103/PhysRevD.87.117301.

[31] G. Cvetič, C. Dib, C. S. Kim, and J. D. Kim. Lepton flavor violation in tau decays. *Phys. Rev. D*, 66:034008, Aug 2002. doi:10.1103/PhysRevD.66.034008.

[32] S.T Petcov, S Profumo, Y Takanishi, and C.E Yaguna. Charged lepton flavor violating decays: leading logarithmic approximation versus full RG results. *Nuclear Physics B*, 676(1–2):453–480, Jan 2004. doi:10.1016/j.nuclphysb.2003.10.020.

[33] Dedes A., Ellis J., and Raidal M. Higgs-mediated B_s,$D^0 \to \mu\tau$,$e\tau$ and $\tau \to 3\mu$,$e\mu\mu$ decays in supersymmetric seesaw models. *Physics Letters B*, 549(1):159–169, 2002. doi:10.1016/S0370-2693(02)02900-3.

[34] M. Kakizaki, Y. Ogura, and F. Shima. Lepton flavor violation in the triplet Higgs model. *Physics Letters B*, 566(3–4):210–216, Jul 2003. doi:10.1016/s0370-2693(03)00833-5.

[35] M. Bordone, C. Cornella, J. Fuentes-Martín, and G. Isidori. Low-energy signatures of the PS3 model: from B-physics anomalies to LFV. *Journal of High Energy Physics*, 2018(10):148, 2018. doi:10.1007/JHEP10(2018)148.

[36] Belle Collaboration. Search for lepton-flavor-violating τ decays into three leptons with 719 million produced $\tau\tau$ pairs. *Phys. Lett. B*, (687):139–143, 2010. doi:10.1016/j.physletb.2010.03.037.

[37] CMS Collaboration. Search for the lepton flavor violating $\tau \to 3\mu$ decay in proton-proton collisions at $\sqrt{s}$=13 TeV. *Phys. Lett. B*, (853):138633, 2024. doi:10.1016/j.physletb.2024.138633.

[38] BABAR Collaboration. Limits on τ lepton-flavor violating decays into three charged leptons. *Phys. Rev. D*, (81):111101, 2010. doi:10.1103/PhysRevD.81.111101.

[39] LHCb Collaboration. Search for the lepton flavour violating decay $\tau \to \mu\mu\mu$. *J. High Energ. Phys.*, (121), 2015. doi:10.1007/JHEP02(2015)121.

[40] ATLAS Collaboration. Probing lepton flavour violation via neutrinoless $\tau \to \mu$ decays with the ATLAS detector. *Eur. Phys. J. C*, (76):232, 2016. doi:10.1140/epjc/s10052-016-4041-9.

[41] Belle Collaboration. The Belle detector. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 479(1):117–232, 2002. Detectors for Asymmetric B-factories. doi:10.1016/S0168-9002(01)02013-7.

[42] BelleII Collaboration. Belle II Technical Design Report, 2010. arXiv:1011.0352.

[43] The BABAR detector. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 479(1):1–116, Feb 2002. doi:10.1016/s0168-9002(01)02012-5.

[44] CMS Collaboration. Search for the lepton flavor violating decay $\tau \to 3\mu$ in proton-proton collisions at $\sqrt{s}$= 13 TeV. *Journal of High Energy Physics*, 2021(1):163, 2021. doi:10.1007/JHEP01(2021)163.

[45] CMS Collaboration. Search for the lepton flavor violating $\tau \to 3\mu$ decay in proton-proton collisions at s=13TeV. *Physics Letters B*, 853:138633, 2024. doi:10.1016/j.physletb.2024.138633.

[46] LHC Design Report Vol.1: The LHC Main Ring. Jun 2004. doi:10.5170/CERN-2004-003-V-1.

[47] ATLAS Collaboration. *ATLAS detector and physics performance*. Technical design report. ATLAS. CERN, Geneva, 1999. URL: https://cds.cern.ch/record/391176.

[48] ALICE Collaboration. The ALICE experiment at the CERN LHC. *Journal of Instrumentation*, 3(08):S08002, aug 2008. doi:10.1088/1748-0221/3/08/S08002.

[49] LHCb Collaboration. LHCb Upgrade II Scoping Document. Technical report, CERN, Geneva, 2024. doi:10.17181/CERN.2RXP.HDK0.

[50] G. Altarelli and G. Parisi. Asymptotic freedom in parton language. *Nuclear Physics B*, 126(2):298–318, 1977. doi:10.1016/0550-3213(77)90384-4.

[51] CMS Collaboration. CMS Luminosity Public Results. https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults#Full_summary_proton_proton_colli, 2025.

[52] CMS Collaboration. Development of the CMS detector for the CERN LHC Run 3. *JINST*, 19(05):P05064, 2024. arXiv:2309.05466, doi:10.1088/1748-0221/19/05/P05064.

[53] CMS Collaboration. Precise Mapping of the Magnetic Field in the CMS Barrel Yoke using Cosmic Rays. *Journal of Instrumentation*, 5, Oct 2009. doi:10.1088/1748-0221/5/03/T03021.

[54] CMS Collaboration. The CMS tracker system project:Technical Design Report. Technical Report CERN-LHCC-98-006, CMS-TDR-5, 1997.

[55] CMS Collaboration. The CMS inner tracker – transition from LHC Run I to Run II and first experience of Run II. page 229, Mar 2016. doi:10.22323/1.234.0229.

[56] CMS Collaboration. The CMS electromagnetic calorimeter project: Technical Design Report. (CERN-LHCC-97-033, CMS-TDR-4), 1997.

[57] P. Adzic et al. Energy resolution of the barrel of the CMS Electromagnetic Calorimeter. *Journal of Instrumentation*, 2(04):P04004, apr 2007. doi:10.1088/1748-0221/2/04/P04004.

[58] CMS Collaboration. ECAL performance with reprocessed 2024 data. 2025. URL: https://cds.cern.ch/record/2938117.

[59] CMS Collaboration. *The CMS hadron calorimeter project: Technical Design Report*. Technical design report. CMS. CERN, Geneva, 1997. URL: https://cds.cern.ch/record/357153.

[60] J. G. Layter. *The CMS muon project: Technical Design Report*. Technical design report. CMS. CERN, Geneva, 1997. URL: https://cds.cern.ch/record/343814.

[61] CMS Collaboration. Particle-flow reconstruction and global event description with the CMS detector. *JINST*, 12(10), 2017. arXiv:1706.04965, doi:10.1088/1748-0221/12/10/P10003.

[62] The CMS trigger system. *Journal of Instrumentation*, 12(01):P01020, Jan 2017. doi:10.1088/1748-0221/12/01/P01020.

[63] CMS Collaboration. CMS TriDAS project. URL: https://cds.cern.ch/record/706847.

[64] CMS Collaboration. CMS Technical Design Report for the Level-1 Trigger Upgrade. (CERN-LHCC-2013-011, CMS-TDR-12), Jun 2013.

[65] A. Bocci, V. Innocente, M. Kortelainen, F. Pantaleo, and M. Rovere. Heterogeneous Reconstruction of Tracks and Primary Vertices With the CMS Pixel Tracker. *Frontiers in Big Data*, Volume 3 - 2020, 2020. doi:10.3389/fdata.2020.601728.

[66] CMS Collaboration. Performance of Run-3 HLT Track Reconstruction. 2022. URL: https://cds.cern.ch/record/2814111.

[67] CMS Collaboration. Performance of the CMS muon trigger system in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Journal of Instrumentation*, 16(07):P07001, Jul 2021. doi:10.1088/1748-0221/16/07/P07001.

[68] CMS Collaboration. Enriching the physics program of the CMS experiment via data scouting and data parking. *Phys. Rept.*, 1115:678–772, 2025. doi:10.1016/j.physrep.2024.09.006.

[69] CMS Collaboration. Evolution of the High-Level Trigger output rates of the CMS experiment from 2012 to 2023. 2024. URL: https://cds.cern.ch/record/2890105.

[70] CMS Collaboration. Test of lepton flavor universality in $B^{\pm} \to K^{\pm}\mu^{+}\mu^{-}$ and $B^{\pm} \to K^{\pm}e^{+}e^{-}$ decays in proton-proton collisions at $\sqrt{s}$=13 TeV. *Reports on Progress in Physics*, 87(7):077802, Jul 2024. doi:10.1088/1361-6633/ad4e65.

[71] CMS Collaboration. Output rates and bandwidth of the High-Level Trigger in the 2023 and 2024 proton-proton runs. 2025. URL: https://cds.cern.ch/record/2931957.

[72] T. Sjostrand, S. Mrenna, and P. Z. Skands. A Brief Introduction to PYTHIA 8.1. *Comput. Phys. Commun.*, (178):852–867, 2008. doi:10.1016/j.cpc.2008.01.036.

[73] CMS Collaboration. Extraction and validation of a new set of CMS PYTHIA 8 tunes from underlying-event measurements. *Eur. Phys. J. C*, (80):4, 2020. doi:10.1140/epjc/s10052-019-7499-4.

[74] NNPDF Collaboration. Parton distributions from high-precision collider data. *Eur. Phys. J. C*, (77):663, 2017. doi:10.1140/epjc/s10052-017-5199-5.

[75] CMS Collaboration. Measurements of the inclusive W and Z boson production cross sections and their ratios in proton-proton collisions at $\sqrt{s}$ = 13.6 TeV, 2025. arXiv:2503.09742.

[76] A. Ryd, D. Lange, N. Kuznetsova, S. Versille, M. Rotondo, D. P. Kirkby, F. K. Wuerthwein, and A. Ishikawa. EvtGen: A Monte Carlo Generator for B-Physics. (EVTGEN-V00-11-07), EVTGEN-V00-11-07, 2005.

[77] J. Alwall et al. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *J. High Energ. Phys.*, (07):079, 2014. doi:10.1007/JHEP07(2014)079.

[78] D. Bertolini, P. Harris, M. Low, and N. Tran. Pileup per particle identification. *Journal of High Energy Physics*, 2014(10), Oct 2014. doi:10.1007/jhep10(2014)059.

[79] https://xgboost.readthedocs.io/en/stable/index.html.

[80] Y. Coadou. *Boosted Decision Trees*, page 9–58. World Scientific, Feb 2022. doi:10.1142/9789811234033_0002.

[81] A. S. Cornell, W. Doorsamy, B. Fuks, G. Harmsen, and L. Mason. Boosted decision trees in the era of new physics: a smuon analysis case study. *Journal of High Energy Physics*, 2022(4), Apr 2022. doi:10.1007/jhep04(2022)015.

[82] G. Louppe. Understanding Random Forests: From Theory to Practice, 2014. doi:10.48550/ARXIV.1407.7502.

[83] A. Takuya, S. Shotaro, Y. Toshihiko, O. Takeru, and Masanori K. Optuna: A Next-generation Hyperparameter Optimization Framework, 2019. arXiv:1907.10902.

[84] L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001. doi:10.1023/A:1010933404324.

[85] J. S. Gábor and L. R. Maria. On the uniqueness of distance covariance. *Statistics Probability Letters*, 82(12):2278–2282, 2012. doi:10.1016/j.spl.2012.08.007.

[86] G. Punzi. Sensitivity of searches for new signals and its optimization, 2003. arXiv:physics/0308063.

[87] CMS Collaboration. Search for the Rare Decay $D^0 \to \mu^+\mu^-$ in Proton-Proton Collisions at $\sqrt{s}$=13.6 TeV. *Physical Review Letters*, 135(15):151803, 2025. doi:10.1103/zc76-rgcp.

[88] M. Pivk and F. R. Le Diberder. sPlots: A statistical tool to unfold data distributions. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 555:356–369, Dec 2005. doi:10.1016/J.NIMA.2005.08.106.

[89] T. Skwarnicki. *A study of the radiative cascade transitions between the Upsilon-Prime and Upsilon resonances*. PhD thesis, Cracow, INP, 1986.

[90] Bernstein, S. Démonstration du théorème de Weierstrass fondée sur le calcul des probabilités. *Communications de la Société Mathématique de Kharkov*, 13:1–2, 1912.

[91] Arfken, G. Chebyshev (Tschebyscheff) Polynomials and Chebyshev Polynomials–Numerical Applications. In *Mathematical Methods for Physicists*, chapter 13.3 and 13.4, pages 731–748. Academic Press, Orlando, FL, 3rd edition, 1985.

[92] CMS Collaboration. Luminosity measurement in proton-proton collisions at 13.6 TeV in 2022 at CMS. Technical Report CMS-PAS-LUM-22-001, CERN, Geneva, 2024. URL: https://cds.cern.ch/record/2890833.

[93] CMS Collaboration. Measurement of the offline integrated luminosity for the CMS proton-proton collision dataset recorded in 2023. 2024. URL: https://cds.cern.ch/record/2904808.

[94] A. L. Read. Presentation of search results: the cls technique. *Journal of Physics G: Nuclear and Particle Physics*, 28(10):2693, Sep 2002. doi:10.1088/0954-3899/28/10/313.

[95] A. Hayrapetyan et al. The CMS statistical analysis and combination tool: Combine. *Comput. Softw. Big Sci.*, 8:19, 2024. arXiv:2404.06614, doi:10.1007/s41781-024-00121-4.

[96] Wilks, S. S. The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. *Annals Math. Statist.*, 9(1):60–62, 1938. doi:10.1214/aoms/1177732360.

[97] G. Cowan, K. Cranmer, E. Gross, and O. Vitells. Asymptotic formulae for likelihood-based tests of new physics. *The European Physical Journal C*, 71(2), Feb 2011. doi:10.1140/epjc/s10052-011-1554-0.

[98] Clopper, C. J. and Pearson, E. S. The use of confidence or fiducial limits illustrated in the case of the Binomial. *Biometrika*, 26(4):404–413, 1934. doi:10.1093/biomet/26.4.404.

[99] P.D. Dauncey, M. Kenzie, N. Wardle, and G.J. Davies. Handling uncertainties in background shapes: the discrete profiling method. *Journal of Instrumentation*, 10(04):4015, Apr 2015. doi:10.1088/1748-0221/10/04/p04015.

[100] The CMS collaboration. Search for the lepton flavor violating decay $\tau \to 3\mu$ in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Journal of High Energy Physics*, 2021(1):163, 2021. doi:10.1007/JHEP01(2021)163.

[101] The ATLAS Collaboration, Belle II Collaboration, CMS Collaboration, and LHCb Collaboration. Projections for Key Measurements in Heavy Flavour Physics. 2025. arXiv:2503.24346, doi:10.48550/arXiv.2503.24346.

[102] CMS Collaboration. The Phase-2 Upgrade of the CMS Muon Detectors. Technical report, CERN, Geneva, 2017. This is the final version, approved by the LHCC. doi:10.17181/CERN.5T9S.VPMI.

[103] CMS Collaboration. The Phase-2 Upgrade of the CMS Level-1 Trigger. Technical report, CERN, Geneva, 2020. Final version. URL: https://cds.cern.ch/record/2714892.

[104] P. de Barbaro et al. The Phase-2 upgrade of the CMS barrel calorimeters. CMS Technical Proposal CERN-LHCC-2017-011, CMS-TDR-015, 2017. URL: https://cds.cern.ch/record/2283187.

[105] CMS Collaboration. The Phase-2 Upgrade of the CMS Endcap Calorimeter. Technical report, CERN, Geneva, 2017. doi:10.17181/CERN.IV8M.1JY2.

[106] CMS Electromagnetic Calorimeter Group. Performance of the front-end electronics of the CMS electromagnetic calorimeter barrel for the High-Luminosity LHC. *Journal of Instrumentation*, 21(01), Jan 2026. doi:10.1088/1748-0221/21/01/P01001.

[107] CMS Collaboration. The Phase-2 Upgrade of the CMS Barrel Calorimeters. Technical Report CERN-LHCC-2017-011. CMS-TDR-015, CERN, Geneva, Sep 2017. URL: https://cds.cern.ch/record/2283187.

[108] CMS collaboration. ECAL performance and ES alignment with early 2025 data. 2025. URL: https://cds.cern.ch/record/2938118.

[109] Olivier Gevin, P. Baron, M. Dejardin, and F. Guilloux. CATIA: APD readout ASIC for the CMS phase 2 ECAL electronics upgrade. In *Proceedings of Topical Workshop on Electronics for Particle Physics — PoS(TWEPP2019)*, volume 370, page 001, 2020. doi:10.22323/1.370.0001.

[110] G. Mazza et al. The LiTE-DTU: A Data Conversion and Compression ASIC for the Readout of the CMS Electromagnetic Calorimeter. *IEEE Transactions on Nuclear Science*, 70(6):1215, 2023. doi:10.1109/TNS.2023.3274930.

[111] J. Troska, F. Vasey, and V. Weidberg. Radiation tolerant optoelectronics for high energy physics. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 1052:168208, 2023. doi:10.1016/j.nima.2023.168208.

[112] Hernandez Montesinos et al. Overview of the production and qualification tests of the lpgbt. *Journal of Instrumentation*, 19(04):C04048, Apr 2024. doi:10.1088/1748-0221/19/04/C04048.

[113] N. Loukas. The Barrel Calorimeter Processor demonstrator board for the Phase II Upgrade of the CMS ECAL Barrel. Technical Report CMS-CR-2018-301, CERN, Geneva, 2018. URL: https://cds.cern.ch/record/2644903.

[114] D. Banerjee et al. The North Experimental Area at the Cern Super Proton Synchrotron. CERN Accelerator Note 2021/0015, 2021. doi:10.17181/CERN.GP3K.0S1Y.

[115] Yu. Barnyakov et al. Response of microchannel plates in ionization mode to single particles and electromagnetic showers. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 879, 2018. arXiv:1707.08503v1, doi:10.1016/j.nima.2017.10.002.

[116] A. C. Marini. H4DAQ: a modern and versatile data-acquisition package for calorimeter prototypes test-beams. *JINST*, 13(02):C02042, 2018. doi:10.1088/1748-0221/13/02/C02042.

[117] S. Ritt, R. Dinapoli, and U. Hartmann. Application of the drs chip for fast waveform digitizing. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 623(1):486, 2010. 1st International Conference on Technology and Instrumentation in Particle Physics. doi:10.1016/j.nima.2010.03.045.

[118] P. Adzic et al. Intercalibration of the barrel electromagnetic calorimeter of the CMS experiment at start-up. *JINST*, 3:P10007, 2008. doi:10.1088/1748-0221/3/10/P10007.

[119] P. Adzic et al. Energy resolution of the barrel of the CMS electromagnetic calorimeter. *JINST*, 2:P04004, 2007. doi:10.1088/1748-0221/2/04/P04004.