# Tau identification algorithms and study of the CP structure of the Yukawa coupling between the Higgs boson and tau leptons in CMS

**Dissertation**

zur Erlangung des Doktorgrades an der Fakultät

für Mathematik, Informatik und Naturwissenschaften

Fachbereich Physik

der Universität Hamburg

vorgelegt von

**Oleg Filatov**

Hamburg

2023

# Eidesstattliche Erklärung / Declaration on oath

Ich erkläre hiermit an Eides statt, dass ich diese Dissertation selbst verfasst und keine anderen als die angegebenen Hilfsmittel oder Quellen benutzt habe.

I hereby declare in lieu of oath that I have written this dissertation myself and that I have not used any auxiliary materials or sources other than those indicated.

Hamburg, March 13, 2023 Oleg Filatov

Unterschrift des Doktoranden

| | |
|---|---|
| Gutachter/innen der Dissertation: | Dr. Alexei Raspereza |
| | Prof. Dr. Elisabetta Gallo |
| Zusammensetzung der Prüfungskommission: | Prof. Dr. Sven-Olaf Moch |
| | Prof. Dr. Elisabetta Gallo |
| | Dr. Alexei Raspereza |
| | Jun.-Prof. Dr. Gregor Kasieczka |
| | Dr. Judith Katzy |
| Vorsitzende/r der Prüfungskommission: | Prof. Dr. Sven-Olaf Moch |
| Datum der Disputation: | 19.01.2023 |
| Vorsitzender des Fach-Promotionsausschusses PHYSIK: | Prof. Dr. Günter H. W. Sigl |
| Leiter des Fachbereichs PHYSIK: | Prof. Dr. Wolfgang J. Parak |
| Dekan der Fakultät MIN: | Prof. Dr.-Ing. Norbert Ritter |

# Abstract

The measurement of the CP properties of the Yukawa coupling of the Higgs boson to $\tau$ leptons is presented. The data set used for the analysis is collected by the CMS experiment at the LHC during the Run 2 data-taking period in proton-proton collisions at $\sqrt{s} = 13$ TeV and corresponds to an integrated luminosity of 137 fb$^{-1}$. The Yukawa coupling between the Higgs boson and $\tau$ leptons is parametrised in terms of the effective mixing angle $\alpha^{H\tau\tau}$, where the value $\alpha^{H\tau\tau} = 0°(90°)$ corresponds to the SM scenario of the pure CP-even (CP-odd) H$\tau\tau$ coupling.

The angle between the decay planes of the $\tau$ leptons is used as the observable encoding the CP nature of the Higgs boson. The measurement is performed in the $\tau_e\tau_h$ channel where one $\tau$ lepton decays into an electron and the other hadronically. The results are combined with the measurement in the $\tau_\mu\tau_h$ and $\tau_h\tau_h$ channels. The observed (expected) value of the effective mixing angle for the combination is measured to be:

$$\alpha^{H\tau\tau} = -1 \pm 19°(0 \pm 21°) \text{ @68.3\% CL.}$$

The results are compatible with the SM expectation and the pure CP-odd hypothesis is rejected at the observed (expected) significance level of 3.0 (2.6) standard deviations.

The improvements to the $\tau$ lepton identification in CMS in the context of the Run 3 preparation are described. Retraining and optimisation of the DeepTau algorithm with the addition of the adversarial fine-tuning procedure are performed. The resulting model improves upon the previous DeepTau model in terms of the background rejection by 10-50% and has a better description of data with simulation in the H $\rightarrow \tau\tau$ selection region.

A new algorithm called Tau Transformer (TaT) is proposed to overcome the limitations of the DeepTau architecture. The TaT core is based on self-attention layers and features the embedding module allowing for the multimodality treatment of the input representation. Comparison of the TaT model with the retrained DeepTau model and a comparable ParticleNet-based architecture shows consistently improved performance by up to 50% in the misidentification rate across the $p_T$, $\eta$, and decay mode ranges of interest.

# Zusammenfassung

Die Messung der CP-Eigenschaften der Yukawa-Kopplung des Higgs-Bosons an $\tau$-Leptonen wird vorgestellt. Der für die Analyse verwendete Datensatz wurde vom CMS-Experiment am LHC während der Datenerfassungsperiode des Run 2 in Proton-Proton-Kollisionen bei $\sqrt{s} = 13$ TeV aufgezeichnet und entspricht einer integrierten Luminosität von 137 fb$^{-1}$. Die Yukawa-Kopplung zwischen dem Higgs-Boson und $\tau$-Leptonen wird durch den effektiven Mischungswinkel $\alpha^{\mathrm{H}\tau\tau}$ parametrisiert, wobei der Wert $\alpha^{\mathrm{H}\tau\tau} = 0°(90°)$ dem SM-Szenario der reinen CP-geraden (CP-ungeraden) H$\tau\tau$-Kopplung entspricht.

Der Winkel zwischen den Zerfallsebenen der $\tau$-Leptonen wird als Beobachtungsgröße verwendet, deren Verteilung die CP-Natur des Higgs-Bosons widerspiegelt. Die Messung wird im $\tau_{\mathrm{e}}\tau_{\mathrm{h}}$-Zerfallskanal durchgeführt, bei dem ein $\tau$-Lepton in ein Elektron und das andere hadronisch zerfällt. Die Ergebnisse werden mit den Messungen in den Zerfallskanälen $\tau_\mu\tau_{\mathrm{h}}$ und $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$ kombiniert. Der beobachtete (erwartete) Wert des effektiven Mischungswinkels für die Kombination wird wie folgt gemessen:

$$\alpha^{\mathrm{H}\tau\tau} = -1 \pm 19°(0 \pm 21°) \ @68.3\% \ \mathrm{CL}.$$

Die Ergebnisse sind mit der SM-Erwartung vereinbar, und die reine CP-ungerade-Hypothese wird auf dem beobachteten (erwarteten) Signifikanzniveau von 3.0 (2.6) Standardabweichungen zurückgewiesen.

Die Verbesserungen bei der Identifizierung von $\tau$-Leptonen in CMS im Zusammenhang mit der Vorbereitung von Run 3 werden beschrieben. Der DeepTau-Algorithmus wird unter Anwendung des Adversarial-Verfahrens neu trainiert und optimiert. Das sich daraus ergebende Modell verbessert das vorherige DeepTau-Modell in Bezug auf die Unterdrückung des Untergrundprozesses um 10-50% und führt zu einer besseren Beschreibung der Daten durch die Simulation im H $\rightarrow \tau\tau$-Selektionsbereich.

Ein neuer Algorithmus namens Tau Transformer (TaT) wird vorgeschlagen, um die Limitierungen der DeepTau-Architektur zu überwinden. Der TaT-Kern basiert auf Self-Attention Schichten und verfügt über ein Einbettungsmodul, das die multimodale Behandlung der Eingabedarstellung ermöglicht. Der Vergleich des TaT-Modells mit dem neu trainierten DeepTau-Modell und einer vergleichbaren ParticleNet-basierten Architektur zeigt eine durchgängig verbesserte Leistung von bis zu 50% bei der Fehlerkennungsrate in den interessierenden $p_{\mathrm{T}}$- und $\eta$-Bereichen sowie $\tau_{\mathrm{h}}$-Zerfallmoden.

*CERCA TROVA*

# Contents

# 1 | Introduction

One of the most fundamental notions in the description of nature is the one of matter. It can be thought of as a substance filling the space and transforming in time. One can wonder why the objects constituting matter are very heterogeneous in properties as defined by one's perception? It turned out to be useful to spatially zoom in into the structure of matter and define the most fundamental objects serving as building blocks. These were called particles and it was the difference in their types and interaction which explained the diversity of matter representations and related phenomena.

However, despite the particle zoo has been continuously expanding, it turned out that there are matter representations which cannot be described as being made of already known particles. An example would be dark matter which is supposed to fill the universe in order to explain various astrophysical observation, e.g. galaxy rotation curves. Furthermore, as antimatter was discovered, one would imagine that there must be an equal amount of matter and antimatter in the universe at any given point in time – a statement which is rooted in another important notion of symmetry as the ruling principle of nature. However, the matter which is observable from the Earth now does not have sizeable fractions of antimatter. This raises the question whether this asymmetry was generated at some point in time, and if yes, then in which way.

Assuming that there was indeed a phase in the evolution of the universe when the matter-antimatter asymmetry was generated (so-called baryogenesis), Andrei Sakharov proposed three conditions which necessary have to be satisfied for the particle interactions to create the asymmetry [1]:

- Baryon number violation,
- C-symmetry and CP-symmetry violation,
- Absence of thermal equilibrium.

The second item in the list corresponds to the violation of the charge (C) and combined charge-parity (CP) symmetries. The charge (parity) symmetry states that a given process occurs

in the same way if the particle charges (spatial coordinates) are flipped.

Therefore, in order to establish whether these conditions hold true in nature, one should search for potential sources of the CP violation (CPV). This search has to be necessarily performed within some theoretical framework which describes the processes occurring at the most fundamental level of matter. This framework has been built over decades and it is called the Standard Model (SM). Chapter 2 provides a brief overview of its essential fundamentals required to pursue the dedicated search for CPV.

One important property which essentially defines matter is mass. It was known experimentally that the discovered particles have it, but before 1960s it was not clear how to introduce it theoretically into the SM framework. The Brout-Englert-Higgs-Hagen-Guralnik-Kibble (BEH) mechanism was one of the proposed solutions, as described in Sec. 2.2.3. However, it took almost 50 years to verify this solution experimentally by observing a particle similar in properties to the Higgs boson – a scalar particle of the Higgs field playing a key role in the proposed mechanism. To date, multiple Higgs boson properties are measured with a good level of precision, as summarised in Sec. 2.3, to give a high confidence that the observed particle is indeed the one predicted by the BEH mechanism. The CP properties are also the important ones to be established as they can provide hints to the sources of CPV in the theory and the matter-antimatter asymmetry problem of the Universe. This is the main motivation behind this work, which is dedicated to the investigation of the CP structure of the interaction between the Higgs boson and tau leptons.

However, to pursue this study one firstly needs to be able to produce Higgs bosons in the laboratory settings. For that purpose, one builds accelerators where particles are brought to the speed close to the speed of light and then collided as the way to make them interact. Chapter 3 introduces the Large Hadron Collider as the world's largest particle accelerator (Sec. 3.1) which can currently be viewed as the only "Higgs factory". The Compact Muon Solenoid (CMS) detector (Sec. 3.2) is nowadays one of the main instruments to get insights into the interaction of the Higgs boson with the other particles. Surrounding the interaction point, it aims to fully capture and observe emerging particles, thus allowing for the full reconstruction of the Higgs boson decay.

As it was mentioned, interaction of the Higgs boson with the tau leptons is the primary focus of this work. The interaction is probed in the Higgs decay, which therefore requires precise reconstruction and identification of tau leptons. The corresponding methods developed and applied in the CMS collaboration are described in Chapter 4. Firstly, the most fundamental Particle Flow (PF) algorithm (Sec. 4.2.1) is used to reconstruct basic particles. The hadron-plus-strips (HPS) algorithm (Sec. 4.2.2) is then used specifically for the tau lepton reconstruction from the PF-derived particles. An identification step usually follows the reconstruction, and in CMS it is performed using the DeepTau algorithm. In this work, version 2.1 (v2.1) of this model was used (Sec. 4.3.1), but several improvements were introduced, referred to as version

2.5 (v2.5), in order to prepare the model for the new Run 3 data-taking period (Sec. 4.3.2). Furthermore, the DeepTau architecture is known to have several intrinsic limitations. New approaches based on the graph- and attention-based models were investigated to further improve the performance of the tau identification step (Sec. 4.3.3).

It is Chapter 5 which describes the analysis of CP properties of the Higgs interaction with the tau leptons, as motivated above. Since there is a multiplicity of options for the tau lepton to decay, this work focuses on the reconstruction of the Higgs boson decaying into a pair of tau leptons, where one lepton decays into the final state with a single electron and neutrinos, while the other lepton decays into hadrons and neutrino ($\tau_e\tau_h$ final state). The framework to quantify the CP effects of interest, the physical observable, and the analysis strategy are outlined in Sec. 5.1 and Sec. 5.3.7, while the other sections cover the other essential steps of the analysis. The measurement performed in the $\tau_e\tau_h$ final state (Sec. 5.8) is combined with the measurement of the $\tau_\mu\tau_h$ and $\tau_h\tau_h$ final states, as described in Chapter 6. Finally, the summary and conclusions are given in Chapter 7.

## Personal contribution

The author of this work contributed to all the stages of the analysis of the $\tau_e\tau_h$ final state. This includes addition of the electron objects to the analysis framework with the necessary corrections, validation of the simulation agreement with the data, training of the neural networks used for event categorisation, addition of systematic uncertainties to the analysis framework and their validation, performing statistical inference, including goodness-of-fit tests and final result extraction.

In the tau lepton identification studies, on the side of DeepTau v2.5, the author contributed to the improvement of the scalability of the framework, architecture optimisation, development of the feature preprocessing and performance evaluation modules, and integration into the CMS software. On the side of Tau Transformer, the contributions include the conceptual design of the model and its implementation, development of the data loading pipeline, training and hyperparameter tuning, performance evaluation, and ablation studies.

# 2 | The Standard Model

It is an inherent part of human nature to wonder about phenomena happening in the world and to ask questions about their origin. It seems to be driven by inexplicable curiosity to understand, almost a demand of the reason to build logical structures, through centuries adding one brick after another to the temple of the humankind knowledge.

Since people discovered the notion of *space* and *time* and formalised it, this became one of the most important frames of human reasoning. Another important notion to be realised and formulated was the one of *matter*, a medium which fills the space, evolves in time, and sensory possesses heterogeneous properties. At this point it was natural to start asking questions about the *structure* of matter, trying to split it into logical categories and find its place in some larger framework of reasoning. Starting from things which one can observe and experience in daily life, the reasoning went farther away to smaller or larger scales in space – a natural frame of reference for reason to explore. It is probably this narrative which contributed to the emergence of *particle physics*, a field of science describing phenomena at the very smallest scale.

Most of the scientific experiments in natural sciences deal with *interacting* objects, which in turn are essentially various representations of matter. These observed phenomena had to be necessarily explained using some toolkit with *fundamental* blocks and categories which matter is made of. Step-by-step, this necessity led to the discovery of *particles* – and in fact a plethora of them. All the visible matter turned out to be composed of them organised in peculiar patterns. It became also possible to explain multiple phenomena in nature by formalising it in terms of dynamics, transformation, and interaction of particles.

However, it is not sufficient to only formulate particles as a set of logical entities. There is always a need of reason to bring structure into it, to organise them in a "beautiful" and "meaningful" way. Furthermore, the concept of interaction and evolution necessarily motivates the composition of rules guiding them. These rules take up the form of *laws*, which value is estimated not only by the precision to describe already explored phenomena but also by the power to predict yet undiscovered ones.

Bringing various pieces of the puzzle together, this is how the *Standard Model* (SM) has

been gradually built. It has always been expanding to incorporate newly discovered particles, interactions, and observed phenomena while also predicting new ones on the way. Admittedly, it is an "Absolutely Amazing Theory of Almost Everything" [2] as it describes the most fundamental blocks which constitute matter and the rules of its transformation. This Chapter gives a brief overview of its fundamental aspects. Sec. 2.1 introduces the particles which have been discovered so far. In Sec. 2.2 the theoretical foundation of SM is laid out. The discovery of the Higgs boson in 2012 [3, 4] was a great milestone bringing yet one more piece of the puzzle into the Standard Model. The corresponding motivation behind its importance is described in Sec. 2.2.3 and the properties as predicted theoretically and measured experimentally are outlined in Sec. 2.3.

## 2.1 Particle content

A particle, being a small localised in space object with certain properties, is called elementary (or fundamental) if there is no other particles which it is composed of. Three important particle properties should be mentioned as the starting point:

- **Spin** – a quantum number which might be informally associated to the internal angular momentum of the particle. Particles taking half-integer spin values are called *fermions*, and particles with an integer spin are called *bosons*.

- **Electric charge** – a quantum number which defines the behaviour of the particle in the electromagnetic field. Taking the positive/negative integer values, it categorises particle to be charged (charge does not equal to 0) or neutral (charge equals to 0).

- **Mass** – an intrinsic property of a body which measures its inertia to acceleration by a net force. In this work, the mass is measured in electronvolt (eV) units, where $1\text{eV}/c^2 = 1.78266192 \times 10^{-36}$ kg and $c = 299792458\ m/s$ is a speed of light. Natural units are used throughout this work which set $c = h/(2\pi) = 1$, where $h$ is the Planck constant.

The Standard Model describes the following elementary fermions and bosons (Fig. 2.1):

- **Fermions:** 6 quarks of different flavour ($q = u, d, c, s, t, b$), 3 charged leptons ($l = e, \mu, \tau$), 3 neutral leptons ($\nu_l = \nu_e, \nu_\mu, \nu_\tau$).

- **Bosons:** 8 gluons of different colour ($g$), $W^{\pm}/Z$ particles, photon ($\gamma$), the Higgs boson ($H$).

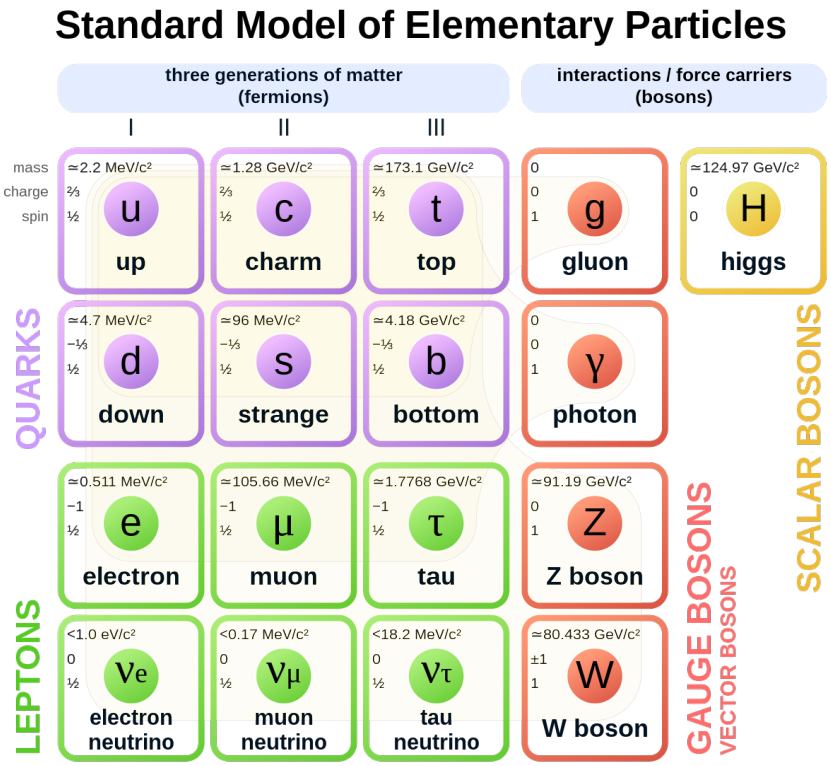


Figure 2.1: Elementary particles in the Standard Model and their properties [5].

Each particle is also accompanied by an *antiparticle* which either coincides with the particle itself or represent a new particle. In the latter case it has the same mass and spin but the opposite electric charge (if different from 0). Fermions can be further categorised into *quarks* and *leptons*. The former can take part in the strong interaction (Sec. 2.2) while the latter cannot. They are furthermore combined into three generations (three left columns on Fig. 2.1).

The bosons are considered to be *force carriers*, i.e. particles which mediate the interactions between other particles and hence guide the evolution of particle systems. The only exception is the Higgs boson, which is responsible for the generation of particle masses. At the moment, four fundamental forces are discovered: electromagnetic, weak, strong, gravitational. The former three are known to have the corresponding boson mediators. The mediator of the gravitational force still remains a mystery to be understood.

The notion of interaction and force is at the very core of the Standard Model, and its formalisation is provided in the next Section.

## 2.2 Formulation

### 2.2.1 Introduction

The Standard Model is a relativistic Quantum Field Theory (QFT) describing dynamics of particles moving at the speed close to the speed of light. It is built upon the notion of *symmetry*, which in this particular case refers to the invariance of physical processes under a certain type of transformations. The transformations are formalised using the language of group theory. The latter introduces a mathematical concept of a *group* – a set of elements $G$ and an operation on this set which maps two elements of the set to another element in the set. The group should also satisfy three axioms: associativity, existence of the identify element, and existence of the inverse element.

For example, one of the most fundamental symmetries which the Standard model should satisfy is the *Lorentz symmetry*. In particular, the theory should be build in a Lorentz-covariant way, so that physical quantities transform accordingly under a given representation of the Lorentz group (indefinite orthogonal group O(1,3)) – i.e. as quantities composed of scalars, four-vectors, four-tensors, and spinors. This allows to properly account for the relativistic nature of particles in a consistent way with the special theory of relativity.

The main part of the SM is a *Lagrangian density L* which is the general starting point to derive laws describing the evolution of a system. If classical Lagrangian mechanics uses Lagrangian to derive equations of motion for a system with finite number of degrees of freedom, field theory extends the approach to *fields* having infinite number of degrees of freedom.

Equations of motion can be derived via the *action principle*. One defines an action functional as:

$$A[\phi(x)] = \int d^4x \; L(\phi(x), \partial_\mu \phi), \tag{2.1}$$

where the integral is taken over the space-time coordinates $x \equiv x_\mu = (t, \vec{x})$ with a Minkowski metric $g_{\mu\nu} = \mathrm{diag}(1, -1, -1, -1)$ and, for the sake of illustration, the Lagrangian density is written for a scalar field $\phi(x)$. Requiring the variation of the functional for every infinitesimal shift of the field $\delta\phi(x)$ to be 0:

$$A[\phi(x) + \delta\phi(x)] - A[\phi(x)] = \int d^4x \left[ \left( \partial_\mu \frac{\partial L}{\partial \partial_\mu \phi} - \frac{\partial L}{\partial \phi} \right) \delta\phi + \partial_\mu \left( \frac{\partial L}{\partial \partial_\mu \phi} \right) \right] = 0, \tag{2.2}$$

one can obtain the equation of motion for the field $\phi(x)$ by plugging-in the known Lagrangian density of the field system.

A simple example would be the Lagrangian density of the free Dirac field describing the motion of a spin-1/2 particle with a mass $m$:

$$L_0 = \bar{\psi}(i\hat{\partial} - m)\psi, \tag{2.3}$$

where $\hat{\partial} \equiv \gamma_\mu \partial^\mu$, $\gamma_\mu$ are the gamma matrices, $\psi(x)$ is a Dirac spinor, $\bar{\psi}(x) \equiv \psi^\dagger(x)\gamma_0$ is the Dirac-conjugated spinor. Using Eq. 2.2 one can obtain the famous Dirac equation:

$$(i\hat{\partial} - m)\psi(x) = 0. \tag{2.4}$$

One can notice that the Dirac Lagrangian density (2.3) is invariant under the *global* transformation of the U(1) group which acts on the spinor $\psi(x)$ as:

$$\psi(x) \mapsto \psi'(x) = e^{ie\omega}\psi(x), \tag{2.5}$$

where $\omega$ is a constant. However, if one assumes a *local* U(1) symmetry $\omega \to \omega(x)$, the Lagrangian is no longer invariant. This symmetry can be restored if an additional interaction of the fermion with a photon field $A_\mu$ is introduced by promoting the usual derivative to a *covariant* one:

$$\partial_\mu \to D_\mu \equiv \partial_\mu - ieA_\mu, \tag{2.6}$$

and defining the transformation of the photon field as $A_\mu \to A'_\mu = A_\mu - \partial_\mu\omega$. The Lagrangian which is invariant under the local U(1) symmetry corresponds to the Lagrangian of Quantum Electrodynamics (QED):

$$L_{\text{QED}} = \bar{\psi}(i\hat{D} - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}, \tag{2.7}$$

where a field-strength tensor $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is also introduced.

If there was no mass term in the Dirac Lagrangian (2.3), then a *chiral symmetry* would also be conserved. Left and right *chiral spinors* are defined as eigenvalues of the $\gamma_5 \equiv i\gamma_0\gamma_1\gamma_2\gamma_3$ operator:

$$\gamma_5\psi_{L/R} = \mp\psi_{L/R}. \tag{2.8}$$

Every spinor decomposes into a sum of the left and right spinors:

$$\psi = \psi_L + \psi_R,\ \psi_{L/R} = \frac{1 \mp \gamma_5}{2}\psi, \tag{2.9}$$

and the Lagrangian (2.3) reads:

$$L_0 = i(\bar{\psi}_L\hat{\partial}\psi_L + \bar{\psi}_R\hat{\partial}\psi_R) - m(\bar{\psi}_L\psi_R + \bar{\psi}_R\psi_L). \tag{2.10}$$

If one considers a transformation $\psi \rightarrow \psi' = e^{i\gamma_5\omega}$ associated with the chiral symmetry, the second term (mass term) in Eq. 2.10 violates this symmetry.

### 2.2.2 Lagrangian

#### Electroweak interaction

The Lagrangian (2.7) was the first theory which combined special relativity with quantum mechanics in a consistent way to describe the observed electromagnetic phenomena. But while it produces very accurate predictions for such quantities as the anomalous magnetic moment of the electron, it does not incorporate such phenomena as, for example, the $\beta$ decay of the neutron.

Fermi's interaction turned out to be a good explanation of the $\beta$ decay proceeding via a *weak interaction*. However, this was still a standalone piece of the puzzle which was not related to the electromagnetic phenomena. More importantly, it had a fundamental problem of *unitarity violation* – the cross section of the reaction was predicted to increase linearly with the energy. This was hinting that the Fermi's theory was an effective theory valid only in the limit of low energies. A more fundamental theory was yet to be formulated.

It was Sheldon Glashow, Abdus Salam, and Steven Weinberg who proposed a theory which would describe in a unified way electromagnetic and weak interactions. It is build on the assumption that the Lagrangian should be invariant under the transformation of the group:

$$\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y. \tag{2.11}$$

The underlying symmetry (also referred to as *gauge* symmetry) is composed of two gauge groups. The *weak-isospin* group $\mathrm{SU}(2)_L$ comes with three gauge fields $W^a_\mu$, $a = 1, 2, 3$. The *weak-hypercharge* group $\mathrm{U}(1)_Y$ comes with one gauge field $B_\mu$, analogously to the QED case with the $A_\mu$ field.

The fields $W^a_\mu$ are allowed to interact only with the left-handed fermions. This behaviour is introduced in order to account for the $(V - A)$ pattern of the Fermi's interaction. The latter is introduced in order to describe the experimental observation of the *parity violation* in the Wu experiment [6]. Parity is another fundamental transformation which corresponds to the inversion of the spatial coordinates $x_\mu = (t, \vec{x}) \mapsto (t, -\vec{x})$. The $(V - A)$ pattern effectively models each of the interacting fermion currents in the Fermi's theory as a difference between the vector $j^V_\mu = \bar{\psi}\gamma_\mu\psi$ and axial $j^A_\mu = \bar{\psi}\gamma_\mu\gamma_5\psi$ currents, where the former is conserved and the latter is not.

The covariant derivatives which act on the left and right fermion fields are:

$$D_\mu^L = \partial_\mu - ig\frac{T_a}{2}W_\mu^a - ig'\frac{Y_L^f}{2}B_\mu, \tag{2.12}$$

$$D_\mu^R = \partial_\mu - ig'\frac{Y_R^f}{2}B_\mu. \tag{2.13}$$

Here $g$ and $g'$ are the gauge couplings, $T_a = \frac{\sigma_a}{2}$ are the generators of the $\mathrm{SU}(2)_L$ group with $\sigma_a$ being the Pauli matrices, $Y_{L/R}^f$ are the hypercharges introduced by the $\mathrm{U}(1)_Y$ group.

As it was mentioned in Sec. 2.2.1, in the Standard Model one needs to incorporate in total 6 quark and 6 lepton fields. In the example of one generation, the corresponding spinors are grouped into doublets and singlets under the $\mathrm{SU}(2)_L$ group:

$$Q_j \equiv \begin{pmatrix} u_L \\ d_L \end{pmatrix}_j,\ u \in \{u, c, t\},\ d \in \{d, s, b\} \tag{2.14}$$

$$L_j \equiv \begin{pmatrix} \nu_L \\ l_L \end{pmatrix}_j, \nu \in \{\nu_e, \nu_\mu, \nu_\tau\},\ l \in \{e, \mu, \tau\} \tag{2.15}$$

$$u_j \equiv (u_R)_j \tag{2.16}$$

$$d_j \equiv (d_R)_j \tag{2.17}$$

$$\tag{2.18}$$

The fermion part of the electroweak (EWK) Lagrangian then can be written by summing the terms of the Dirac Lagrangian (2.3) across the fermion generations $j = 1, 2, 3$ while taking into account the $\mathrm{SU}(2)_L$ singlet/doublet structure and the corresponding covariant derivatives:

$$L_f^{\mathrm{EWK}} = \sum_j i\bar{Q}_j\hat{D}^L Q_j + i\bar{u}_j\hat{D}^R u_j + i\bar{d}_j\hat{D}^R d_j + i\bar{L}_j\hat{D}^L L_j + i\bar{l}_j\hat{D}^R l_j. \tag{2.19}$$

Furthermore, kinetic terms for the $W_\mu$ and $B_\mu$ fields should also be included into the EWK Lagrangian:

$$L_k^{\mathrm{EWK}} = -\frac{1}{4}W_a^{\mu\nu}W_{\mu\nu}^a - \frac{1}{4}B^{\mu\nu}B_{\mu\nu} \tag{2.20}$$

$$W_{\mu\nu}^a = \partial_\mu W_\nu^a - \partial_\nu W_\mu^a + g f^{abc} W_\mu^b W_\nu^c \tag{2.21}$$

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu, \tag{2.22}$$

where $f^{abc}$ are the structure constants of the SU(2) group.

One can notice that Eq. 2.19 does not contain terms with the right-handed neutrinos since experimental results indicate that the observed neutrinos are left-handed. Furthermore, there is no fermion and vector boson mass terms included since they violate $SU(2)_L$ symmetry. This problem is solved by introducing the interaction with the Higgs field which generates these masses after the symmetry breaking mechanism (Sec. 2.2.3).

In order to account for QED, one needs to obtain terms where the photon field couples to the fermions. This is not explicit yet in the Lagrangian (2.19) and the desired interaction can be achieved by rotating the initial gauge fields $W^a_\mu$, $B_\mu$ to a physical basis:

$$W^\pm_\mu = \frac{1}{\sqrt{2}}(W^1_\mu \pm iW^2_\mu), \tag{2.23}$$

$$\begin{pmatrix} Z_\mu \\ A_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_W & -\sin\theta_W \\ \sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} W^3_\mu \\ B_\mu \end{pmatrix}. \tag{2.24}$$

Here the *Weinberg angle* $\theta_W$ is introduced. The requirement that $A_\mu$ in Eq. 2.23 corresponds to the photon field adds further constraints on the hypercharges $Y^f_L$. Noting also that $Y^u_L = Y^d_L \equiv Y^Q_L$, $Y^\nu_L = Y^l_L \equiv Y^L_L$ and fixing the normalisation $Y^Q_L = 1/3$ one can obtain the following relations between the electric charge $e$ and gauge couplings $g, g'$:

$$e = g\sin\theta_W, \tag{2.25}$$

$$e = g'\cos\theta_W, \tag{2.26}$$

and the following relation in terms of the eletromagnetic and $SU(2)_L \times U(1)_Y$ charge operators:

$$Q^f = (T^f_3 + \frac{Y^f}{2}), \tag{2.27}$$

where $T^f_3$ return $\pm 1/2$ for the left up/down fermions and 0 for the right fermions. Finally, the fermionic part of the Lagrangian takes the following form:

$$L^{\text{EWK}}_f = L_{CC} + L_{NC} \tag{2.28}$$

$$L_{CC} = \frac{g}{\sqrt{2}}(J^+_\mu W^{+\mu} + J^-_\mu W^{-\mu}), \quad J^+_\mu = \sum_f \bar{f}_u \gamma_\mu \frac{1-\gamma_5}{2} f_d \tag{2.29}$$

$$L_{NC} = eJ^A_\mu A^\mu + \frac{g}{\cos\theta_W} J^Z_\mu Z^\mu, \quad J^A_\mu = \sum_f Q^f \bar{f}\gamma_\mu f, \quad J^Z_\mu = \sum_f \bar{f}\gamma_\mu (v^f - a^f\gamma_5) f \tag{2.30}$$

$$v^f = \frac{T^f_3}{2} - Q^f \sin^2\theta_W, \quad a^f = \frac{T^f_3}{2} \tag{2.31}$$

One can observe that the (V-A) structure of the Fermi's interaction is present in the CC component of Eq. 2.28. Moreover, now the theory predict one new particle – the neutral $Z$ boson. It was a great success of the EWK theory when the Gargamelle experiment at CERN reported in 1973 the observation of the neutral current, a first indication of the $Z$ boson existence. The $W$ and $Z$ boson were directly observed at the Super Proton Synchrotron in 1983.

Later on, both $W^{\pm}$ and $Z$ boson properties were measured at the Large Electron Proton (LEP) Collider and at the Standford Linear Collider (SLAC). One of these properties is their mass, which is now measured to be $m_W = 80.377 \pm 0.012$ GeV, $m_Z = 91.1876 \pm 0.0021$ GeV [7]. However, similarly to fermions the EWK theory does not have explicit mass terms for the boson since they would violate the underlying gauge symmetry. The minimalistic approach of leaving them as massless particles in the theory would contradict the experimental observations. Furthermore, one can show that the unitarity principle is violated in the gauge boson scattering amplitudes, which hence leaves the theory with issues to be solved.

### Quantum Chromodynamics

One more piece of the puzzle which has not been mentioned so far is the theory describing the strong interaction. This theory is called *Quantum Chromodynamics* (QCD) and it describes the interaction between quarks. The motivation for a dedicated theoretical foundation is that experimentally the quarks have not been found in a free state contrary to leptons. Instead, they are confined within bound states of quark-antiquark pairs $q_i\bar{q}_j$ (called *mesons*), quark triplets $q_iq_jq_k$ (called *baryons*), tetraquarks $q_i\bar{q}_iQ_j\bar{Q}_j$, and pentaquarks $q_iq_jq_kq_l\bar{q}_m$, where $i, j, k, l, m$ refers to different quark flavours. Mesons and baryons are collectively called hadrons. These states are colour singlets, where the colour $C$ is a charge introduced by the $\mathrm{SU}(3)_C$ group. As of now, six quarks have been discovered: up ($u$), down ($d$), charm ($c$), strange ($s$), top ($t$), bottom ($b$).

Another consequence of the theory, similarly to the EWK case, was the appearance of the gauge bosons associated to the group – gluons. These are the mediators of the strong force which is responsible for formation of quark bound states. Experimentally, gluons were discovered at the electron-positron collider PETRA at DESY in three-jet events [8]. The jet is a cone of particles produced after the hadronisation (hadron formation) of a quark or a gluon. In this search, two jets would be initiated by two quarks appearing from the lepton annihilation, while one jet would be initiated by the gluon radiated from one of the quarks.

Theoretically, gluons are introduced by the covariant derivative corresponding to the $\mathrm{SU}(3)_C$ group:

$$D_\mu = \partial_\mu - ig_s\frac{\lambda_a}{2}G^a_\mu. \tag{2.32}$$

Here, $G^a_\mu$ are the gluon fields with the index $a = 1..8$, $\lambda_a$ are the Gell-Mann matrices, $g_s$ is the gauge constant of the $SU(3)_C$ group. These terms from the covariant derivative are then added to the covariant derivatives in the quark terms of the EWK Lagrangian (2.19). Lastly, a kinetic term for the gluons also has to be introduced:

$$L_k^{\text{QCD}} = -\frac{1}{4} G^a_{\mu\nu} G_a^{\mu\nu} \tag{2.33}$$

$$G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu + g_s f^{abc} G^b_\mu G^c_\nu, \tag{2.34}$$

where $f^{abc}$ are the structure constants of the $SU(3)_C$ group.

### 2.2.3 Spontaneous symmetry breaking

As discussed up to this point, the SM Lagrangian has the following form, invariant under the local $SU(2)_L \times U(1)_Y \times SU(3)_C$ gauge symmetry:

$$L_0^{\text{SM}} = L_f^{\text{EWK+QCD}} + L_k^{\text{EWK}} + L_k^{\text{QCD}} + L_{\text{Gauge-fixing}} + L_{\text{Ghosts}}, \tag{2.35}$$

where the terms corresponding to the gauge fixing and Fadeev-Popov ghosts are added as required by the theory quantisation and renormalisation procedures. The theory thus assumes that the fermions and the gauge bosons are massless, since explicit mass terms violate the underlying symmetries. This clearly contradicts the experimental observations, which poses a question of completeness of the SM in such formulation.

It was the Brout-Englert-Higgs-Hagen-Guralnik-Kibble mechanism [9–12] which was proposed in 1964 by the authors as the solution to the mass generation problem. It describes what is commonly referred to as *"spontaneous symmetry breaking"* and it proceeds as follows. A doublet $\Phi$ of complex scalar fields under the $SU(2)_L$ group – referred to as the *Higgs field* – is introduced:

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}, \tag{2.36}$$

with the corresponding Lagrangian consisting of the kinetic and potential terms:

$$L^{\text{Higgs}} = \partial_\mu \Phi^\dagger \partial^\mu \Phi - V(\Phi^\dagger \Phi), \tag{2.37}$$

where the potential reads:

$$V(\Phi^\dagger \Phi) = \mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2. \tag{2.38}$$

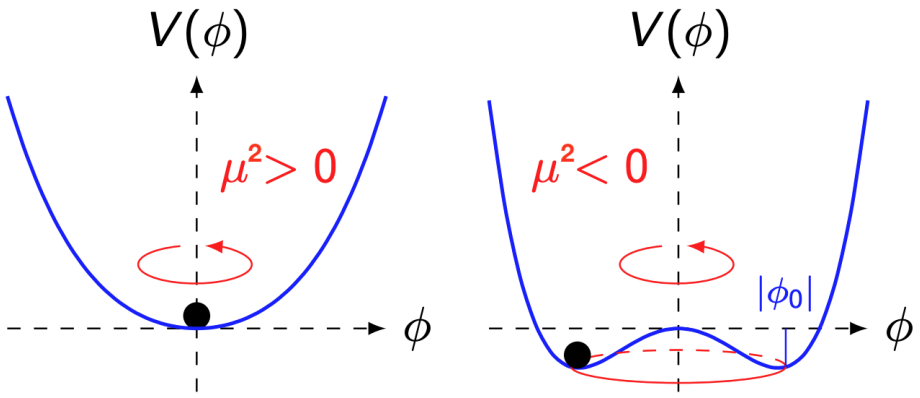


Figure 2.2: Illustration of the Higgs potential (Eq. 2.38) in case of a single complex scalar field $\phi$ for two scenarios: $\mu^2 > 0$ (left) and $\mu^2 < 0$ (right) [2].

Considering a simpler example of only one complex scalar field $\phi$ instead of a $SU(2)_L$ doublet $\Phi$, one can distinguish either a scenario $\mu^2 > 0$ with a single minimum of the potential at $\phi_0 = 0$ or a scenario $\mu^2 > 0$ with a valley of degenerate minima at $\phi_0 \neq 0$ (Fig. 2.2). In the latter case the non-trivial minima correspond to the value of the field:

$$\phi_0 \equiv \langle 0|\phi(x)|0\rangle = \frac{v}{\sqrt{2}} e^{i\beta}, \quad v = \frac{\mu}{\sqrt{\lambda}}. \tag{2.39}$$

The value $\phi_0$ is interpreted as the *vacuum expectation value*, a characteristic of the vacuum state. One observes that $\phi_0$ transforms under U(1) symmetry, however this symmetry is *implicitly broken* when one selects a particular minimum:

$$\phi_0 \overset{\beta=0}{=} \frac{v}{\sqrt{2}}. \tag{2.40}$$

Considering again the $\Phi$ doublet, one can parametrise the Higgs field around the vacuum expectation value similarly to Eq. 2.39 as:

$$\Phi(x) = \frac{1}{\sqrt{2}} \exp\left(i\frac{\zeta_j(x)\sigma^j}{2v}\right)\begin{pmatrix}0\\ v + h(x)\end{pmatrix}, \quad \Phi_0 \equiv \langle 0|\Phi(x)|0\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\ v\end{pmatrix}. \tag{2.41}$$

Here, $\zeta_j$ fields are the Nambu-Goldstone bosons, massless scalar particles necessarily appearing in theories where a continuous symmetry is spontaneously broken, i.e. the ground state is not invariant under the action of the underlying group. The physical state $h$ corresponds to the *Higgs boson*.

In order to link the Lagrangian (2.37) with the SM, one firstly introduces the covariant derivative $\partial_\mu \rightarrow D_\mu$ subject to the $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ group of the SM. This adds interaction between the Higgs field and the gauge bosons. One can further show that after the symmetry breaking (2.41) $W^\pm$ and $Z$ fields can be redefined to include $\partial_\mu \zeta_j$ terms. This is interpreted as the "absorption" of the Goldstone bosons resulting in the longitudinal polarisation of the $W^\pm$ and $Z$ bosons.

Moreover, the following terms appear after the symmetry breaking (2.41):

$$L^{\mathrm{SM}} \supset M_W^2 W_\mu^+ W^{\mu-}, \quad M_W^2 = \frac{g^2 v^2}{4} \tag{2.42}$$

$$L^{\mathrm{SM}} \supset \frac{1}{2} M_Z^2 Z_\mu Z^\mu, \quad M_Z^2 = \frac{(g+g')^2 v^2}{4}. \tag{2.43}$$

These are naturally interpreted as the mass term for the $W^\pm$ and $Z$ gauge bosons. The requirement that the photon has to remain massless leads to the following condition:

$$g \sin\theta_W - g' \cos\theta_W = 0, \tag{2.44}$$

and the mass of the Higgs boson $h$ itself from the corresponding $m_H hh$ term is found to be:

$$m_H = \sqrt{2\lambda} v. \tag{2.45}$$

Given that in the limit of small energies one should obtain Fermi's interaction, the vacuum expectation value can be directly linked to the Fermi constant:

$$\frac{G_F}{\sqrt{2}} = \frac{g^2}{8 M_W^2} \Rightarrow v^2 = \frac{1}{\sqrt{2} G_F} \simeq 246 \text{ GeV}. \tag{2.46}$$

One therefore concludes that the spontaneous symmetry breaking mechanism $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y \rightarrow \mathrm{U}(1)_{\mathrm{em}}$ and the added Higgs potential $\mathrm{V}(\Phi^\dagger \Phi)$ solve the issue with the generation of the gauge boson mass terms. One can also show that, due to the terms with trilinear and quartic coupling between $h$ and $W^\pm/Z$, the unitarity violation in the scattering amplitudes vanishes. Finally, it predicts the existence of the new scalar Higgs boson. A particle which so far meets the required properties to a very good precision was discovered in 2012 by the CMS and ATLAS experiments, as described in more detail in Sec. 2.3.

There is one more aspect which has not been covered yet – the mass generation for the fermions. While for neutrinos the process of acquiring mass is still not solved, for the other fermions it proceeds with the addition of *Yukawa terms*, preserving the $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ symmetry:

$$L^{\mathrm{Yukawa}} = -\sum_{ij} \left( Y_d^{ij} \bar{Q}^i \Phi d^j + Y_u^{ij} \bar{Q}^i \Phi^c u^j + Y_l^{ij} \bar{L}^i \Phi l^j \right) + h.c. \tag{2.47}$$

The notation here is identical to the one introduced in (2.14) and the sum is taken over the three fermion generations. The charge conjugate of the Higgs field is defined as $\Phi^c = i\sigma_2\Phi^*$. $Y_u^{ij}$, $Y_d^{ij}$, and $Y_l^{ij}$ are the general complex $3 \times 3$ matrices. The expression (2.47) is written in the flavour basis and one can make the corresponding transformation to the mass basis by rotating the fields with a unitary matrix:

$$Q \to V_Q Q, \quad u \to V_u u, \quad d \to V_d d, \tag{2.48}$$

so that:

$$Y_d = \mathrm{diag}(y_d, y_s, y_b), \quad Y_u = V_{\mathrm{CKM}}^{\dagger}\mathrm{diag}(y_u, y_c, y_t), \tag{2.49}$$

where $V_{\mathrm{CKM}}$ is the Cabibbo-Kobayashi-Maskawa (CKM) matrix. After performing this diagonalisation and the symmetry breaking (2.41) one obtains the following terms describing the quark interaction with the gauge bosons and the Higgs field:

$$L^{\mathrm{SM}} \supset L_{NC} + \frac{g}{\sqrt{2}}\bar{u}_L^i \hat{W}^+ \frac{1-\gamma_5}{2} V_{\mathrm{CKM}}^{ij} d_L^j + \frac{y_{u_i} v}{\sqrt{2}} \bar{u}_L^i u_R^i \left(1 + \frac{h}{v}\right) + \frac{y_{d_i} v}{\sqrt{2}} \bar{d}_L^i d_R^i \left(1 + \frac{h}{v}\right) + h.c., \tag{2.50}$$

where $L_{NC}$ is defined in Eq. 2.28. One can observe that now quarks acquired the mass:

$$m_{u_i/d_i} = \frac{y_{u_i/d_i} v}{\sqrt{2}}, \tag{2.51}$$

and they couple to the Higgs boson with a strength which is linearly proportional to the quark mass. The CKM matrix further modifies the interaction of the quarks with the $W$ boson compared to (2.28) by accounting for the mixing across the generations.

It should be noted that the same diagonalisation from the flavour to the mass basis can be performed in the lepton sector, yielding the same mass generation $m_{l_i} = \dfrac{y_{l_i} v}{\sqrt{2}}$ and Higgs coupling structure as for the quarks. In the basis where the Yukawa couplings of the charged leptons are diagonal, flavour states of neutrinos are linked to the mass states by the rotation with the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix. The PMNS matrix enters the charged current for the neutrino in place of the CKM matrix in the quark case. Furthermore, it plays an important role in the description of neutrino oscillations.

Lastly, it is the addition of the Yukawa terms (2.47) followed by their diagonalisation to the mass basis which introduces the violation of the *CP symmetry* into the SM. This is the symmetry under the combined charge conjugation and parity transformations. The former corresponds to the change of particle to their corresponding antiparticles, while the latter corresponds to the change of the sign of the spatial coordinates. These discrete symmetries are not explicitly introduced into the theory as the foundation, but are observed in nature and also implicitly

manifest itself in the SM Lagrangian. Due to the fact that there are exactly three generations of fermions, $Y_{ij}^{*} \neq Y_{ij}$ and there is one complex phase parameter in the CKM matrix and, in case of the Dirac nature of neutrinos, also in the PMNS matrix. In nature, this manifests itself as direct/indirect CP violation, for example, in the kaon [13], B-meson [14], and D-meson sectors [15].

This work continues the investigation of the Yukawa coupling sector from the perspective of searching CP anomalous effects. In particular, the structure of the Yukawa coupling between tau leptons and the Higgs boson is experimentally studied, as described in Sec. 5.

## 2.3 Higgs boson properties

In 2012 the CMS and ATLAS experiments at the Large Hadron Collider (Sec. 3.1) announced the observation of a new particle at a mass of 125 GeV [3,4]. In the CMS experiment the search was performed in $\gamma\gamma$, $ZZ$, $WW$, $\tau\tau$, and $b\bar{b}$ final states using the proton-proton collision at the center-of-mass energy $\sqrt{s}$ = 7 and 8 TeV. A local observed significance of the signal combined for all the decay modes was 5.0 standard deviations ($\sigma$). In the ATLAS experiment, using the same decay modes the local observed significance of 5.9$\sigma$ was reached.

Although the properties of the newly observed particle were reasonably consistent with the SM expectations of the Higgs boson, more studies had to be performed to better understand the nature of the new particle. Since 10 years of its discovery, multiple measurements have been done in order to understand deeper its properties [16, 17].

From the SM perspective, the Higgs boson is predicted to be dominantly produced in the following processes (Fig. 2.3, top left panel):

- Gluon-gluon fusion (ggH),
- Vector boson fusion (VBF),
- Associated production with a W or Z boson (VH or Higgsstrahlung),
- Associated production with top (ttH) or bottom (bbH) quarks.

In the proton-proton collisions at $\sqrt{s}$ = 13 TeV, for the mass of the Higgs boson near 125 GeV the most dominant production mode is ggH with the predicted cross section $\sigma_{\mathrm{ggH}} = 48.3 \pm 2.4$ pb (Fig. 2.4, top left panel). This accounts for approximately 87% of the total predicted production cross section $\sigma_{\mathrm{tot}} = 55.4 \pm 2.6$ pb [18]. The next most important contributions come from the VBF mode (7%, $\sigma_{\mathrm{VBF}} = 3.8 \pm 0.8$ pb), the VH mode (4%, $\sigma_{\mathrm{WH}} = 1.36 \pm 0.03$

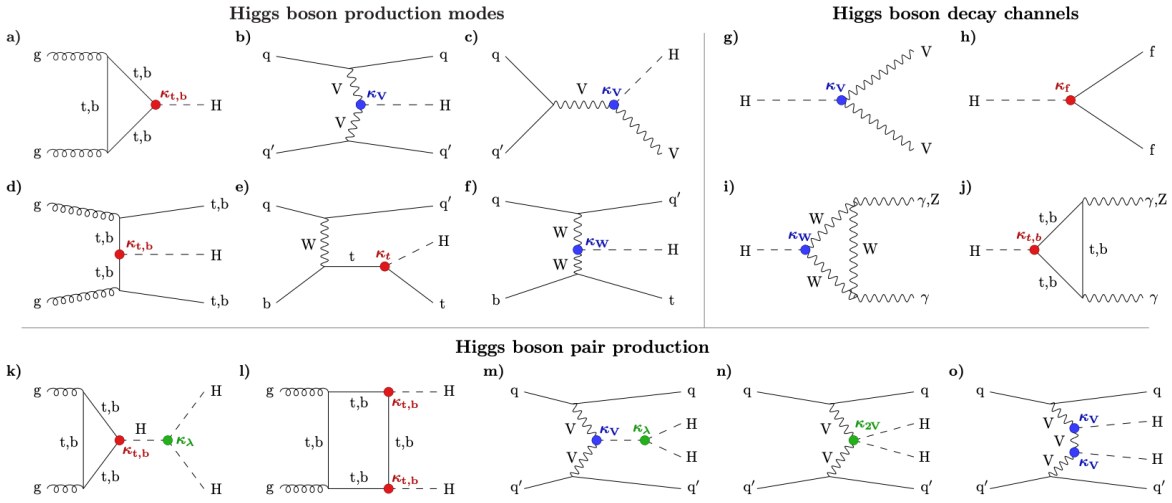


Figure 2.3: Feynman diagrams for the leading Higgs boson interactions: single Higgs boson production modes (top left), Higgs boson decay channels (top right), and Higgs boson pair production modes (bottom) [16].

pb, $\sigma_{\mathrm{ZH}} = 0.88 \pm 0.04$ pb). Experimentally, in CMS these values were confirmed to a large degree as measured by the signal strength modifiers $\mu_i$ (Fig. 2.4, bottom left panel). The latter corresponds to the ratio of experimentally observed signal yields to those predicted by the SM. The signal strength obtained after the fit to all the production modes and decay channels of all Run 2 data (Sec. 3.1) with a common single parameter results in $\mu = 1.002 \pm 0.057$, which is in excellent agreement with the SM prediction.

The mass of the Higgs boson is measured to be $m_{\mathrm{H}} = 125.25 \pm 0.17$ GeV as a combination of the CMS and ATLAS results in the most precisely reconstructed $\gamma\gamma$ and $ZZ \rightarrow 4l$ final states [7]. The width for the Higgs boson with such mass is predicted to be $\Gamma_{\mathrm{H}} = 4.14 \pm 0.02$ MeV [18]. Using off-mass-shell and on-mass-shell production, the CMS experiment measured the width to be $\Gamma_{\mathrm{H}} = 3.2^{+2.4}_{-1.7}$ MeV [20], which is in agreement with the SM expectation.

As discussed in Sec. 2.2.3, the Higgs field and the corresponding Higgs boson, appearing after the spontaneous symmetry breaking, couples to fermions (via the Yukawa terms) and gauge bosons $W^{\pm}/Z$ (via the covariant derivative terms). Therefore it is expected that the SM Higgs boson decays into the corresponding particle-antiparticle pairs as well as into the massless bosons (photons and gluons) via loops at the quantum level. Furthermore, the Higgs coupling strength is proportional to the mass (mass squared) of the fermion (vector boson). This sets the hierarchy of the expected decay probabilities with the third fermion generation being more preferred comparing to the second and first generations (Fig. 2.4, top right panel).

The Higgs boson decays which have been observed so far are:

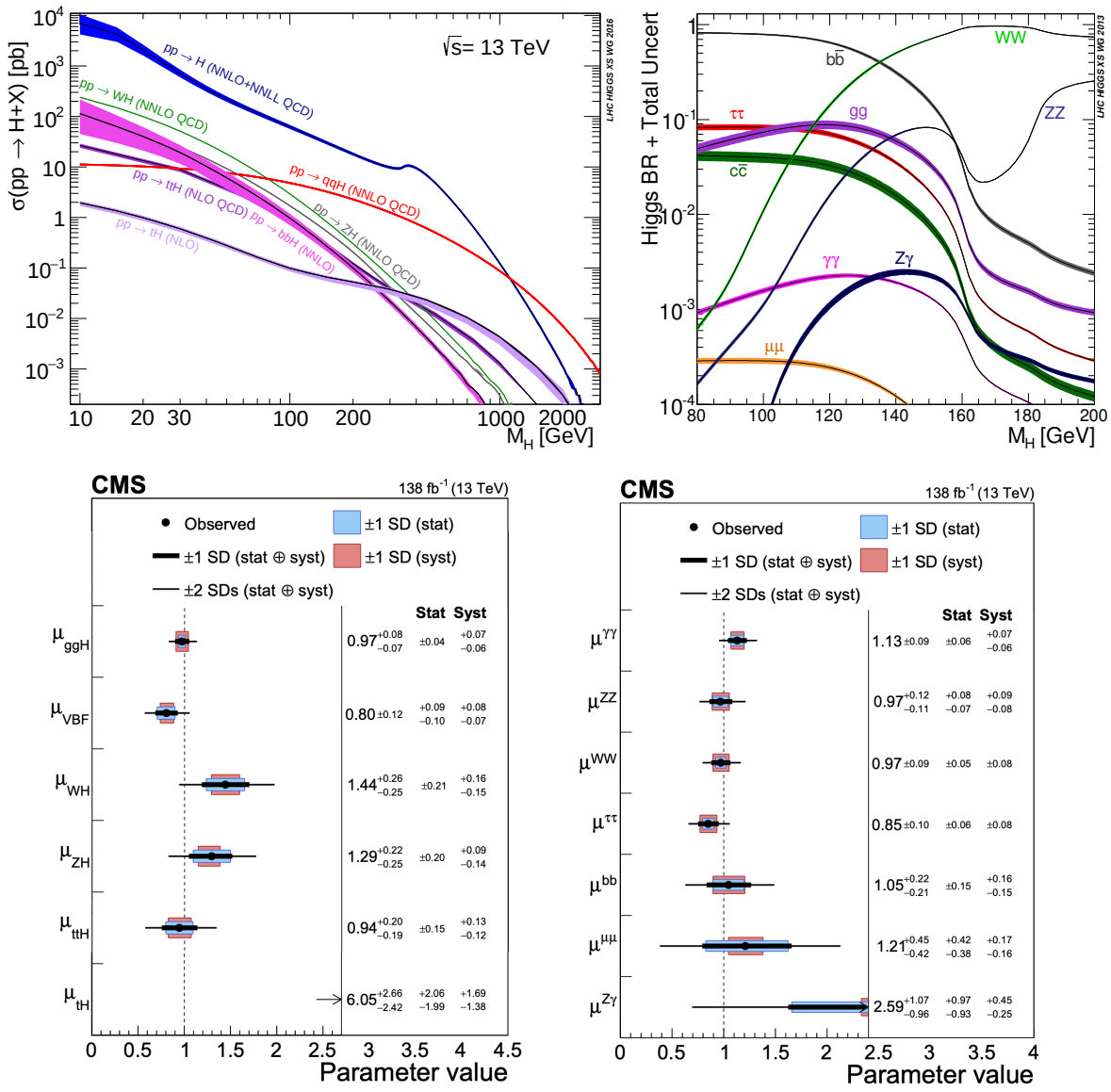


Figure 2.4: (Top left panel) Predicted Higgs boson production cross sections for various production modes as a function of the Higgs boson mass in proton-proton collisions at $\sqrt{s}$ = 13 TeV [18]. (Top right panel) Predicted Higgs boson branching fractions for various decay modes as a function of the Higgs boson mass in proton-proton collisions at $\sqrt{s}$ = 13 TeV [19]. (Bottom left panel) Signal strength modifiers for each of the Higgs production modes as measured by the CMS experiment [16]. (Bottom right panel) Signal strength modifiers for each of the Higgs decay modes as measured by the CMS experiment [16].

- Bosons: $\mathrm{H} \to \gamma\gamma$, $\mathrm{H} \to ZZ$, $\mathrm{H} \to WW$.

- 3rd generation fermions: $H \to \tau\tau$, $H \to bb$.
- 2rd generation fermions: $H \to \mu\mu$ (evidence, $3\sigma$ observed significance).

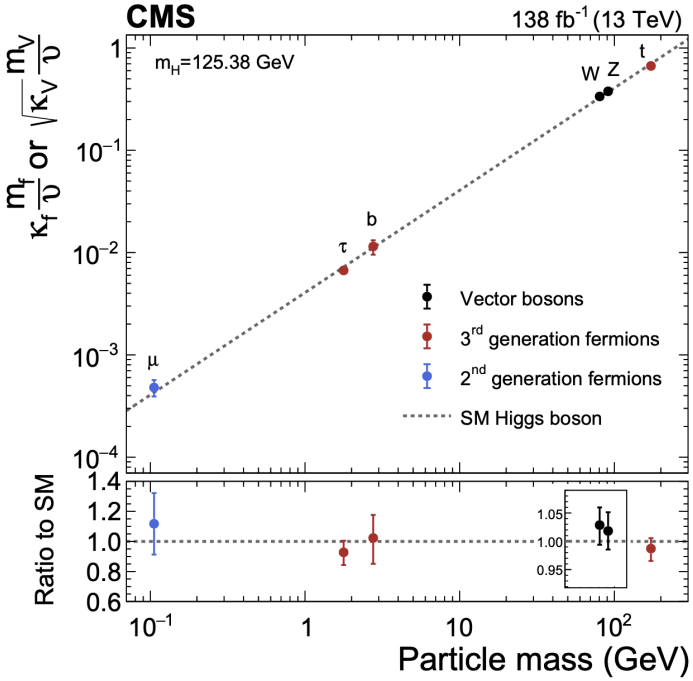


Figure 2.5: The measured $\kappa$-coupling modifiers for various Higgs boson decay modes as a function of the corresponding fermion/gauge boson mass [16].

To date, the signal strength modifiers associated to each of the decay modes also agree well with the SM predicted values within uncertainties (Fig. 2.4, bottom right panel). One can also perform the analysis in the $\kappa$-framework [19], which introduces $\kappa$ parameters scaling the interaction of the Higgs boson with a given particle both in production (affecting the cross section) and decay (affecting the decay width). The $\kappa$ value of one corresponds to the SM scenario. Experimental results for the $\kappa$-parametrised couplings of the Higgs boson with the fermions and gauge bosons indicate perfect agreement with the SM expectation over three orders of magnitude of mass (Fig. 2.5).

Due to the mentioned mass-dependent coupling of the Higgs boson, it is experimentally challenging to probe the coupling of the Higgs boson to the second and first generations fermions as having at least one order of magnitude lower mass than the third-generation fermions. However, it is of high importance to observe and measure these processes, either in production or in decay, to further verify that the observed particle with the mass $m_H \simeq 125$ GeV is indeed the SM Higgs boson. Recent evidence for the $H \to \mu\mu$ decay [21] and search for the $VH \to cc$ decays [22]

have made a promising step in this direction. Furthermore, the parameters of the Higgs potential (2.38) are also yet to be measured experimentally. This is where the ongoing quest to discover the Higgs pair production (Fig. 2.3, bottom panel) will shed more light on the Higgs boson nature [16].

The last item to be covered is the CP properties of the Higgs boson. The SM predicts the Higgs boson to be even under the charge-parity inversion, i.e. to have the quantum numbers of the pure CP scalar particle $J^{CP} = 0^{++}$. Experimentally, the hypotheses of a pure pseudoscalar as well as a spin-1 and spin-2 particle were excluded [23, 24] at the confidence level of more than $3\sigma$. However, it is possible that the observed Higgs boson is a mixture of scalar and pseudoscalar hypotheses. Such anomalies in the CP sector can be searched either in the Higgs coupling to the fermions or to the vector bosons. For the former, the Higgs interaction with the top quark and the tau lepton plays the dominant role. Pure CP-odd hypothesis was rejected at the confidence level of more than $3\sigma$ and upper limits were set on the anomalous Yukawa couplings to the top quark in the studies of tH and ttH processes [25–28]. The coupling with the vector bosons has also been probed in the VBF and VH production modes in the $H \to ZZ, \gamma\gamma, \tau\tau$ decays [29–32]. Overall, the results are compatible with the SM expectation of the pure CP scalar hypothesis. This work complements these studies by exploring the Yukawa coupling of the Higgs boson with tau leptons.

# 3 | CMS experiment at LHC

In order to study the physical properties of fundamental particles one should have access to their source with the possibility to construct a laboratory environment and carry out experiments in a controlled way. In this work, as it will be described in the following chapters, one needs to be able to produce Higgs bosons and to be able to measure their decays into a pair of tau leptons. The former is done using proton-proton ($pp$) collisions at the Large Hadron Collider (LHC) as described in Sec. 3.1. A Compact Muon Solenoid (CMS) detector detailed in Sec. 3.2 serves as a laboratory for the reconstruction of the Higgs boson decays into a pair of tau leptons and for the measurement of their properties.

## 3.1 LHC facility

The Large Hadron Collider [33] is a hadron accelarator built by the European Organization for Nuclear Research (Conseil Européen pour la Recherche Nucléaire or CERN for short) which started its operation in 2008. It is installed in a 3.8-metre wide circular tunnel with 26.7 km in circumference which was previously used for the operation of the Large Electron-Positron (LEP) Collider. The tunnel lies between 50 m and 175 m below the surface at the border between France and Switzerland near Geneva.

Original design for the LHC provides collision of protons with protons, protons with heavy ions (p-Pb), and heavy ions with heavy ions (Pb-Pb, Xe-Xe). The particles in each of the two beams are grouped in *bunches* in order to maximise the rate of interactions in a unit of time. For the proton-proton collisions, firstly a centre-of-mass energy of $\sqrt{s}$ = 7 TeV was reached in 2010 and 2011, followed by an increase to $\sqrt{s}$ = 8 TeV in 2012, where both periods are referred to as Run 1. During a Run 2 period corresponding to the years from 2016 to 2018 the LHC was operating at a centre-of-mass energy of $\sqrt{s}$ = 13 TeV. During the time of writing this work, a Run 3 period was ongoing at a centre-of-mass energy of $\sqrt{s}$ = 13.6 TeV.

In between the periods, two Long Shutdowns (LS) took place (in 2013-2015 and 2018-2022

years) which introduced major upgrades to the LHC and its detectors to allow for their operation at the higher centre-of-mass energy and collision rate. The latter is referred to as instantaneous luminosity and it is defined by the beam parameters with the following formula:

$$\mathscr{L} = \frac{N_b^2 n_b f_{\text{rev}} \gamma}{4\pi \varepsilon_n \beta^*} F. \tag{3.1}$$

Here, $N_b$ is the number of particle per bunch, $n_b$ is the number of bunches per beam, $f_{\text{rev}}$ is the beam revolution frequency, $\gamma$ is the beam relativistic gamma factor, $\varepsilon_n$ is the normalised transverse beam emittance, $\beta^*$ is the beta function at the point of collision, $F$ is the geometric reduction factor:

$$F = \left(1 + \left(\frac{\theta_c \sigma_z}{2\sigma^*}\right)^2\right)^{-1/2}, \tag{3.2}$$

where $\theta_c$ is the bunch crossing angle at the interaction point, $\sigma_z$ is a root mean square (RMS) of the bunch in the longitudinal direction, $\sigma^*$ is the RMS of the beam in the direction transverse to the beam trajectory.

The nominal design luminosity for the LHC was $10^{34}$ cm$^{-2}$ s$^{-1}$ and twice this value was reached during the Run 2 period. An integrated luminosity can be derived from the instantaneous one by integrating over a given period of time $[t_1, t_2]$:

$$L = \int_{t_1}^{t_2} \mathscr{L}\, dt. \tag{3.3}$$

Then, the expected number of *events* (i.e. occurrences in the detector) of a given process can be obtained by multiplying the integrated luminosity by the cross section of the process:

$$N_{\text{events}} = L \cdot \sigma_{\text{process}}. \tag{3.4}$$

Particles in the LHC are accelerated in two separate beam pipes each kept at the vacuum with the pressure ranging from $10^{-10}$ to $10^{-11}$ mbar. 1232 dipole magnets, each 15 m long, are used to bend the particle beams. 392 quadrupole magnets, each 5-7 m long are used to focus the beams. Sextupole, octupole, and decapole magnets are also used to correct slight imperfections in the magnetic field. The superconducting magnets are made of copper-clad niobium-titanium and are kept at the temperature of 1.9 K. This is achieved by using superfluid helium-4 and dedicated vacuum systems to create an insulating environment.

The CERN accelerator complex is used to inject protons into the LHC (Fig. 3.1). The energy of protons is progressively increased in stages by several accelerators. In order to obtain protons

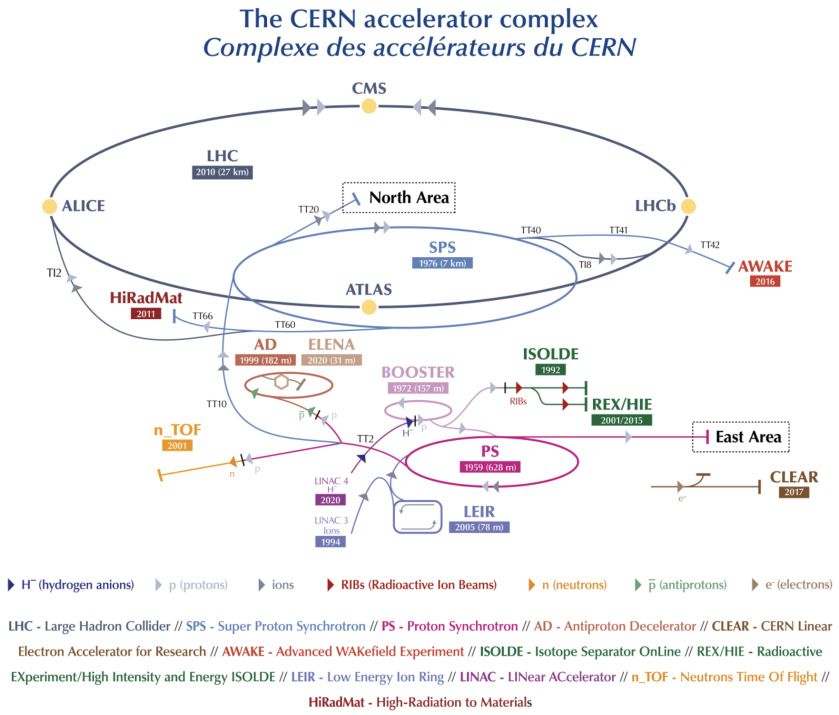


Figure 3.1: The CERN accelerator complex [34].

in the first place, hydrogen ions ($H^-$) with the energy of 160 MeV are generated with a Linac 4 linear accelerator. It started operation in 2020 and substituted the previously used Linac 2. The ions are further fed into a Proton Synchrotron Booster (PSB) where the electrons are removed from the atom leaving only a nucleus with a single proton. These protons are accelerated up to 2 GeV and fed into a Proton Synchrotron (PS) which brings their energy up to 25 GeV. After the PS, the protons are fed into a Super Proton Synchrotron (SPS) which increases their energy up to 450 GeV and feeds them into the main ring of the LHC, where they are accumulated and accelerated up to the target center-of-mass energy.

Nine experiments are installed at the LHC in order to study a broad range of particle physics phenomena. Four of them are the large experiments placed at the LHC beam intersection points in dedicated underground caverns:

- CMS ("Compact Muon Solenoid") [35] and ATLAS ("A Toroidal LHC ApparatuS") [36]. These are general-purpose detectors installed at the Point 5 (P5) and Point 1 (P1), respectively. The detectors are designed to study a wide range of phenomena from the Higgs mechanism to searches for new physics, e.g. dark matter candidates and extra dimensions.

- LHCb ("Large Hadron Collider beauty") [37]. It is placed at Point 8 (P8) and its specialisation is the study of heavy flavour physics ranging from CP violation to exotic hadron spectroscopy.
- ALICE ("A Large Ion Collider Experiment") [38]. It is installed at Point 2 (P2) with the specific design to study heavy-ion collisions and thus explore the properties of the quark-gluon plasma.

The other five experiments are smaller at scale and share the cavern with the ones mentioned above. These are:

- TOTEM ("Total, elastic and diffractive cross-section measurement") [39]. It is installed in the very forward region from P5 (CMS) along the beam trajectory. The detector is designed to study the proton structure and to measure the proton-proton interaction cross section complementary to the other general-purpose detectors.
- LHCf ("Large Hadron Collider forward") [40]. It is installed few hundred meters away from P1 (ATLAS) with the goal of simulating the environment of ultra-high-energy cosmic rays which is helpful for calibration of large-scale cosmic-ray experiments.
- MoEDAL-MAPP (Monopole and Exotics Detector at the LHC, MoEDAL Apparatus for Penetrating Particles) [41]. The detector is installed near P8 (LHCb) with the goal to directly search for the magnetic monopole and other highly ionizing stable massive particles.
- FASER (ForwArd Search ExpeRiment) [42]. It is located 480 metres downstream of P1 (ATLAS) and it is designed to search for new light and weakly interacting particles, e.g. dark photons, axion-like particles and sterile neutrinos. Furthermore, a dedicated sub-detector FASER$\nu$ is installed to detect and study neutrinos with TeV energies.
- SND@LHC (Scattering and Neutrino Detector at the LHC) [43]. It is a recently approved experiment to be installed close to P1 (ATLAS) which will cover the pseudorapidity range $7.2 < \eta < 8.4$ complementary to the other experiments. Its primary goal is to measure the process $pp \rightarrow \nu X$ to provide insights into the charmed-hadron production and search for feebly interacting particles.

The analysis presented in this work uses the data collected with the CMS detector during the Run 2 data-taking period (Fig. 3.2). The data set size corresponds to a total integrated luminosity of 137.62 fb$^{-1}$. In the future, starting from 2026 year, the LHC will undergo a High-Luminosity upgrade (HL-LHC) in order to extend its discovery potential and maintain the performance at the instantaneous luminosity increased by a factor of 10 comparing to the originally designed value [45]. After the upgrade HL-LHC is expected to deliver a total integrated luminosity of

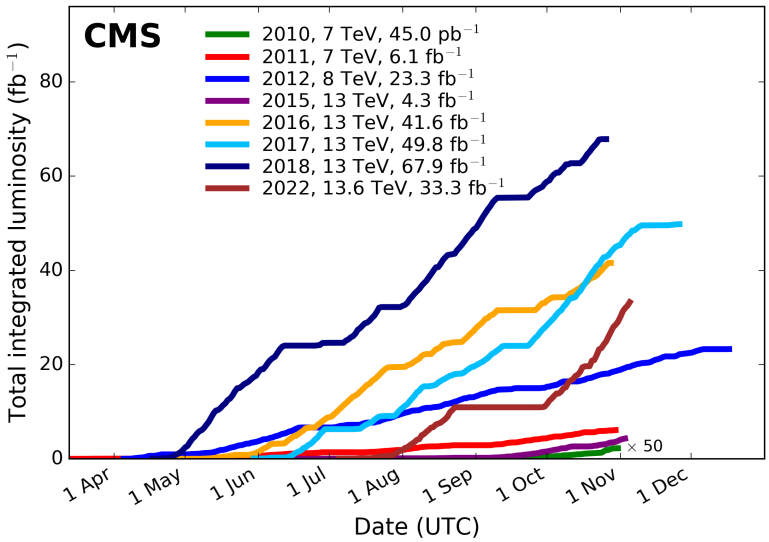


Figure 3.2: Delivered luminosity for *pp* collisions versus time for 2010-2012 (Run 1), 2015-2018 (Run 2), and 2022 (Run 3) years [44].

up to 4000 fb$^{-1}$ by the year 2040. The upgrade is expected to significantly improve the reach of physical searches and the measurements precision, also for the study presented in this work, as it is mentioned in Sec. 6.

## 3.2 CMS detector

The Compact Muon Solenoid detector [35] is a general-purpose apparatus designed to cover a broad spectrum of particle physics research at the TeV scale. It is placed about 100 meters underground at the Point 5 collision area of the LHC near Cessy, France. The detector weights about 14000 tonnes and it is about 15 metres high and 21 metres long (Fig. 3.3).

The detector consists of various subsystems each designed to meet a specific purpose, as described in the following subsections. In general, the concept of being suitable for performing a wide range of experiments drives the complexity of the design and the subsystems. This task is also complicated by the fact that the expected number of observed particles per *pp* bunch crossing is of the order of thousands, and the bunch crossings appear approximately every 25 ns. Therefore, the detector subsystems must be granular enough to provide the desired level of resolution.

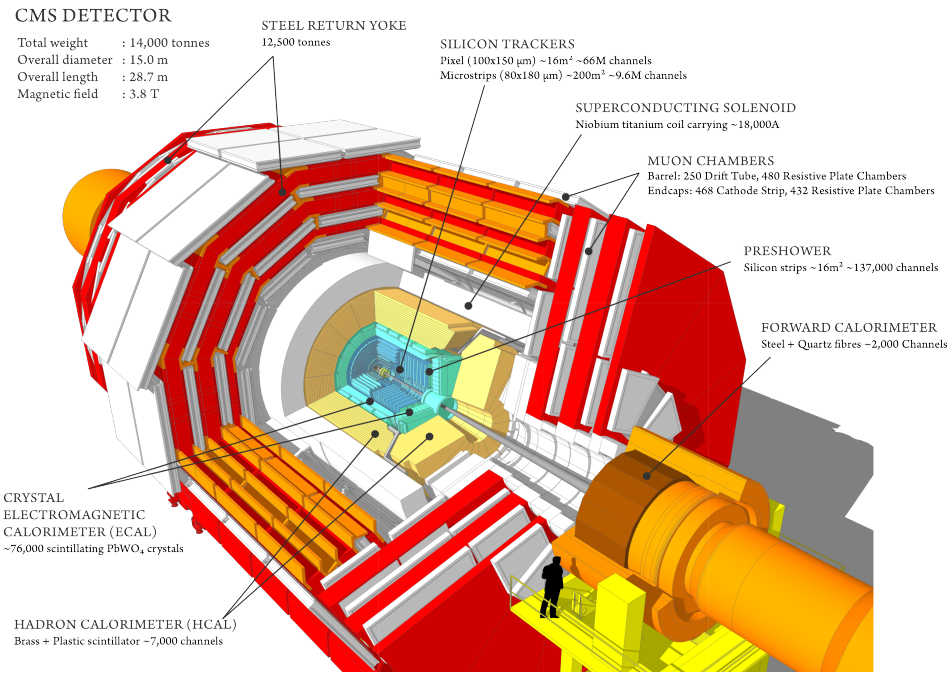


Figure 3.3: Schematic illustration of the CMS detector [46].

Overall, the detector design should provide physical measurements having an excellent quality and precision of the following items:

- Muon identification, (di)muon momentum resolution up to 1 TeV and precise charge assignment. This is largely achieved by the design of the magnet system (Sec. 3.2.1).
- Electron and photon reconstruction within the large geometrical acceptance, good (di)electron/photon energy resolution. This is largely achieved by the design of the electromagnetic calorimeter (Sec. 3.2.3).
- Charged-track momentum resolution and reconstruction efficiency. It is crucial, for example, for the jet and tau lepton reconstruction and it is ensured by the design of the tracking system (Sec. 3.2.2).
- Hermeticity for containment of all the particles appearing in the collision. This relies on the proper design of the hadron calorimeter (Sec. 3.2.4) and has direct impact on the precision of the missing transverse energy reconstruction (MET) (Sec. 5.3.4).

The following coordinate system is defined to describe positions and directions in the detector. The origin of the coordinate system is placed at the nominal $pp$ collision point. The $y$-axis points

vertically upward, the $x$-axis points radially inward towards the centre of the LHC ring, and the $z$-axis points along the tangent of the LHC ring from the Point 5 towards the Jura mountains. The azimuthal angle $\phi$ is measured from the $x$-axis in the $x - y$ plane as well as the radial distance $r$. Variables referred to as transverse (e.g. transverse momentum $p_{\mathrm{T}}$) correspond to the projections onto the plane transverse to the beam direction, i.e. $x - y$ plane. The polar angle $\theta$ is measured from the $z$-axis. The *pseudorapidity* $\eta$ is defined as:

$$\eta \equiv -\ln \tan(\theta/2). \tag{3.5}$$

### 3.2.1 Solenoid magnet

The superconducting magnet is at the core of the CMS detector. It is designed to provide a magnetic field of up to 4 T (with the nominal value of 3.8 T) which is crucial for precise measurement of the charged particles momentum, especially those of muons. The magnet has a cold bore diameter of 6.3 metres, a magnetic length of 12.5 metres and a mass of 220 tonnes. Its distinctive feature is a 4-layer winding (instead of the previously maximum of 2 layers e.g. in BaBar coils [47]) needed to provide the magnetic field of the nominal strength. A 10000-tonne yoke consisting of 6 endcap disks and 5 barrel wheels is used to return the magnetic flux. All the detector subsystems are placed inside of the main solenoid except for the muon chambers, which are installed within the return yoke.

### 3.2.2 Tracking system

One of the most fundamental steps of particle object reconstruction in the detector is the reconstruction of tracks. In the presence of a magnetic field it serves as the basis for measuring particle properties such as charge and momentum. Furthermore, tracks are fundamental blocks for the actual particle reconstruction (Sec. 4.2.1) as well as the reconstruction of the primary interaction point (Sec. 5.3.5). Therefore, efficient design of the tracking system responsible for the measurement of tracks is necessary.

As mentioned earlier, the expected number of particles appearing per $pp$ bunch crossing is extremely large. This puts strict requirement on the tracking system with granularity, speed, and radiation hardness being the most crucial points. These requirements motivate the choice of the silicon technology for the whole tracking system. The tracker constitutes the innermost subsystem of the CMS detector as it directly surrounds the proton collision point. It has a cylindrical shape with an outer radius of 1.2 m and a length of 5.6 m. It consists of two main parts: the pixel detector and the silicon strip tracker (Fig. 3.4).

The Phase-0 configuration (before 2017 year) of the pixel detector consists of three barrel layers at radii of 44, 73, and 102 mm and two endcap disks from both sides at distances 345

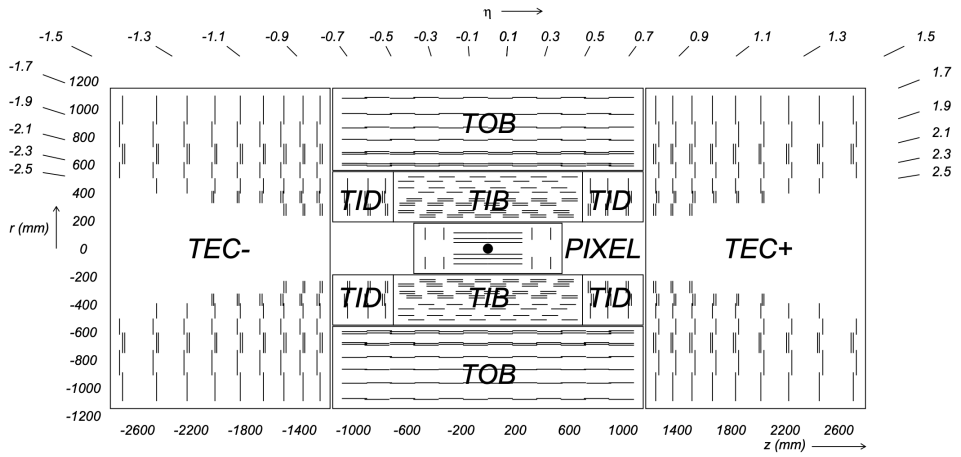


Figure 3.4: Schematic view of the CMS tracker for the Phase-0 configuration [35]

and 465 mm from the collision point. The pixel cell size is $100 \times 150$ $\mu$m$^2$ in $r - \phi$ and $z$, respectively. It is chosen to provide the same track resolution in both directions which in turn allows for a precise 3D reconstruction of primary/secondary vertices. The nominal coverage for this configuration is $|\eta| < 2.5$ in terms of pseudorapidity and 1.1 m$^2$ in terms of area, with about 66 million pixels in total. For the Phase-1 configuration [48], the pixel detector was upgraded to account for the future increase of the instantaneous luminosity leading to an increased multiplicity of tracks in the tracker. Barrel layers were arranged more closely to the collision point at radii 29, 68, 109, and 160 mm with one more layers added. One more endcap disk was also added to both sides with the placement distances being 291, 396, and 516 mm from the collision point. The total coverage area was therefore extended to 1.9 m$^2$ which corresponds to the pseudorapidity range $|\eta| < 3.0$. Overall, the upgrade leads to an improved transverse track impact parameter resolution from $25 - 90$ $\mu$m to $20 - 65$ $\mu$m for tracks with the transverse momentum $p_{\mathrm{T}} \in [1, 10]$ GeV and $|\eta| < 1.4$ (3.0) for the Phase-0 (Phase-1) configuration [49].

The silicon strip tracker spans the radial distance between 20 cm and 116 cm and consists of three subsystems:

- Tracker Inner Barrel and Disks (TIB/TID). These occupy the region radially up until 55 cm with 4 barrel layers and 3 endcap disks at each end. The mean strip pitch varies from 80 (100) $\mu$m to 120 (141) $\mu$m for TIB (TID).

- Tracker Outer Barrel (TOB). It extends up to 116 ($\pm$118) cm in radius ($z$) and consists of 6 barrel layers with a pitch varying from 183 $\mu$m to 122 $\mu$m from the innermost layers outwards.

- Tracker Endcaps (TEC+/TEC-). The modules cover the region 124 cm $< |z| <$ 282 cm

and 22.5 cm < $|r|$ < 113.5 cm. Each TEC consists of 9 disks with radial strip pitch varying from 97 $\mu$m to 184 $\mu$m.

Additionally, the modules in the first two layers of TIB and TID, the rings of the first two layers of TOB and layers 1, 2, and 5 of TECs include a second micro-strip module mounted back-to-back with a stereo angle of 100 mrad. This allows for the measurement of the second coordinate in the corresponding modules (either $z$ in the barrel or $r$ in the endcap).

### 3.2.3 Electromagnetic calorimeter

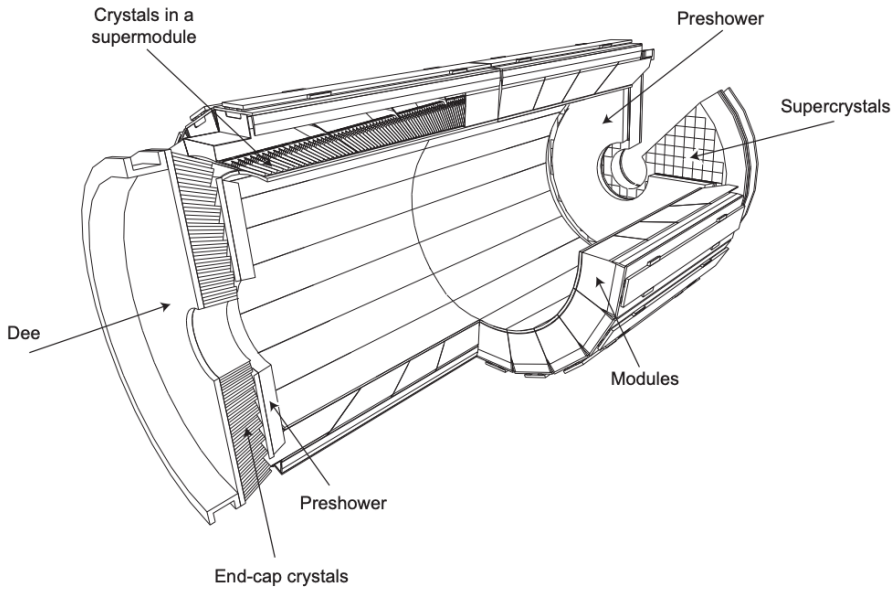


Figure 3.5: Schematic layout of the CMS electromagnetic calorimeter [35].

The electromagnetic calorimeter in CMS is an hermetic homogeneous calorimeter covering the pseudorapidity range of $|\eta| < 3$ (Fig. 3.5). The main driving motivation behind its design is to be able to detect the decay of the Higgs boson into a pair of photons which requires a good energy resolution for the reconstructed photons. Also, general LHC conditions put constraints on the fast scintillation time, fine granularity, and radiation resistance. Altogether, the choice of $PbWO_4$ crystals for the main calorimeter material meets all of these criteria. The scintillation time is of the same order as the time between two consecutive bunch crossing time ($\sim$ 25 ns). Moreover, the small Moliere radius (2.2 cm) and the short radiation length $X_0$ (0.89 cm) allow for the construction of the compact calorimeter with fine granularity.

The barrel part of ECAL (EB) contains $360 \times 2 \times 85 = 61200$ crystals with 8.14 $m^3$ in volume (67.4 tonnes in weight) covering the range $|\eta| < 1.479$ in pseudorapidity. The front faces of the crystals are positioned at a radius of 1.29 m from the collision point. The cross-section of crystals is approximately $0.0174 \times 0.0174$ in $\eta - \phi$ plane. The length of crystals is 25.8 $X_0$ (230 mm) which allows for more than 98% energy containment for electrons and photons with the energy up to 1 TeV.

The endcap part of ECAL (EE) for both sides consists of two parts (so-called *dees*) with 3662 crystals each. The crystals are arranged in a rectangular $x - y$ grid with off-pointing angles ranging from 2 to 8 degrees. The front (rear) cross-section for the crystals is $28.62 \times 28.62$ $mm^2$ ($30 \times 30$ $mm^2$) and the length is 24.7 $X_0$ (220 mm). The total EE crystal volume is 2.90 $m^3$ with a weight of 24 tonnes.

A fine-grained preshower detector is installed before each of the endcap disks with the goal of distinguishing $\pi^0$ decays from prompt photons. It consists of two layers where each layer is a lead radiator followed by a silicon strip sensor plane. The two radiators are of the size of approximately one and two radiation lengths. The pitch of the silicon sensors is 1.9 mm ($61 \times 61$ $mm^2$ divided into 32 strips).

The scintillation light is collected with avalanche photodiodes (vacuum phototriodes) in the EB (EE) due to their high radiation tolerance, fast response, and ability to operate in the 4-T magnetic field. The ECAL subsystems are precisely maintained at the operating temperature of 18° with a dedicated water cooling system to mitigate fluctuations in the number of scintillated photons.

### 3.2.4 Hadron calorimeter

The hadron calorimeter (HCAL) plays an important role not only in the precise measurement of jet properties, but also in the reconstruction of the missing transverse energy, which is sensitive to the level of containment of hadronic activity in the detector. The CMS HCAL is a hermetic sampling calorimeter made of alternating layers of a brass absorber and plastic scintillator tiles (Fig. 3.6). It is placed after the tracking system and ECAL (at a radius of 1.77 m) before the solenoid coil (at a radius of 2.95 m).

The barrel part of HCAL (HB) covers the pseudorapidity range of $|\eta| < 1.3$. The thickness of the absorber reaches up to six interaction lengths $\lambda_I$ and increases up to ten interaction lengths at larger pseudorapidities. ECAL in front of HB adds $1.1\lambda_I$ of the material. The plastic scintillator is divided into 16 sectors in pseudorapidity which results in the $0.087 \times 0.087$ segmentation in the $\eta - \phi$ plane. The light from the scintillators is passed with wavelength shifting fibres to hybrid photodiodes.

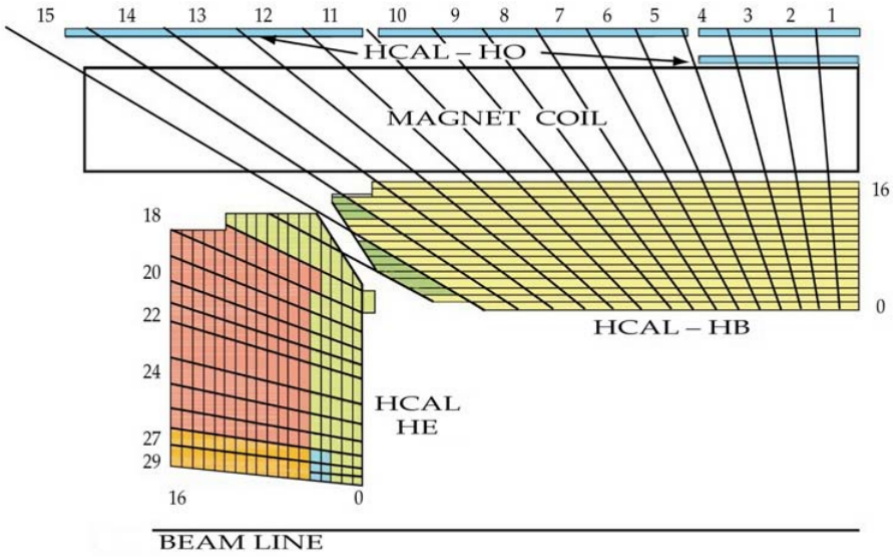


Figure 3.6: Schematic layout of one-fourth of the CMS hadronic calorimeter in the $r - z$ plane [35].

While the HB part is placed within the magnet solenoid, it is also complemented by a tail catcher (HO) outside of the solenoid. This module is added since EB together with HB do not provide sufficient containment of hadron showers. HO is placed as the first layer in each of the five rings (along the $z$ axis) which form an iron yoke returning the magnetic field. In the most central ring ($|\eta| \sim 0$) two layers of scintillators are placed, while all the other rings have a single layer. The cross-section of tiles in HO approximately matches those of HB with the granularity $0.087 \times 0.087$ in the $\eta - \phi$ plane. It was studied that the energy response of the calorimeter is more Gaussian for the HCAL configuration with HO comparing to those without HO [50]. Furthermore, the effect of leakage is reduced which propagates to the improved measurement of the missing transverse energy.

The endcap part of HCAL (HE) covers the pseudorapidity range of $1.3 < |\eta| < 3.0$. Similarly to HB, it is also composed of alternating layers of brass absorber and plastic scintillator tiles. In total, the number of tiles for both HE parts is 20 916 with the granularity $\approx 0.17 \times 0.17$ in the $\eta - \phi$ plane. The depth of the entire calorimeter system (ECAL and HCAL) reaches about 10 $\lambda_I$ in the endcap and 12 $\lambda_I$ in the barrel regions.

The last part of the CMS hadron calorimeter system is the forward calorimeter (HF). It is installed from both sides of the interaction point at the distance of ±11 m along the beam line. It is designed to cover the pseudorapidity range up to $|\eta| \approx 5$ and to survive under harsh radiation conditions. HF is made of grooved plates of a steel absorber. Quartz fibres are chosen as active

material of the calorimeter due to their radiation resistance and fast response. They are placed in the grooves with the long fibres (over the full thickness of the absorber) alternating with the short ones (covering the back of the absorber). The towers cross-section is $\approx 0.175 \times 0.175$ in the $\eta - \phi$ plane. The signals from the short and long fibres initiated by the Cherenkov light from the charged particles as they pass through the HO medium are used to measure separately the electromagnetic and hadronic components of the shower.

### 3.2.5 Muon system

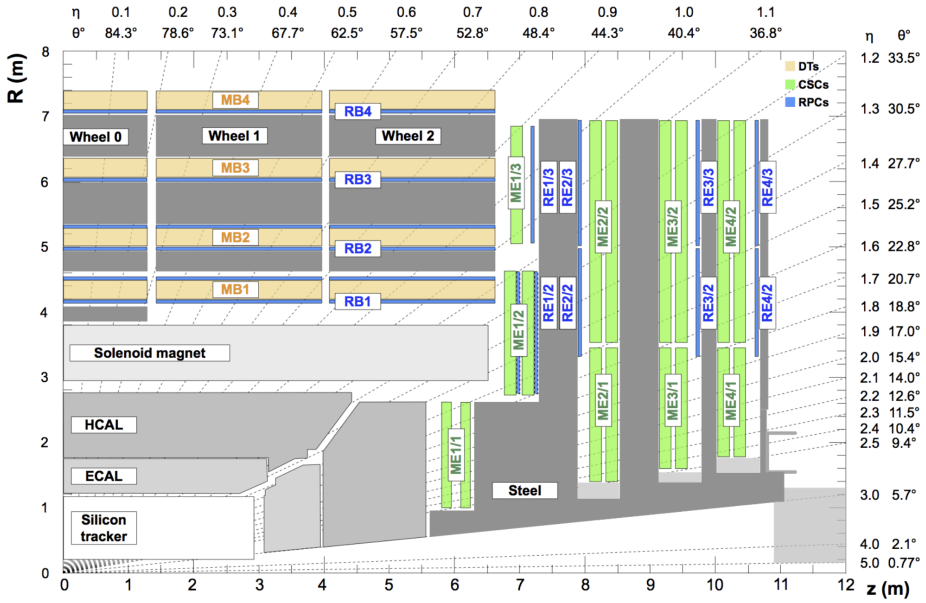


Figure 3.7: Cross section of the CMS detector in the $r - z$ plane illustrating the subdetectors of the muon system [51]. The drift tube stations are shown in yellow (labelled as MB), the cathode strip chambers are shown in green (labelled as ME), the resistive plate chambers are shown in blue (labelled as RB and RE).

The CMS detector is originally designed to precisely identify and measure muons across the whole kinematic experimental range. This is achieved by a dedicated structure of the muon system. Since it is placed outside of the solenoid, it necessarily has to cover a large area which puts forward the requirements for the system to be inexpensive and robust.

In the barrel region, the low particle rate is expected and drift tube (DT) chambers are used. There are 4 stations in total placed between the layers of the flux return plates covering the pseudorapidity range $|\eta| < 1.2$. The first 3 stations contain 8 chambers grouped into a pair of

4 chambers. The first 4 chambers measure the coordinate in the $r - \phi$ plane, while the other 4 stations measure the $z$ coordinate. The fourth station measures only the $r - \phi$ coordinate.

In the endcap region, a high particle rate is expected and therefore cathode strip chambers (CSC) are used. Covering the pseudorapidity range $0.9 < |\eta| < 2.4$ they provide high radiation resistance and fast time response. In total there are 4 CSC stations in each endcap placed between the flux return plates. The cathode strips within each chamber are organised radially and measure the $r - \phi$ coordinate. The anode wires are arranged perpendicularly to the strips and measure the $\eta$ coordinate and the beam crossing time of the muon.

Both DT and CSC can be used for event triggering independently of the rest of the detector due to a relatively good $p_T$ resolution and background rejection. However, due to a low resolution in the beam crossing time, resistive plate chambers (RPC) are added both to the barrel and endcap regions to improve the triggering capabilities. These are double-gap chambers operating in avalanche mode and covering the pseudorapidity range $|\eta| < 1.6$. There are 6 RPC layers in the barrel region with 2 layers placed in each of the first 2 stations and 1 layer placed in each of the last 2 stations. In the endcap region, an RPC plate is added to each of the first 3 stations. Overall, they provide improved time resolution and fast response at the expense of having lower coordinate resolution.

Lastly, for the Phase-2 upgrade of the detector, several updates to the muon system are being performed [52] in order to be able to operate efficiently at higher luminosity. For the DT chambers and CSC, electronics will be changed to maintain the same level of latency and readout rate for the trigger system. For the forward region, two new sections of improved RPCs are added to extend the RPC pseudorapidity coverage from 1.9 to 2.4. Moreover, three sets of detectors based on Gas Electron Multiplier (GEM) technology will be installed. This will extend the pseudorapidity range of the offline muon reconstruction up to $|\eta| < 2.8$ with an improved trigger capabilities.

### 3.2.6 Trigger and data facility

LHC delivers proton-proton collisions at the very high rate of 40 MHz (1 bunch crossing per 25 ns), which makes it extremely challenging to analyse data from various CMS subdetectors online. Furthermore, the high granularity of the CMS detector results in a large amount of data which is not possible to store. A trigger system is designed to tackle this problem by significantly reducing the size of original data by at least a factor of $10^6$.

The trigger system consists of two main stages [53]. The first stage consists of a Level-1 (L1) trigger and a Data Acquisition (DAQ) system which aim to reduce the initial rate from 40 MHz to 100 kHz. It uses raw information available separately from the calorimeter and muon systems to coarsely determine high-level physics objects (e.g. jets or isolated/non-isolated

$e/\gamma$). This information is further used to define criteria whether a given collision should be rejected or accepted for further processing. All the computation are performed in hardware with field programmable gate arrays (FPGA), application-specific integrated circuit (ASIC) and programmable memory lookup tables (LUT) technologies to allow for processing of large scale data online at high rates.

The second stage is a High-Level Trigger (HLT) and it aims to reduce the L1 rate from 100 kHz to 1 kHz. Contrary to the L1 trigger, software computations are performed on a large farm of processors at the HLT level. Reconstruction algorithms similar to the ones used offline are being run for each of the L1-selected collisions to reconstruct physics objects. For that purpose, several so-called HLT paths are defined each targeting to select collisions with a specific predefined topology. Software modules in HLT paths are arranged sequentially and gradually build objects of increasing complexity from raw detector-level data. In the end of the path, a decision is made based on the information about reconstructed HLT physics objects whether a given collision should be recorded.

Events accepted by the trigger system are sent to a "storage manager" software process which handles the transfer of data from local disks to a CMS Tier-0 computing center at CERN. In general, the CMS computing system is structured hierarchically into three tiers combining data centres worldwide:

- Tier-0 (CERN). It handles transfer of online data to a permanent storage, performs first reconstruction of raw data and exports it to Tier-1 centers.
- Tier-1 (few national computing facilities). The tier provides a long-term safe second storage of CMS raw and simulated data outside of Tier-0 and rapid transfer of data to Tier-2 centers. Additionally, it performs the second reconstruction of the raw data and provides the analysis of vary large data samples.
- Tier-2 (several research institutes). The tier is responsible for the local storage of data transferred from Tier-1. It also provides support for the final-stage analysis of data sets and production of the Monte Carlo (MC) simulated data with its transfer to Tier-1 centers.

Lastly, resources of all the data centres across the tiers are integrated into a single Worldwide LHC Computing Grid (WLCG) system relying upon Grid middleware to standardise access to storage and CPU facilities [54]. It provides an interface for the CMS-specific distributed computing software to remotely access data and perform job submission for the data analysis at scale.

# 4 | Tau lepton reconstruction & identification

The tau lepton, being the heaviest of the three discovered charged leptons in the Standard Model, plays a crucial role in understanding of matter at the most fundamental level. For example, in the context of the minimal supersymmetric extension of the SM (MSSM) [55, 56], there is a special interest in searches for neutral and charged Higgs bosons decaying into a pair of tau leptons [57]. Furthermore, in the light of testing the Lepton Flavour Universality (LFU) several observed tensions with the SM predictions are yet to be understood [58–61].

The precision of such analyses heavily relies on the ability in a given experiment to accurately reconstruct and separate tau leptons from background processes. However, the tau lepton stands out from the other leptons in its properties, which poses several challenges in this endeavour. In particular, it is the only lepton known to decay into hadrons, which makes it difficult to distinguish such decays (hereafter labelled as $\tau_h$) from jets originating from QCD processes. Therefore, this Chapter will take an experimentalist's perspective and will describe, with a particular emphasis on the CMS experiment, challenges and achievements accomplished so far in the $\tau_h$ reconstruction and identification (hereafter also referred to as "tau lepton reconstruction/identification").

The Chapter is organised as follows. In Section 4.1 a brief overview of the tau lepton's history and its properties measured to this date is given. Section 4.2.2 introduces methods to reconstruct tau leptons in the CMS experiment, in particular a hadron-plus-strip (HPS) algorithm built on top of the Particle Flow (PF) algorithm, described in Section 4.2.1. After the reconstruction of tau lepton candidates, an identification step has to be performed to categorise whether a given candidate originates from a genuine tau or a jet/lepton faking tau. An algorithm named DeepTau was developed for that purpose and its details – including the recent improvement in the context of the Run 3 data taking – will be described in Sections 4.3.1 and 4.3.2. However, the algorithm has several intrinsic limitations in its design, and ongoing efforts to overcome them with new Machine Learning (ML) models will be detailed in Section 4.3.3.

## 4.1 Discovery & Properties

The tau lepton was observed for the first time in the Mark I experiment at the SPEAR $e^+e^-$ storage ring at the Stanford Linear Accelerator Center (SLAC) in 1974 by Martin L. Perl et al. [62]. Fundamentally, the motivation [63] behind the analysis was to solve an electron-muon problem, which manifests itself in two questions [64]:

- Why is the muon 206.8 times heavier than electron?
- Why doesn't the muon decay through the process $\mu \rightarrow e + \gamma$?

One of the ideas to understand this difference was to change the perspective and search for additional heavy leptons, which, in case of their existence, could help to gain insights into the initial problem. The theoretical framework to search for such leptons was a sequential heavy lepton model due to its elegance, symmetry and simplicity [65]. The minimalistic and main assumption it makes is the existence of pairs $(L_\alpha, \nu_\alpha)$ of charged leptons and associated neutrinos with the lepton masses larger than those of the electron and the muon.

Additionally, the sequential heavy lepton model builds upon the concept of the lepton number conservation. In the original formulation, it postulates that electron and muon each possess a unique property not possessed by other particle, a lepton family number, meaning that electron $e^-$ and its associated neutrino $\nu_e$ are assigned a lepton number $n_e = +1$, $\mu^-$ and $\nu_\mu$ receive a number $n_\mu = +1$ and antiparticles have the corresponding number negative. This lepton number should be preserved in reactions separately for each of the lepton family.

Assuming the lepton family conservation, the sequential heavy lepton model expands this principle to other lepton families with higher masses. From these principles it follows that given a high enough mass of a heavy charged lepton $L^-$, there should exist the following decays:

D1 $L^- \rightarrow e^- \bar{\nu}_e \nu_L$,

D2 $L^- \rightarrow \mu^- \bar{\nu}_\mu \nu_L$,

D3 $L^- \rightarrow \pi^- \nu_L$,

D4 $L^- \rightarrow \pi^- \pi^+ \pi^- \nu_L$,

where the former two are exactly analogous to the corresponding decay of the muon into electron and two neutrinos via the weak interaction.

After making an additional ansatz that heavy charged leptons can be produced similarly to electrons and muons in reactions $e^+e^- \rightarrow L^+L^-$ , Martin L. Perl et al. proposed an elegant idea

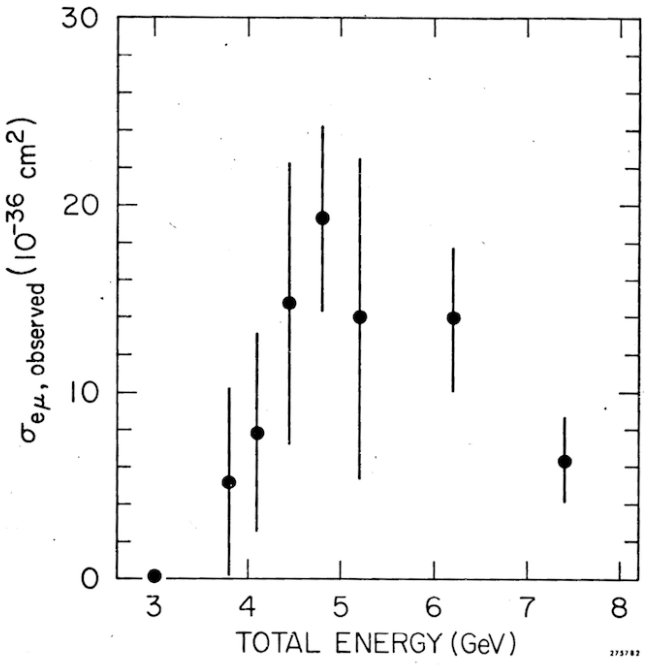


Figure 4.1: The observed background-subtracted cross section versus center-of-mass energy in the Mark I experiment within the detector acceptance for events with the $e^{\pm}\mu^{\mp}$ signature. [62]

to search for them in this production mode by looking into a process where one $L$ would decay via D1 and the other via D2. Such an unusual final state consisting of $e^{\pm}$ and $\mu^{\mp}$ of opposite charge and a missing energy due to neutrinos escaping detection would hint to anomalous processes appearing in the detector. An excess of such events over the small background expectations (Fig. 4.1) for which the analysts had "no conventional explanation" was exactly what was observed at SPEAR. After the follow-up studies [66–71] it was finally concluded, that the observed excess indeed can be attributed to a lepton with the mass 3600 times the electron mass and 17 times the muon mass, later called *tau* (from greek $\tau\rho\iota\tau\sigma\nu$, "third"). These results together with an observation of the tau neutrino by the DONUT collaboration [72] therefore established the existence of the third generation of leptons.

Since the era of its discovery, the properties of the tau lepton has been extensively studied in several experiments including Belle, BaBar, BESIII, CLEO, KEDR, LEP experiments, PLUTO, and others. They can be summarised as follows [7]:

- Mass $m_\tau = 1776.86 \pm 0.12$ MeV.
- Mean lifetime $\tau = (290.3 \pm 0.5) \cdot 10^{-15}$ s, with the lifetime difference between $\tau^+$ and $\tau^-$: $(\tau_{\tau^+} - \tau_{\tau^-})/\tau_{\text{average}} < 7.0 \times 10^{-3}$ at 90% C.L.

- Decay modes (DM): notable feature is the existence of hadronic decays, not present for the other leptons. A brief summary of those decay modes relevant to this study is presented in Table 4.1, inspired by [73].

Table 4.1: Decay modes of the tau lepton with the corresponding branching fractions $\mathscr{B}$ [7]. If applicable, intermediate known resonances contributing to decay modes are mentioned. $h^{\pm}$ denotes a charged hadron and the same numbers apply for the charge-conjugated decays.

| Decay mode | Resonance | $\mathscr{B}$ (%) |
|---|---|---|
| Leptonic decays | | 35.2 |
| $\tau^- \to e^- \bar{\nu}_e \nu_\tau$ | | 17.8 |
| $\tau^- \to \mu^- \bar{\nu}_\mu \nu_\tau$ | | 17.4 |
| Hadronic decays | | 64.8 |
| $\tau^- \to h^- \nu_\tau$ | | 11.5 |
| $\tau^- \to h^- \pi^0 \nu_\tau$ | $\rho(770)$ | 25.9 |
| $\tau^- \to h^- \pi^0 \pi^0 \nu_\tau$ | $a_1(1260)$ | 9.5 |
| $\tau^- \to h^- h^+ h^- \nu_\tau$ | $a_1(1260)$ | 9.8 |
| $\tau^- \to h^- h^+ h^- \pi^0 \nu_\tau$ | | 4.8 |
| Other | | 3.3 |

## 4.2 Reconstruction in CMS

### 4.2.1 Particle Flow algorithm

In order to perform physical measurements in the particle physics context, one usually operates with an abstract notion of a *physics object*, which is an entity reconstructed from the signals observed in the detector and representing a particle candidate of a particular kind. The goal of the reconstruction process is to build physics objects which are as close and as representative as possible of the genuine particles appearing in the detector. Since the precision of the object reconstruction has a direct impact on the precision of the physical measurement, it therefore plays a crucial role in every particle physics analysis.

Historically, the reconstruction of particles of a given type was primarily based on the information of detector's subsystems which were specifically built to identify them. For example, the reconstruction of electrons and photons was primarily based on the ECAL response and was aimed to capture rather isolated particles. This approach can be referred to as *local*, because it does not make a full use of the signals across all detector subsystems due to a technical granularity limitation.

With detectors becoming more fine-grained one could turn from a local to a *global* approach of the physics objects' reconstruction. With that it became possible to build a *holistic* image of an event in the detector by linking information from various detector subsystems. This is exactly the core idea behind a *particle-flow (PF) algorithm* [74], developed in the CMS experiment, which aims at tracing the entire "flow" of particles as they are traversing the detector.

### Basic elements

The PF algorithm follows a hierarchical approach in the reconstruction of physics objects. The first step in the algorithm is to construct basic PF elements which will later serve as a basis for building more complex high-level objects. The main PF elements being constructed at this point are:

- Charged-particle tracks,
- Electron and muon tracks,
- Preshower, ECAL and HCAL energy clusters.

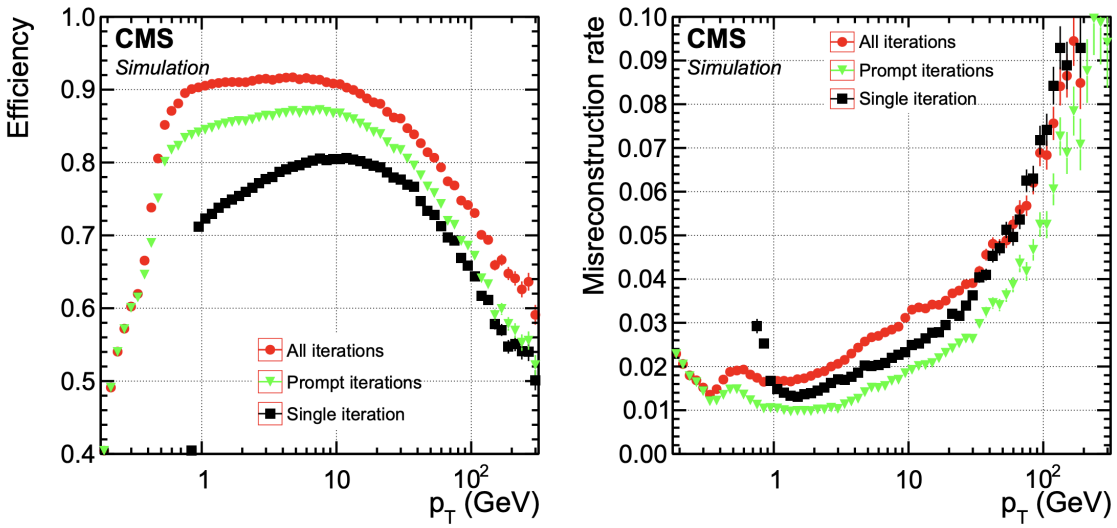


Figure 4.2: Efficiency (left) and misreconstruction rate (right) as a function of the reconstructed track $p_T$ for the charged hadrons in a sample of simulated QCD multijet events for the track reconstruction algorithm in CMS [74]. Black squares correspond to the global combinatorial track finder [75], green triangles to the prompt iterations of the iterative procedure [76] seeded by at least one hit in the pixel detector, red dots to all the iterations of the procedure.

For the reconstruction of **charged-particle tracks**, a pattern recognition approach using a combinatorial Kalman filter has been an indispensable tool among experimentalists for decades, also within the CMS experiment [77]. To improve the overall track reconstruction efficiency while keeping the misreconstructed rate at the same level, an iterative approach is taken [76]. With each fitting iteration, it targets to recover inefficiencies for a specific track type by e.g. tailoring the seed construction to a given track type, which allows the loosening of the requirement on the number of hits. The track types include prompt or displaced high/low $p_{\mathrm{T}}$ tracks, tracks inside high $p_{\mathrm{T}}$ jets, and muon tracks. Overall, the iterative procedure brings a significant recovery in efficiency across the $p_{\mathrm{T}}$ range while also performing twice faster compared to a single iteration approach due to the optimised seed construction (Fig. 4.2).

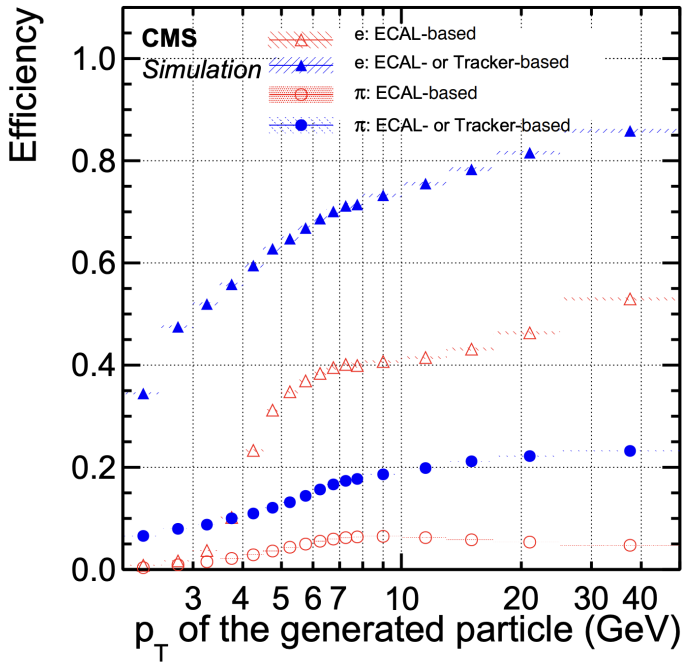


Figure 4.3: Efficiency for electrons (triangles) from simulated b quark jets and charged hadrons (circles) to give rise to an electron seed as a function of $p_{\mathrm{T}}$. Efficiencies for both ECAL-only seeding (hollow symbols) and ECAL-only with added tracker-based seeding (solid symbols) are displayed.

Reconstruction of **electron tracks** largely profits from the iterative tracking procedure, since the latter allows to efficiently reconstruct electrons radiating bremsstrahlung photons. It therefore forms the basis of a tracker-based seeding method, in contrast to a conventional ECAL-based, which suffers from misreconstruction of non-isolated electrons as well as radiating electrons. Generally, for electrons with a small fraction of radiated energy, the tracks are usually well-reconstructed and therefore can be extrapolated to the ECAL surface and matched with the

closest ECAL cluster to form an electron candidate seed. However, when energetic photons are radiated, the pattern recognition may result in tracks fitted with large $\chi^2$ values. For such tracks, an additional fitting is performed with a Gaussian-sum filter (GSF) [78], which essentially is an adaptation of the Kalman filter fitting accounting for possible sudden and substantial energy losses along the electron's trajectory. Finally, both tracker-based and ECAL-based electron seed are passed to an extended GSF algorithm for a full track reconstruction. Overall, this procedure significantly improves electron reconstruction efficiency by up to a factor of two compared to an ECAL-based only approach, while also extending the allowed $p_{\mathrm{T}}$ range down to 2 GeV (Fig. 4.3).

While electron track reconstruction was largely improved in the context of the PF algorithm development, **muon tracking** is not its specific part but a standalone step [51]. The output of this procedure is a collection of muon track candidates of three types:

- *Standalone-muon tracks* built by running the pattern recognition on the hits exclusively in the muon spectrometer subsystems (DT, CSC, RPC).
- *Tracker muon tracks* reconstructed with an "inside-out" approach, where firstly the tracker tracks are extrapolated to the muon system. If at least one muon segment matches to extrapolated track, the inner tracks is declared as a tracker muon.
- *Global muon tracks* reconstructed with an "outside-in" approach, where standalone muon tracks are being matched with tracker tracks. The matching is performed by propagating both tracks to a common surface and is followed by a combined fit with the Kalman filter.

The resulting approach yields an excellent reconstruction of muon with about 99% of the muons produced in the acceptance region of the detector being classified as either global or tracker muon tracks.

**Calorimeter clusters** reconstruction is an essential part of the PF algorithm. On the one hand, it plays a crucial role in the identification of neutral particles for which no track information is available, and their separation from the charged particles. On the other hand, it brings additional information to the reconstruction of electrons radiating bremsstrahlung photons and also charged hadrons. Once combined with the track information, it helps to sizeably increase the purity of the final collection of physics objects.

The building of calorimeter clusters, performed separately in each subdetector (ECAL barrel/endcap, HCAL barrel/endcap, preshower layers), consists in two steps:

1. *Clustering*. Firstly, seeds are formed from the calorimeter cells with the deposited energy larger of the neighbouring cells and above a given seed threshold. The neighbouring cells

are either the four closest cells sharing a side with the seed candidate, or the eight closest cells sharing either a side or a corner with the seed candidate. Then, topological clusters are grown on the principle of recursively adding neighbouring cells, passing a certain set of energy requirements, to the current cell, starting from the seeds candidates.

2. *Energy attribution.* Due to a potential overlap of topological clusters, a dedicated approach was developed to allow for sharing of the cells energy across several clusters. Performed with an expectation-maximisation algorithm based on a Gaussian-mixture model, it assumes that the energy deposited in a topological cluster is a composition of as many spatially Gaussian-distributed energy clusters, as there are seeds in the topological clusters. The model parameters (location and total energy of each cluster in the Gaussian mixture) obtained after convergence are used as cluster parameters in the downstream reconstruction.

After the clusters are formed, it is important to calibrate the response of the calorimeters to have a correct energy scale and identification of neutral and charged particles. Performed for ECAL [79] and HCAL [80] by fitting a parametric function which maps true values of energy and pseudorapidity of the cluster to the calibrated ones, the calibration procedure was successfully validated on data and showed overall improved energy response compared to a raw one.

### Linking & Object construction

Having formed the fundamental PF elements, the PF algorithm proceeds to linking them between each other in order to form *PF blocks*, essentially representing chains of PF elements. In order to reduce the computing time, the linking is performed as a search for the nearest neighbours in the $\eta - \phi$ plane via a k-dimensional tree [81].

With the PF blocks being built, the PF algorithm proceeds to the identification of physics objects on the per block basis. The objects are formed in the following order: muons, electrons together with isolated photons (converted or unconverted), hadrons (charged or neutral) and nonisolated photons (e.g. from $\pi^0$ decays). To account for possible nuclear interactions in the tracker material, secondary charged-particle tracks which are linked via a common nuclear-interaction vertex are merged into a single primary particle. Lastly, events are post-processed to resolve rare cases of anomalously large $p_{\mathrm{T}}^{\mathrm{miss}}$ coming from e.g. the misreconstruction of muons. A detailed description of requirements applied to PF elements within a block at each of the steps can be found in the original paper [74].

Once all PF blocks are emptied and physics objects of the aforementioned types are formed, one can proceed to grouping them into more complex objects. One of them is the tau lepton object, which reconstruction with a *hadron-plus-strips (HPS) algorithm* is described in the next section.

### 4.2.2 HPS algorithm

Since the tau lepton in about 65% of all cases decays into the final state with hadrons (Section 4.1), it is important to efficiently identify such topologies in the detector. While the leptonic decays of the tau lepton in the CMS experiment are handled by the usual techniques for muon [51] and electron [82] reconstruction and identification, the hadronic decays pose a challenge of separating them from an overwhelming background of QCD jets. To tackle this, a hadron-plus-strips (HPS) algorithm was designed, originally for the LHC operation at $\sqrt{s}$ = 7 TeV [83] and 8 TeV [84], followed by improvements for the data taking at $\sqrt{s}$ = 13 TeV [85] and for the tau lepton identification with a DeepTau algorithm [73]. Below, the most recent overview of the HPS algorithm is provided with relevant references to the original implementation.

As it was previously mentioned, the main challenge in reconstructing hadronic tau decays is that of efficiently distinguishing them from a large amount of jets originating from quarks or gluons. However, hadronic decay products of the tau are usually more collimated compared to those of the QCD jets. In addition, $\pi^0$ in the final state coming from intermediate $\rho(770)$ or $a_1(1260)$ resonances provide a unique handle to identify genuine $\tau_\mathrm{h}$ as well as its corresponding decay modes (DM). In this work, DMs (Table 4.1) are enumerated according to the formula: $\mathrm{DM} = 5 \times (N(\mathrm{h}^\pm) - 1) + N(\pi^0)$, where $N(\mathrm{h}^\pm)$ and $N(\pi^0)$ are the numbers of the reconstructed (or identified, depending in the context) charged prongs and strips/$\pi^0$, respectively.

Motivated by these observations, the HPS algorithm starts from constructing so-called *strips*, which serve as a proxy for $\pi^0$ particles. In the detector, a $\pi^0$ promptly decays into a pair of photons, which consequently, due to a sizeable amount of the tracker material, are very probable to convert to a pair of electrons, which can furthermore radiate bremsstrahlung photons, etc. In the presence of the magnetic field, the electrons trajectories are bent and therefore, on the ECAL surface in the $\eta - \phi$ plane, the clusters associated to $\pi^0$ decay products have an extended "strip" shape in the $\phi$ direction.

In order to construct a strip, an iterative clustering procedure with the following steps is performed:

1. In an event, hadronic jets are reconstructed by clustering PF particles using the anti-$k_T$ algorithm [86] (Sec. 5.3.4) with the distance parameter $\Delta R = 0.4$. For each jet, all PF particles in the cone of radius $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} = 0.5$ around the axis are passed as an input to the next step.

2. Within a jet, a strip is seeded by a highest $p_\mathrm{T}$ photon or electron that is not yet included in any strip. The algorithm proceeds with a one-by-one aggregation of electrons/photons with $p_\mathrm{T} > 0.5$ GeV within a $(\Delta\eta, \Delta\phi)$ window in the $\eta$-$\phi$ plane of the dynamically adjusted size (originally, of the fixed size [84]). The size of the strip window is a parametrised

function of $p_T$ of the strip at the current iteration and the $e/\gamma$ to be included in the strip [85]. The functional form is derived from simulated single tau events with a uniform $p_T$ spectrum with the goal of capturing 95% of possible $e/\gamma$ in $\tau_h$ decay products. In case of adding an $e/\gamma$ candidate to the strip, its position is recomputed as a $p_T$-weighted average of the coordinates in the $\eta - \phi$ plane of all the strip's constituents, and the strip momentum is set to a sum of the strip's constituents momenta. The procedure is terminated if there is no other $e/\gamma$ within a $(\Delta\eta, \Delta\phi)$ window and the clustering of a new strip continues with selecting a new seed.

3. For each jet seed, $\tau_h$ hypotheses are formed by combining reconstructed strips with the charged PF candidates. Combinations are formed on the basis of decay modes to be targeted: $h^\pm$, $h^\pm\pi^0$, $h^\pm\pi^0\pi^0$, $h^\pm h^\mp h^\pm$ , $h^\pm h^\mp h^\pm\pi^0$, $h^\pm h^{\pm/\mp}(\pi^0)$, where $\pi^0$ represents a reconstructed strip, $h^\pm$ a charged PF candidate and the last category targets $\tau^- \rightarrow h^- h^+ h^- \pi^0$ with one of the charged hadrons ($\pi^0$) escaping detection. The latter two were included into the reconstruction workflow together with the DeepTau algorithm [73]. In order to be assigned to a DM category, each combination is required to pass a mass window constraint to be compatible with the corresponding intermediate resonance (Table 4.1). Originally, the mass window was statically defined but later it was updated to be dynamically dependant on the strip $p_T$. In the following, $h^\pm\pi^0$, $h^\pm\pi^0\pi^0$ DMs are analysed together and referred to as $h^\pm\pi^0$.

4. Among the $\tau_h$ hypotheses formed at the previous step, a set of further requirements is applied. $\tau_h$ candidates should have a charge $\pm 1$, except for the DMs with the missing charged hadron, where the $\tau_h$ charge is set to the charge of the charged hadron with the highest $p_T$. All reconstructed $h^\pm$ and strips in the combination should be located within the tau signal cone defined by the radius $R_{sig} = 3.0/p_T$ (GeV), limited to the range 0.05-0.10, with respect to the $\tau_h$ momentum. Finally, for each seeding jet a single $\tau_h$ candidate with the highest $p_T$ is selected.

Overall, more than a half of each of the most significant $\tau_h$ decay modes ($h^\pm$, $h^\pm\pi^0$, $h^\pm h^\mp h^\pm$) is reconstructed in the targeted DMs (Fig. 4.4). Although $h^\pm h^{\pm/\mp}(\pi^0)$ category helps to recover 19% (13%) of $h^\pm h^\mp h^\pm$ ($h^\pm h^\mp h^\pm\pi^0$) DMs, it is not considered in the main $\tau_h$ reconstruction routine due to its large charge mis-assignment probability. Despite the fact that DM reconstructed efficiencies are naturally bounded by the 90% efficiency of the charged track reconstruction and even lower efficiency for photons coming from $\pi^0$ decays, one can observe that there is still room for improvement in the reconstruction of all DMs. This is particularly true for the DMs with one charged prong $h^\pm$ and at least one $\pi^0$, where the HPS algorithm fails to reconstruct 25% of these DMs, which amounts to $\approx$8% of all possible tau decays. Therefore, this motivates future studies in the direction of improvement of the HPS algorithm.

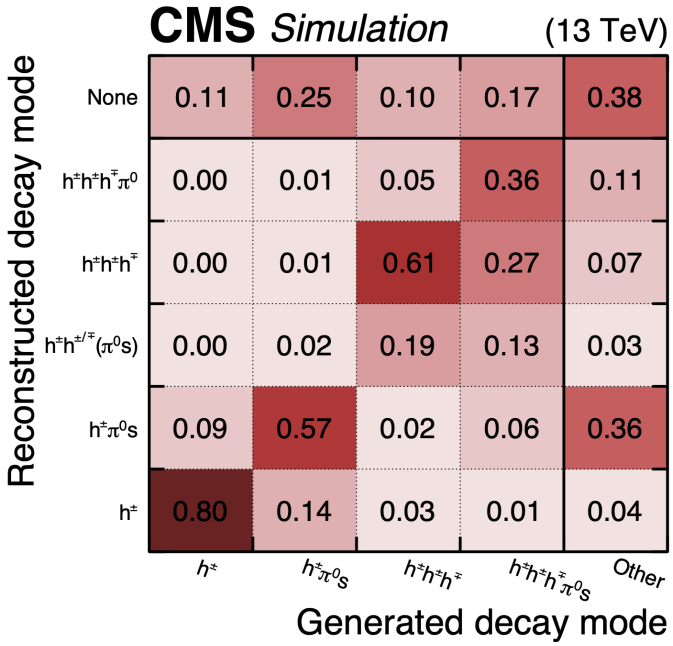


Figure 4.4: Fraction of $\tau_h$ candidates with a given generated decay mode to be reconstructed by the HPS algorithm in different decay modes [73].

## 4.3 Identification in CMS

Conceptually, the reconstruction step, starting from the PF algorithm (Section 4.2.1) and going hierarchically to more complex algorithms, e.g. jet clustering or the HPS algorithm (Section 4.2.2), aims at providing physics objects as inclusively as possible, i.e. maximising the efficiency of capturing original genuine particles. This approach inherently creates a collection of physics objects which is not pure in the objects of interest and is contaminated by background objects. Therefore, an additional step is needed to refine the purity of the collection.

This step is usually referred to as *identification* (ID), and its goal is to identify the types of objects appearing in the collection of reconstructed objects among the categories which are expected. This two-staged "RECO-ID" paradigm of building physics objects has been a standard in high-energy physics for years. However, with the emergence of powerful ML techniques, novel end-to-end approaches unifying two steps into a single one proved to be a promising and efficient solution to the problem of reconstructing physics objects [87, 88].

In the RECO-ID paradigm, ML-based algorithms have also proved to bring significant improvement to the ID step. The historical evolution pattern of ID methods is moving from a so-called *cut-based* (or rule-based) set of criteria to algorithms based on *linear classifiers* or

an ensemble of *decision trees* and then finally to algorithms based on *Deep Learning* (DL). The hadronic decays of the tau lepton also fit into this historical pattern, where the RECO step with the HPS algorithm was initially followed by a set of isolation criteria targeting predefined misidentification probabilities of $\tau_h$ against quark/gluon jets [83]. Later on, algorithms based on boosted decision trees (BDT) were introduced [84, 85] each trained to distinguish $\tau_h$ from either jets, or electrons, or muons. Lastly, a DeepTau algorithm [73] combined previously separate classifiers into a single neural network, providing an excellent discrimination power between $\tau_h$, jets, electrons and muons altogether.

The DeepTau architecture in its original implementation (Section 4.3.1), referred to as DeepTau v2.1, showed a significant improvement in $\tau_h$ identification against jets and leptons, compared to the previous approaches. Building upon this milestone, several improvements have been made in the context of the Run 3 preparation as outlined in Section 4.3.2, with the corresponding model being referred to as DeepTau v2.5.

### 4.3.1 DeepTau v2.1

There is one important aspect, in addition to the already mentioned unification of jet, electron and muon discriminants, which motivated a switch towards more advanced techniques for $\tau_h$ identification – the usage of low-level information. While hand-crafted high-level variables (also called *features*), provided as an input to a BDT, generally encapsulate the object to be identified, they are still limited in the representation power by the domain knowledge of the one who designed them. Since jets, being an input to a given model for $\tau_h$ identification, inherently exhibit complex hadronisation patterns, it is expected that their behavior cannot be fully described in terms of only several variables.

In the field of Computer Vision (CV) it has been shown that Convolutional Neural Networks (CNN) trained on images learn notions of growing complexity, starting from simple patters at the first layers and capturing more complex abstract concepts at the deeper layers [89]. Since pixels in the image carry only low-level intensity information, one can therefore view the process of training a CNN model as an *automated feature engineering*: more complex features are automatically learnt based on the low-level inputs. Moreover, the performance of ML models has been shown to increase as the model size grows [90–96]. That hints towards a large scope of high-level features which models can learn without explicit guidance. Furthermore, it is yet to be understood if (and how) it is possible to design such automatically learnt features manually.

One of the perspectives on a particle detector is when it is viewed as a "camera" imaging collisions. That makes it natural to use an *image representation* to describe the physics objects and the activity in the detector [97–99]. Despite the fact that this representation comes with certain limitations (Section 4.3.3), it has proved to be very performative in tasks like jet tagging

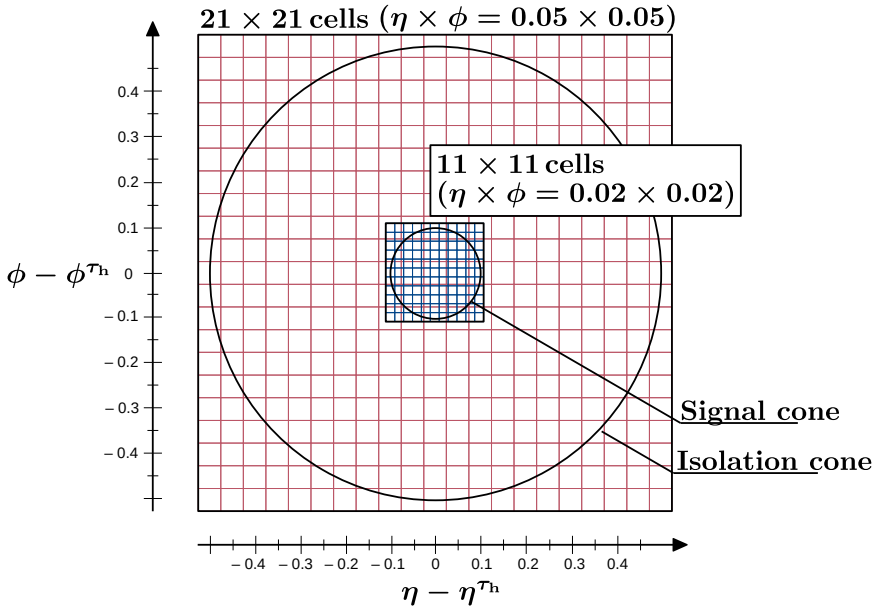


Figure 4.5: The grid representation in $\eta$-$\phi$ plane passed as the input to the DeepTau model [73]. The signal cone ($R = 0.1$) and the isolation cone ($R = 0.5$) are also shown to motivate the choice of the inner and outer grids. The former enters the isolation requirement on the PF candidates within the HPS algorithm and has a higher cell granularity to better capture the hadronisation activity within the core of the $\tau_h$ candidate, especially in boosted scenarios. The latter enters in the computation of the high-level isolation variables used for the training and aims to capture the hadronisation activity at a larger scale to specifically identify quark and gluon jets.

[100–102], particle reconstruction and identification [103–106], and particle shower generation [107–109].

All these aspects motivate the usage of convolution layers as the main building blocks of the DeepTau architecture and an image-like structure of the $\tau_h$ **input representation** (Fig. 4.5). The latter is constructed by defining in $\eta$-$\phi$ space an inner grid with $11 \times 11$ cells of size $0.02 \times 0.02$, and an outer grid with $21 \times 21$ cells of size $0.05 \times 0.05$ (in the $\eta/\phi$ units of measurement). The grids overlap and are centered around the HPS-reconstructed direction of flight of the $\tau_h$ candidate (Section 4.2.2). Seven types of particles in the vicinity of the $\tau_h$-axis are taken as an input:

- PF-reconstructed: muons, electrons, photons, charged hadrons, and neutral hadrons (Section 4.2.1).
- Standalone-reconstructed (RECO): electrons, muons.

The latter category uses dedicated standalone reconstruction algorithms which provide additional information about electrons and muons compared to those available from the PF algorithm. Each of the particles is attributed to a cell on both inner and outer grids according to its position in $\eta$-$\phi$ space, and the corresponding cell is filled with the features specific for a given particle type. Generally, the features describe the track quality, the quality of the associated PV or SV, the particle kinematics, the calorimeter and PU information. If several particles which are attributed to a grid of the given type {inner, outer} $\times$ $\{e^{\pm}/\gamma, \mu^{\pm}, \mathrm{h}^{\pm}/\mathrm{h}^{0}\}$ (described below) enter the same cell, the features of the one with the highest $p_{\mathrm{T}}$ are filled into the cell.

In addition, high-level features are also provided as an input, as described below, to improve the discriminating power of the model. These handcrafted variables, describing the $\tau_{\mathrm{h}}$ isolation, kinematic properties, associated vertex information, information about the associated strips, have been successfully used previously for $\tau_{\mathrm{h}}$ identification with the MVA classifiers. Although in theory, the model should be able to learn these variables, in practise this is often not the case due to a limited number of training data. Therefore, these variables are added explicitly to augment the model with expert knowledge.

The overall **architecture** is illustrated on Fig. 4.6. The model hyperparameters are described in detail in the original paper and below a conceptual overview of the model structure is provided. It starts from three streams, each processing its inputs independently:

- *Global:* a set of fully connected layers which processes high-level features.
- *Inner:* a set of one-dimensional (1D) followed by two-dimensional (2D) convolutional layers which processes inputs from the inner cone around the reconstructed tau direction of flight.
- *Outer:* a set of 1D followed by 2D convolutional layers which processes inputs from the outer cone around the reconstructed tau direction of flight.

1D section of both inner and outer streams are further split into three subsections individually processing three particle blocks:

- $e^{\pm}/\gamma$: to process inputs from combined PF electrons, PF photons, RECO electrons.
- $\mu^{\pm}$: to process inputs from combined PF muons and RECO muons.
- $\mathrm{h}^{\pm}/\mathrm{h}^{0}$: to process inputs from combined PF charged hadrons and PF neutral hadrons.

After being processed individually, the three particle blocks are concatenated and passed to another set of 1D convolutional layers before being passed to a 2D section. In general, the idea of using 1D convolutions is to encode input features on the per-particle level into a more

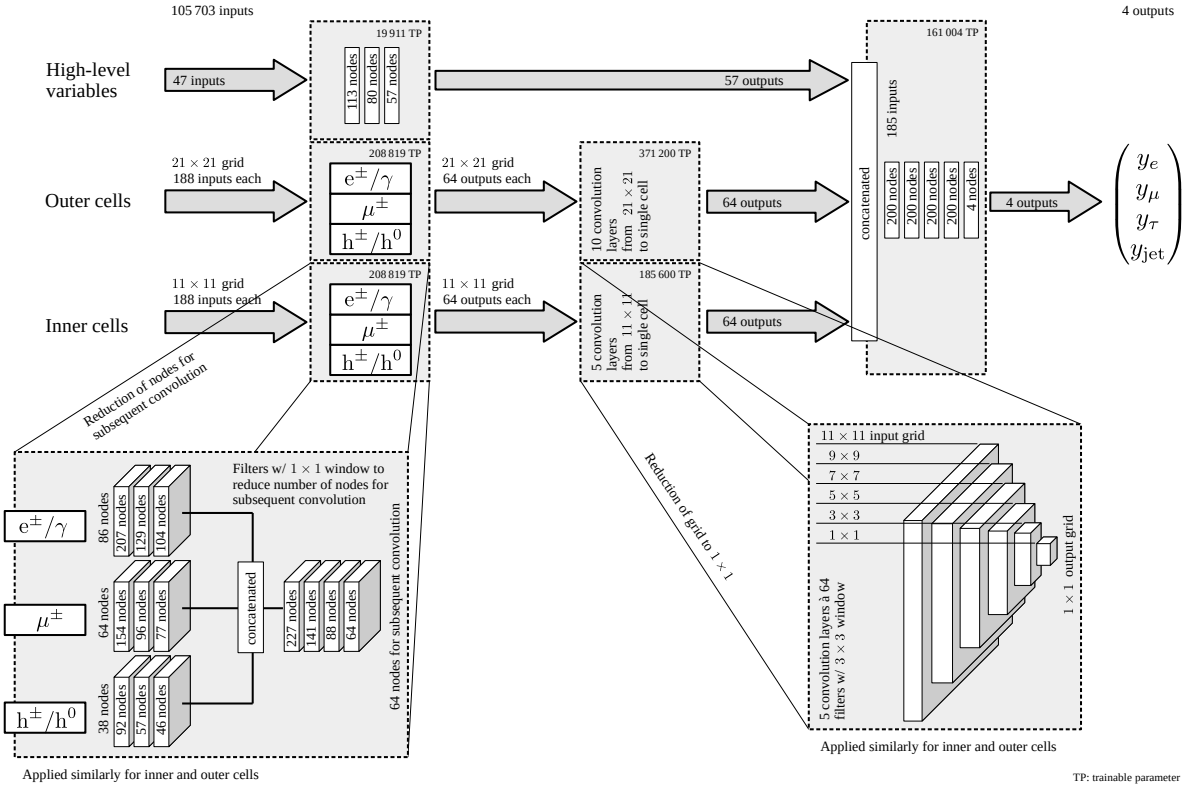


Figure 4.6: The DeepTau v2.1 architecture. Three processing streams (inner, outer, global) are illustrated as arrows, three particle subsections ($e^{\pm}/\gamma$, $\mu^{\pm}$, $\mathrm{h}^{\pm}/\mathrm{h}^{0}$) per inner/outer streams are illustrated as rectangular blocks in white. Additionally, the detailed information about the number of input features per particle block/stream, the hyperparameter values, the evolution of the input tensor shape at various stages within the model as well as the number of trainable parameters (TP) for the different subsections of the model is provided.

compact representation compared to the dimensionality of the input space. Otherwise, the usage of 2D convolutional layers directly on the input feature space would make the training task computationally hard to perform. Lastly, for each of the inner and outer streams, the 1D section is followed by the 2D section, where $3\times3$ filters extract spacial correlations between cells across the grid while also downsampling the spatial dimensions from $11x11$ ($21 \times 21$) cells for the inner (outer) grid to a single cell (Fig. 4.6).

Overall, the three-stream part of the DeepTau architecture can be viewed as an *encoder*, which extracts high-level features from the low-level ones while operating on a physically-motivated representation of $\eta$-$\phi$ space. From this perspective, the following processing part of the architecture can be viewed as a *decoder*, which maps the learned features for a given $\tau_{\mathrm{h}}$ candidate to a class probability. First, it concatenates extracted features from the inner and

outer streams with the handcrafted high-level features after being processed in the global stream. Second, it processes them by a set of fully-connected layers, finally followed by a fully-connected layer with four output nodes with a softmax activation function. The latter outputs the probability $y_\alpha$ of the given $\tau_h$ candidate to belong to one of the four classes: electron, muon, genuine $\tau_h$, quark or gluon jet.

In order to measure the model performance and also derive working points (WPs), the final discriminators against electrons, muons and jets are defined as:

$$D_\alpha(\boldsymbol{y}) = \frac{y_\tau}{y_\tau + y_\alpha}, \tag{4.1}$$

where $\boldsymbol{y} = (y_e, y_\mu, y_\tau, y_{\text{jet}})$ is the output of the softmax layer of the model.

To perform the training, a **loss function** is constructed and minimized with Nesterov-accelerated adaptive momentum estimation (NAdam) [110]. The loss function consists of three terms (Appendix A.2):

1. Binary cross-entropy term for $\tau_h$ class against all the other ($e$, $\mu$, jet) classes combined.
2. Focal-loss [111] term for $\tau_h$ class against all other classes combined.
3. Focal-loss terms separately for each of the ($e$, $\mu$, jet) classes, smoothed by a step function to target only $\tau_h$ candidates which are likely to be classified as $\tau_h$.

The composition of the loss function is designed to guide the training to have better performance in the regions which are important for most of the analyses. Namely, it aims to provide better performance in the range 50-80% of $\tau_h$ efficiency, while on the other hand to not focus on the identification of background classes in the high-purity regime.

The **data set** used for the training consists of events from the following simulated processes: Z+jets (NLO), W+jets, $t\bar{t}$, $Z' \to \tau\tau$, $Z' \to ee$, $Z' \to \mu\mu$, (with $m(Z')$ ranging from 1 to 5 TeV),and QCD multijet production. For testing, additional event samples from $H \to \tau\tau$ and Z+jets (LO) are used. To ensure that no additional biases are introduced, the $\tau_h$ candidates are sampled from the input samples such that the contribution of each class ($e$, $\mu$, $\tau_h$, jet) in different ($p_T$, $\eta$) bins is the same. Furthermore, during the training additional weights are applied to make the distribution of classes uniform within each ($p_T$, $\eta$) bin. In total, around 140 million $\tau_h$ candidates are used for the training, while around 10 million are used for the validation. The model implementation and the training are done using the TensorFlow library [112].

Overall, large gains in the **performance** are reported across various regions of the phase space with respect to the previous cut-based/tree-based $\tau_h$ identification approaches (Fig. 4.7). In summary, at a given $\tau_h$ efficiency, the DeepTau discriminator consistently reduces the misidentification probability against jets by more then a factor of 2. Against electrons, the improvement

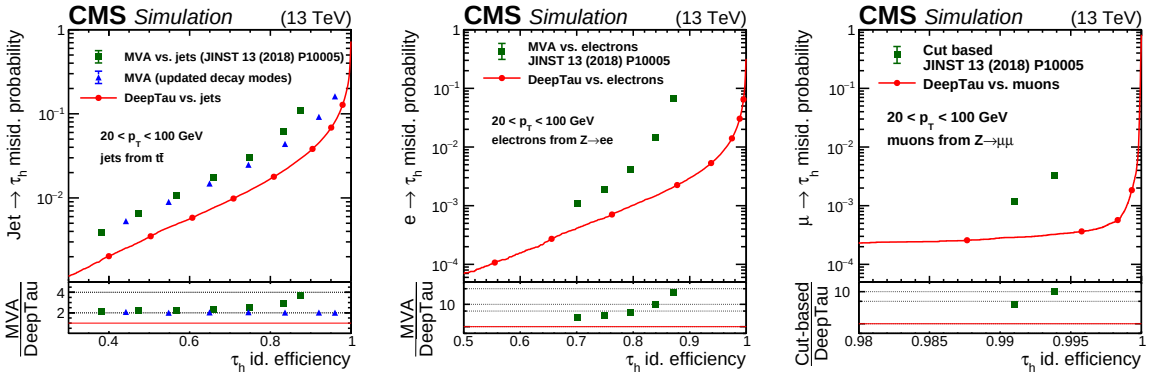


Figure 4.7: Efficiency for jets (left), electrons (middle), and muons (right) versus efficiency for genuine $\tau_\mathrm{h}$ to pass various tau identification discriminators as well as the corresponding $D_\alpha$ discriminators (Eq. 4.1). $\tau_\mathrm{h}$ are selected with the requirements $20 < p_\mathrm{T} < 100$ GeV, $|\eta| < 2.3$, and should additionally pass the loosest WPs against other tau types, with the thresholds as defined in the original paper. For evaluation, genuine $\tau_\mathrm{h}$ are taken from the $H \to \tau\tau$ event sample, electrons and muons are taken from the $Z \to ll$ event sample, and jets are taken from the $t\bar{t}$ event sample.

in misidentification probability ranges from by a factor of 2 for a $\tau_\mathrm{h}$ efficiency of 70% and goes up to a factor of 10 for $\tau_\mathrm{h}$ efficiencies larger than 88%. Against muons, the misidentification probability is reduced by almost a factor of 10 in the region of $\tau_\mathrm{h}$ efficiency around 99%. This expected performance improvement, obtained from the simulated events, was successfully validated on collision data, therefore establishing a new milestone for the tau identification task.

### 4.3.2 DeepTau v2.5

As the LHC and the CMS detector in particular are continuously being upgraded and are moving on with collecting new data, every particle identification model deployed in production largely benefits from retraining on the data corresponding to the new detector conditions. It ensures that the quality of the model does not degrade due to the lack of robustness to the gradual changes in the underlying data (so-called *data drift*) and the model performance stays at the nominal level.

This applies also to DeepTau v2.1 (Sec. 4.3.1), which is originally trained on the data samples simulated and reconstructed with the 2017 data-taking conditions. With the start of the new Run 3 period, it is expected that the model performance will become suboptimal on the newly collected data as well as on the Run2 Ultra Legacy (UL) reprocessed data. Therefore, a dedicated DeepTau retraining has been performed with the following set of updates, further described in more details:

- Data preparation:
  - Updated data samples,
  - Shuffle & Merge procedure,
  - Feature preprocessing.
- Training:
  - Parallel data loading,
  - Hyperparameter optimisation,
  - Adversarial training.

### Data preparation

The global tau identification task remains unchanged w.r.t. DeepTau v2.1 in terms of the input and output spaces: for a given HPS-reconstructed $\tau_h$ candidate the goal is to predict its probability to be either of (jet, $e$, $\mu$, $\tau_h$) classes, with the class assignment criteria described below. The $\tau_h$ candidate is represented with the surrounding PF and RECO electron/muon candidates placed onto inner and outer grids in $\eta$-$\phi$ space centered around $\tau_h$ axis, with the grid parameters being the same as defined in Sec. 4.3.1. The set of input features describing particle collections is the same as it was used for the DeepTau v2.1 training.

In order to compose the training data set, $\tau_h$ candidates of the given classes are sourced from the following data samples, simulated under the detector conditions corresponding to the 2018 year:

- DY: inclusive, jet/$H_T$-binned $\Rightarrow$ jet, $e$, $\mu$, $\tau_h$.
- QCD: $p_T$/$H_T$-binned $\Rightarrow$ jet.
- $t\bar{t}$: leptonic, semileptonic $\Rightarrow$ jet, $e$, $\mu$, $\tau_h$.
- $t\bar{t}$: fully-hadronic $\Rightarrow$ jet, $e$, $\mu$.
- W+jets: jet/$H_T$-binned $\Rightarrow$ jet, $e$, $\mu$, $\tau_h$.
- Higgs: $ZH \to \tau\tau$, $W^{\pm}H \to \tau\tau$, $H \to \tau\tau$ (vector-boson fusion), $HH \to bb\tau\tau$ (gluon-gluon fusion) $\Rightarrow \tau_h$.
- $Z' \to e^+e^-$: $m(Z') \in [1000, 4000]$ GeV $\Rightarrow e$.
- TauGun: $p_T(\tau) \in [15, 3000]$ GeV $\Rightarrow \tau_h$.

A $\tau_{\mathrm{h}}$ candidate is selected and assigned to one of the classes based on the following criteria:

- $e(\mu)$: matching to a prompt $e(\mu)$ or $\tau_{\mathrm{e}}(\tau_{\mu})$ at the generator level within a cone of radius $R = 0.2$. The associated generated lepton should pass the requirement $p_{\mathrm{T}}^{\mathrm{vis}}(l) > 8$ GeV.
- $\tau_{\mathrm{h}}$: matching to a hadronically decaying tau lepton at the generator level within a cone of radius $R = 0.2$. The associated generated lepton should pass the requirement $p_{\mathrm{T}}^{\mathrm{vis}}(l) > 15$ GeV.
- jet: absence of an associated generator-level lepton and matching to a generator-level jet within a cone of radius $R = 0.4$. Additionally, in the selection procedure jets with the reconstructed $p_{\mathrm{T}} < 80$ GeV are randomly rejected with the probability $p = 1 - \exp\{-0.05 \cdot (80 - p_{\mathrm{T}})\}$ in order to balance the contributions from low- and high-$p_{\mathrm{T}}$ ranges in the training data set.

After the initial selection of $\tau_{\mathrm{h}}$ candidates from the data samples, a training data set has to be formed as a set of $\tau_{\mathrm{h}}$ candidates grouped into batches, later fed into the model. In order to have a stable training procedure, it is beneficial for batches in the data set to have the following properties:

- Homogeneous: every two randomly selected batches are statistically similar.
- Unbiased: every batch should be sampled in an unbiased way from the predefined target distribution $p_{\mathrm{target}}(x_s)$ over a set of *spectrum variables* $x_s$. In the case of DeepTau v2.5, these are selected to be: $x_s = \{p_{\mathrm{T}}(\tau_{\mathrm{h}}), \eta(\tau_{\mathrm{h}}), \mathrm{class} \in \{e, \mu, \tau_{\mathrm{h}}, \mathrm{jet}\}\}$. The target distribution is chosen to be a uniform histogram (referred to as a spectral histogram) in a predefined binning for the spectrum variables.

The following procedure, referred to as **Shuffle & Merge** (S&M), is used to form the training data set satisfying these properties. Performed in a memory-efficient manner, it allows for a better control over the spectrum variables compared to the balancing approach used for the DeepTau v2.1 training. It follows a stochastic approach, where a data group is firstly sampled from the categorical distribution, where the categories are the groups of the data samples used for the training as listed above, and the probability of sampling from a given data group is proportional to the number of entries in the group. Secondly, a random $\tau_{\mathrm{h}}$ candidate is sampled from the data group, and it is kept for the training with the probability:

$$p(\tau_{\mathrm{h}}) \sim \frac{p_{\mathrm{target}}(x_s(\tau_h))}{N(\mathrm{bin}, \mathrm{group})}. \tag{4.2}$$

Here, $x_s(\tau_h)$ are the values of the spectrum variables for the given $\tau_{\mathrm{h}}$ candidate, $N(\mathrm{bin}, \mathrm{group})$ is the total number of events in a bin of the spectral histogram corresponding to the given

$x_s(\tau_h)$ value and the data group. An additional correction to the probabilities in the spectrum histograms is made to keep the ratio between the number of $\tau_\mathrm{h}$ candidates in the last $p_\mathrm{T}$ bin and the other $p_\mathrm{T}$ bins per each class more than 0.001.

The procedure is distributed in a parallel manner and for a given thread a random subset of $\tau_\mathrm{h}$ candidates of each of the data group is provided. The condition to terminate the procedure for a thread is when there is no more $\tau_\mathrm{h}$ to select from one of the data groups. Upon the completion of the S&M procedure, a Kolmogorov-Smirnov test is performed in order to validate the compatibility of randomly selected subsets of the formed data set between each other. Overall, an acceptable level of homogeneity is observed across the spectrum variables for each of the data group. The final data set comprised of around 100M $\tau_\mathrm{h}$ candidates, with 70% of them used for the training and 30% for the validation. This data set is further referred to as a *S&M data set*.

The last step before the training is the **feature standardisation**. Its goal is to bring the values of the input features to a common domain, which is achieved by subtracting the mean and dividing by the standard deviation, followed by a clamping to a range $[-5, 5]$. Categorical features and features derived from $p_\mathrm{T}, \eta, \phi$ are normalised by clamping to a predefined range, followed by a mapping to a range $[-1, 1]$. The clamping procedure also removes outlying values, which together with the feature standardisation provides stable gradient updates during the training. The mean and the standard deviation for each of the features are derived in a cumulative manner by aggregating the sums and the counts of the feature values over the input data set. The validation is performed to check that the clamping procedure does not distort the original distributions of the input features.

### Training

After the training data set is formed, for each $\tau_\mathrm{h}$ candidate an image representation is constructed as described in Sec. 4.3.1. The corresponding tensors are combined into batches which are subsequently fed into the model during the training. Since the batch shaping procedure is performed on the fly, there is a challenge of how to make the data loading procedure time-efficient during the gradient updates. This is solved by introducing a multiprocessing queue and a set of workers filling/taking batches to/from the queue, thus allowing for a concurrent loading of batches into the model as the training is on-going. A PyTorch [113] implementation of the multiprocessing queue is used as providing a factor of 3 speed-up compared to a default implementation in Python libraries. Moreover, a better scaling of computational performance with the number of workers and the queue size is observed for the PyTorch implementation.

The **hyperparameter optimisation** is performed in several stages in order to select the model with the most optimal performance. For the first stage (stage 0), the training of each of the trials is performed on 20% of the S&M data set for one epoch, which takes approximately one day

on NVIDIA™ Tesla V100. For validation, another 20% of the S&M data set is used. The same loss function is used as for the DeepTau v2.1 training (Appendix A.2) and it is minimised with a NAdam optimiser with the following parameters: learning rate = $10^{-3}$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 10^{-7}$. Training weights, derived from the spectral histograms of the S&M data set, are added to the loss function in order to make the contributions from different ($p_T$, $\eta$, class) bins uniform. The hyperparamater values varied during the first stage of the optimisation as well as the resulting performance are summarised in Table 4.2. The trial names correspond to the following configurations of the hyperparameters (changes are with respect to the baseline):

- Baseline: DeepTau v2.1 architecture (Sec. 4.3.1).
- (1): number of 2D filters is reduced by 1.8 from one layer to another (*) + number of nodes in the decoder's dense layers form a progression from n/2 to 32 with reduction factor 2, where n is the input dimensionality to the decoder (**).
- (2): (*) + (**) + number of filters in 1D convolutions operating on the merged $e^{\pm}/\gamma$, $\mu^{\pm}$, and $h^{\pm}/h^0$ streams forms a progression: $227 \rightarrow 141 \rightarrow 128$.
- (3): (*) + (**) + number of filters in 1D convolutions operating on the merged $e^{\pm}/\gamma$, $\mu^{\pm}$, and $h^{\pm}/h^0$ streams forms a progression: $227 \rightarrow 141 \rightarrow 88 \rightarrow 64 \rightarrow 40 \rightarrow 32$.
- (4): (**) + number of 2D filters in each layer is reduced by 2 from one layer to another.
- (5): (**) + number of 2D filters in each layer is reduced by 1.6 from one layer to another.
- (6): (*) + number of nodes in the decoder's dense layers forms a progression from n/3 to 32 with the reduction factor 3.
- (7): (*) + number of nodes in the decoder's dense layers forms a progression from n/1.5 to 32 with the reduction factor 1.5.

Overall, the architecture with the hyperparameter values of the DeepTau v2.1 model gives the best performance if measured by the value of the loss function on the validation data set averaged across 5 independent trials.

While the first stage involves the variations of the network structure, the second stage (stage 1) targets the choice of the optimiser and the learning rate. Furthermore, since at this point 70% (30%) of the S&M data set is used for the training (validation), the goal is to reach better convergence compared to the short training of the first stage. The best performing model from the first stage is therefore taken and the training is continued for additional two epochs with the following optimisers and learning rates values being probed:

- NAdam: $10^{-3}$ (acbad), $10^{-4}$ (5371f), $10^{-5}$ (08f84), $10^{-3}$ w/ cross-entropy (6945b).

Table 4.2: Performance comparison for the models from the stage 0 of the DeepTau v2.5 hyperparameter optimisation. The changes to the hyperparameters in the trials are described in the text and the number of trainable parameters (TPs) is shown in the middle column. Each trial is run 5 times and the mean and standard deviation values of the loss function on the validation set are reported in the right column.

| Trial | TPs $\times 10^3$ | $L_{\text{val}}$ |
|---|---|---|
| Baseline | 1,151 | $\mathbf{0.273 \pm 0.002}$ |
| (1) | 1,197 | $0.277 \pm 0.003$ |
| (2) | 3,211 | $0.274 \pm 0.002$ |
| (3) | 653 | $0.287 \pm 0.008$ |
| (4) | 768 | $0.280 \pm 0.002$ |
| (5) | 2,570 | $0.287 \pm 0.019$ |
| (6) | 1,097 | $0.276 \pm 0.004$ |
| (7) | 1,383 | $0.324 \pm 0.080$ |

- Adam [114]: $10^{-3}$ (38c05), $10^{-4}$ (59162).

where a unique hash value for the corresponding entry in the legend of Fig. 4.8 is specified in brackets. For all of the trials, the loss function from the first stage is used in the minimisation, except for the trial NAdam ($10^{-3}$), where the loss function consists only of the weighted sum of binary cross-entropy terms for $\tau_{\text{h}}$ versus the other classes.

To evaluate the performance at the second stage, receiver operating characteristic (ROC) curves are derived for each of the discriminators (Eq. 4.1) in bins of $p_{\text{T}}$, $|\eta|$ and HPS-reconstructed decay mode of the $\tau_{\text{h}}$ candidate. For evaluation, genuine $\tau_{\text{h}}$ candidates are sourced from the $\text{H} \to \tau\tau$ (gluon-gluon fusion production mode) data sample. Jets are sourced from the fraction of the semileptonic $t\bar{t}$ data sample not used in the training. Electrons and muons are sourced from the fraction of the DY data sample not used in the training. The results of the evaluation are shown on Fig. 4.8 for the region $p_{\text{T}}(\tau_{\text{h}}) \in [20, 100), |\eta(\tau_{\text{h}})| < 2.3, \text{DM}(\tau_{\text{h}}) \in \{0, 1, 10, 11\}$. While Adam ($10^{-3}$) performs the best in the classification against muons, it shows inferior performance compared to the other models in the classification against electrons (low tau efficiency region) and against jets. Likewise, the model used as the starting one for the second stage tuning (referred to as "default" on Fig. 4.8), shows the best performance against jets but performs 10-20% worse against electrons and muons. As a trade-off, the model which corresponds to the trial NAdam ($10^{-4}$) is chosen at this stage as the one compromising the performance against all of the three classes.

The last stage (stage 2) in the training pipeline introduces an **adversarial approach** to fine-tuning the model. The motivation for that is to eliminate the discrepancies between data and simulation observed in the high-score region of $D_{\text{jet}}$ for the model after the second stage of the

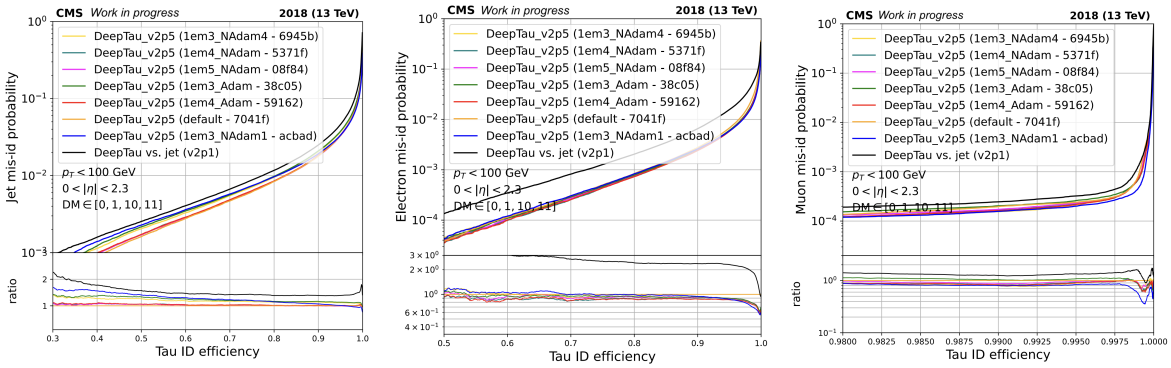


Figure 4.8: Efficiency for jets (left), electrons (middle), and muons (right) versus efficiency for genuine $\tau_h$ to pass the corresponding $D_\alpha$ discriminators for each of the trials at the second stage (stage 1) of the hyperparameter optimisation. "Default" is the model used as a starting model for all of the trials. The panel at the bottom of each of the figures shows the ratio of the ROC curves evaluated for each of the trial to the ROC curve evaluated for the starting model.

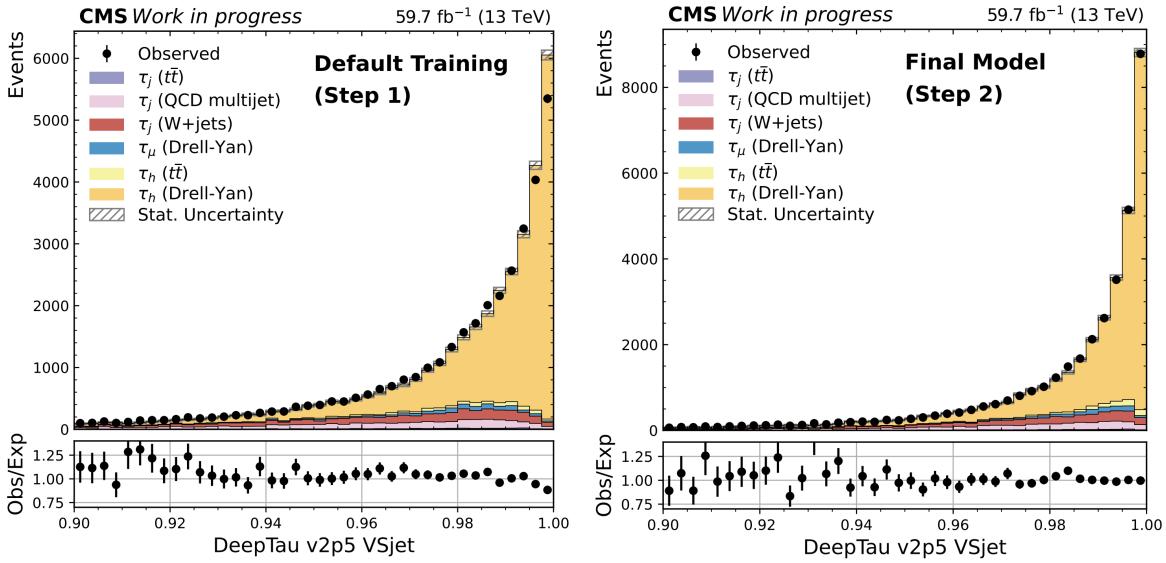


Figure 4.9: Comparison of data and MC simulation agreement for $D_{\text{jet}}$ variable in the adversarial region as defined in the text for DeepTau v2.5 at stage 1 (left) and DeepTau v2.5 (stage 2)

hyperparameter optimisation (Fig. 4.9, left). This can be viewed as a *domain shift* problem, meaning that the model performance does not transfer from the source domain on which it was trained (simulated data) to the target domain on which it is eventually being applied (collision data).

Approaching this problem from the perspective of domain adaption [115], one of the tech-

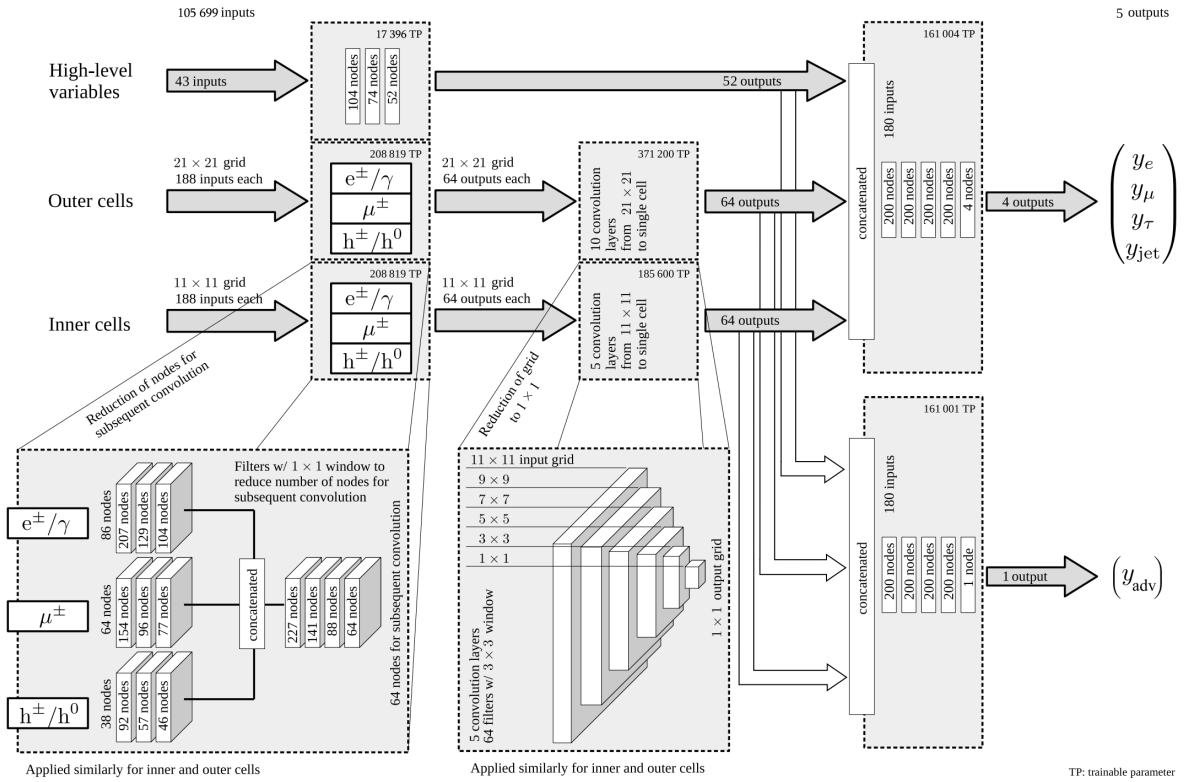


Figure 4.10: The DeepTau v2.5 architecture. The core encoder and decoder parts along with their hyperparameters are the same as in DeepTau v2.1. The key difference is an addition of another adversarial stream in parallel to the decoder which computes the probability of a $\tau_\mathrm{h}$ candidate to originate from either collider or simulated data.

niques to align the model's performance between the source and target domains is to augment the loss function with an additional adversarial term. Initially proposed in [116, 117], the technique showed an improved modelling of displaced jets in a search for new long-lived particles in CMS [118] without significant decrease in performance.

To achieve this goal, the DeepTau v2.5 (stage 1) model is extended to have an additional stream in parallel to the decoder (Fig. 4.10). This branch is responsible for producing predictions of whether a given $\tau_\mathrm{h}$ candidate originates from a collider data event or from a simulated data event. Conceptually, the idea is to continue training an augmented version of the model in a way that the decoder stream still tries to classify between the four classes as good as possible, while the adversarial stream tries to predict the underlying $\tau_\mathrm{h}$ domain as bad as possible.

This intuition is implemented by adding a binary cross-entropy over two classes (collider data vs. simulated data) $L_\mathrm{adv}$ as an additional negative term to the classification loss function

$L_{\text{class}}$ to penalise correct identification of the data domain:

$$L_{\text{tot}} = k_1 \cdot L_{\text{class}} - k_2 \cdot L_{\text{adv}}, \tag{4.3}$$

where $k_1, k_2 > 0$ are the hyperparameters to trade off between adversarial regularisation and classification performance. For the final model, $k_1 = 1$ and $k_2 = 10$ are chosen and each of the loss components receives its own Adam optimiser (for the classification term, the optimiser inherits its state from the original model). The initial learning rates are 0.001 for the classification component and 0.01 for the adversarial component and are decayed exponentially throughout the training. Lastly, the relative class importance constants in the classification loss (Appendix A.2) are modified: $[\kappa_e, \kappa_\mu, \kappa_\tau, \kappa_j] = \frac{4}{10}[1, 2.5, 5, 1.5] \rightarrow \frac{4}{14}[2, 5, 6, 1.]$.

In order to perform fine-tuning with the adversarial component, an adversarial data set is formed, consisting of an equal number of collider data and simulated events. The former are taken from 2018 collider data, while the latter are sourced from Drell-Yan, $t\bar{t}$, QCD, W+jets simulated samples with the same detector conditions. The events are required to pass the selection requirements of $\tau_\mu\tau_{\text{h}}$ channel used in this work (Sec. 6.1) with an additional requirement $D_{\text{jet}}(\text{v2.1}) > 0.9$, where the discriminator is taken from the DeepTau v2.1 model. The region corresponding to this selection is referred to as an adversarial region. The resulting data set comprises 1.9k batches of 100 $\tau_{\text{h}}$ candidates (50 from collider data, 50 from simulation).

The training step firstly proceeds with passing a batch from the S&M data set, used at the previous stages of training, and computing the gradients for the classification loss $L_{\text{class}}$ with respect to the encoder and decoder weights. Then, a batch from the adversarial data set is passed and the gradients for the adversarial loss $L_{\text{adv}}$ with respect to the encoder and the adversarial stream weights are computed. Next, the weights are updated with the computed gradients for each of the model parts according to Eq. 4.3.2 in order: decoder $\rightarrow$ adversarial stream $\rightarrow$ encoder.

The training proceeds until the convergence of the loss function (4.3.2) on the validation data set, with the final validation accuracy of data vs. MC classification task equal to 0.51. This indicates that the model reaches sufficient level of not being able to distinguish between the two domains. Moreover, the agreement of data with simulation improves significantly in the adversarial region for the stage 2 model, compared to the stage 1 model before adversarial fine-tuning (Fig. 4.9, right).

The model performance at stage 2 is evaluated analogously to stage 1. In addition to the simulated samples with the detector conditions of the 2018 year (referred to as Run 2 performance), the model is evaluated on Run 3 samples with the detector conditions corresponding to the early Run 3 data taking period. Electrons, muons and genuine $\tau_{\text{h}}$ candidates are sourced from DY sample, while jets are sourced from a $t\bar{t}$ (semileptonic) sample. The resulting ROC curves

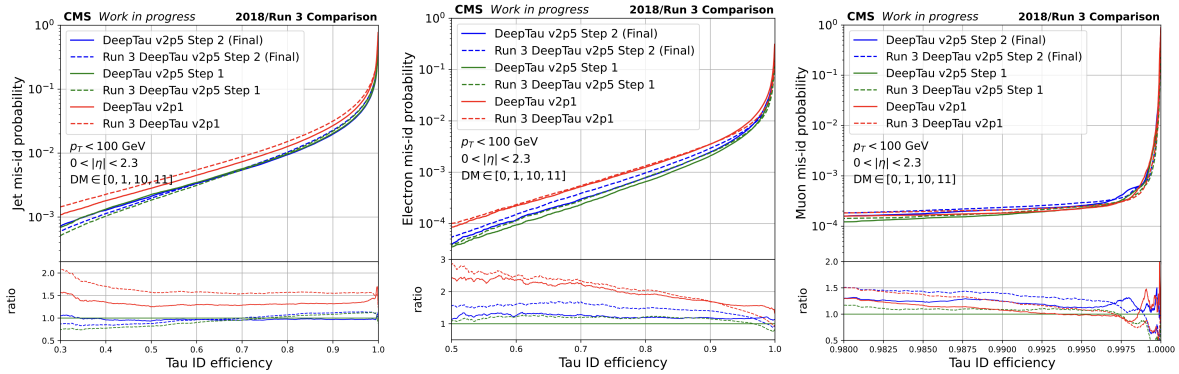


Figure 4.11: Efficiency for jets (left), electrons (middle), and muons (right) versus efficiency for genuine $\tau_h$ to pass the corresponding $D_\alpha$ discriminators at the third stage (stage 2) of the hyperparameter optimisation. Three models are compared: DeepTau v2.1 (red), DeepTau v2.5 at stage 1 (green), DeepTau v2.5 at stage 2 (blue). Solid lines correspond to the performance as measured on the samples with Run 2 detector conditions which are used in the training. Dashed lines correspond to the samples with Run 3 detector conditions. The panel at the bottom of each of the figures shows the ratio of the ROC curves evaluated for each of the trial to the ROC curve evaluated for the DeepTau v2.5 at stage 1 on Run 2 samples.

evaluated in the region $p_T(\tau_h) \in [20, 100), |\eta(\tau_h)| < 2.3, \text{DM}(\tau_h) \in \{0, 1, 10, 11\}$ are shown on Fig. 4.11. The performance of the model at stage 1 as well as those of DeepTau v2.1 is also shown. No significant degradation in performance is observed between the model performance at stage 1 and stage 2. The agreement between data and simulation is improved by 5-15% as can be seen from the data/simulation ratio in the two highest score bins (Fig. 4.9). This indicates the effectiveness of the procedure and motivates its generalisation to the discriminators against electrons and muons. However, it can be seen that the performance generally does degrade if the model is applied on the early Run 3 samples on which it was not trained. This motivates a dedicated training or fine-tuning of the model on those samples to keep the performance at the nominal level. Overall, the final DeepTau v2.5 model delivers a reduced fake rate at a given $\tau_h$ efficiency by 10-50% across the regions of interest and sets a new improved baseline for the tau identification task.

### 4.3.3 Tau Transformer

As mentioned in Sec. 4.3.1, an image representation and a traditional convolutional approach to process it come with certain limitations. Despite yielding good results across various research domains, for the specific task of the $\tau_h$ identification there is a set of design issues:

- **Information is not represented compactly.** On average, the DeepTau grid is filled

with zeros in 90% (99%) cases for the outer (inner) grids which makes the data loading procedure not memory-efficient. The only information stored in empty cell is an implicit positional one which can be passed to the model in a more efficient way.

- **Convolutional layers might not encode information optimally.** Since there is no explicit communication of positional/relational information to the model, one relies on the model as capable to optimally learn the relationships between particles in the spatial 2D frame. This might happen in the limit of infinite data, but in practise, other approaches to encode information can yield better results on the limited data sets.

- **Translational equivariance is not applicable.** The key feature of 2D convolutional layers is that they produce representations which are equivariant with respect to translational shifts. This is not applicable to the tau identification domain, where a symmetry breaking is induced with the choice of the image centering axis (HPS-reconstructed direction of flight) and furthermore with the topology of the CMS detector. Potentially, it may result in undesired behaviour with respect to small spatial perturbations of inputs.

- **Scaling with the number of PU interactions is limited.** Because of the finite cell size, in case of several particles entering the same cell only the one with the highest value of $p_{\mathrm{T}}$ is kept. On the one hand, it might provide a natural regularisation and improved robustness to the increased number of PU interactions. On the other hand, it assumes that the particles with higher $p_{\mathrm{T}}$ are more important for the $\tau_{\mathrm{h}}$ identification, which might not be the case and therefore a significant loss of information can take place.

To put the image representation into the context, there are ongoing studies of various representations in the particle physics area. Historically, representing activity in the detector as an image was one the first ideas together with a classical approach of using handcrafted features dating back to [119]. Then, partially due to the reasons described above, sparse representations were becoming more prominent. These include sets [120], sequences [121], graphs [122], and polynomials [123] with each of these representations coming up with its own way to extract information. Furthermore, physics-motivated representations in particular aiming to preserve the underlying symmetries were also proposed and studied [124–127].

### Architecture

In this work, a $\tau_{\mathrm{h}}$ candidate is represented as a set of input particles. Taking inspiration from a Natural Language Processing (NLP) domain, this representation is tightly linked with a sentence-based perspective, where a $\tau_{\mathrm{h}}$ candidate, being a set of particles (referred to as constituents or tokens), is viewed as a sentence consisting of multiple words with underlying grammar rules (for the $\tau_{\mathrm{h}}$ case, decay history) which are not observed directly. In order to extract information from

this representation, a self-attention mechanism is used as proposed in the original paper [128] introducing a Transformer model. The proposed concept of attention kick-started a revolution in the ML field due its dramatic improvement in the performance across multiple domains and due to its excellent scalability with the size of the input data set [129].

In HEP domain, models built around various implementations of attention also showed noticeable improvement [130, 131]. The most recent Particle Transformer (ParT) [132] model builds upon the original Transformer model and augments it with an interaction mechanism. Notably, the importance of using larger data sets for training in the jet tagging domain is additionally emphasized.

However, the unique feature of the $\tau_h$ identification task is the heterogeneity of the input space. Transformer models in the particle physics domain so far assumed that the inputs constitute only particles of one specific kind. Furthermore, no specific treatment of global variables is proposed. This is to be contrasted with the four collections used in the DeepTau v2.1 and v2.5 training to describe a $\tau_h$ candidate: global variables, PF-reconstructed particles, RECO electrons and muons. This can be viewed from a *multimodality* perspective, where an input object is described by several various modalities which cannot be *a priori* combined into a single one. The notion of multimodality is strictly speaking not fully applicable to the $\tau_h$ case, since RECO electrons/muons and PF constituents are both particles in their essence. However, they have various input features, which makes their treatment as of the same kind not straight-away possible. Furthermore, global variables certainly stand out from the "particle" modality. A multimodality perspective therefore provides a convenient language to describe the input $\tau_h$ representation, also for future studies where additional collections (for example, secondary vertices or tracker hits) can be included into the input representation to make it more informative.

A Particle Embedding module is therefore introduced (Fig. 4.12) in order to unify the four collections together into a single representation. The idea is to bring the dimensionality of tokens of each of the modality to a common one and combine the modalities before propagating them to attention layers. It is achieved by firstly using a categorical embedding of an additionally introduced modality variable (one-dimensional, categorical variable) with $N_{\text{in}} = 10$ input values (7 PF types: $e$, $\mu$, $\gamma$, $h^{\pm}$, $h^0$, HF tower identified as an hadron, HF tower identified as an EM particle; 1 RECO muon; 1 RECO electron; 1 global variables) to an output dimensionality $d_{\text{cat}} = 2$ (two-dimensional, real valued). Performed on a per token basis (embedding matrix is shared across the modalities), the result is concatenated to the other features and is processed by embedding blocks. Each embedding block is defined separately for each of the modalities and consists of two feed-forward layers with dimensionalities (number of output nodes) $d_{\text{ff}} = 256$ and $d_{\text{model}} = 64$ (Fig. 4.12), also operating on the per token basis. Afterwards, all the embedded tokens of all the modalities are concatenated together per $\tau_h$ candidate and passed through a dropout layer [133] with $p = 0.1$ to self-attention layers.

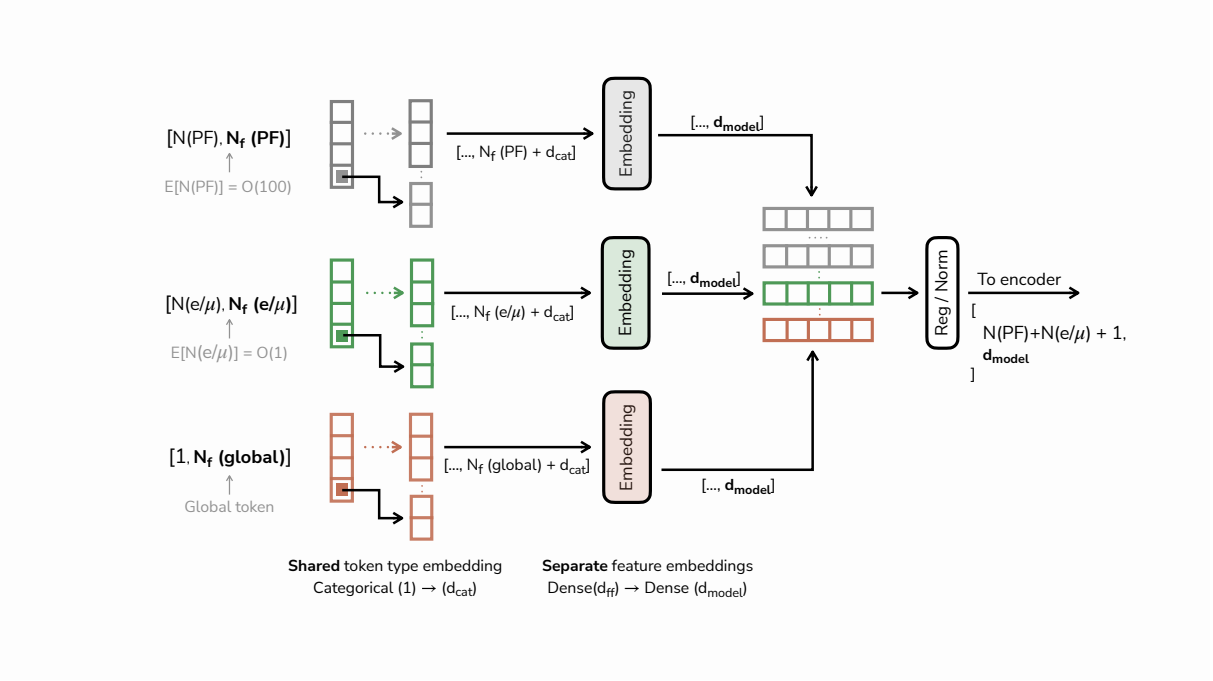


Figure 4.12: A Particle Embedding block. Processing streams for RECO electrons and RECO muons are illustrated together for visualisation purposes. For a single $\tau_h$ candidate each modality is represented as a tensor of the shape $[N(*), N_f(*)]$ and the change of the shape throughout the block is shown next to the arrows. Three tokens illustrating a vector of features are shown with the grey, green, brown colors corresponding to a PF constituent, a RECO electron/muon candidate and global variables, respectively. A dotted cell represents an additionally introduced modality variable. The values of the hyperparameters ($d_{cat}$, $d_{ff}$, $d_{model}$) are described in the text.

It is worth mentioning that in this approach global variables are treated as a "context" token: it is allowed to interact with the other constituents (both PF and RECO particles) on the equal basis as constituents interact with each other. This interaction is guided via attention mechanism and is learned by the model during the optimisation procedure. However, while the proposed approach of combining modalities is straight-forward in its intuition, it might not be optimally representing the underlying relationships between them [134]. This question of optimally encoding multiple modalities in the context of jet tagging is left for future research.

After the particle embedding, each $\tau_h$ candidate is represented as a tensor of the shape $[N(PF) + N(e) + N(\mu) + 1, d_{model}]$. This structure can now be processed with the encoder, which consists of the $N_l = 6$ self-attention layers having the same structure of two sub-layers as in the original Transformer paper. The first sub-layer consists of a multi-head attention block with $N_h = 8$ heads of dimensionality $d_{head} = 8$ and a layer normalisation. The second sub-layer consists of two feed-forward layers with the dimensionalities $d_{ff,\,1} = 256$ and $d_{ff,\,2} = d_{model} = 64$.

Residual connections are employed after each of the sublayers, followed by the dropout layer with $p = 0.1$ and the layer normalisation. After the encoding layers, the learned embeddings are globally pooled by summing the embedding values across the token axis. Lastly, the decoder part proceeeds with $N_{ff} = 3$ feed-forward layers with the decreasing dimensionality $256 \rightarrow 128 \rightarrow 4$ followed by the Softmax layer.

The resulting model is referred to as Tau Transformer (TaT) and its architecture is illustrated on Fig. 4.13 (top) along with the ParticleNet [135] architecture (bottom). The ParticleNet is chosen for a benchmark on the $\tau_h$ identification task as being one of the most prominent models at the moment, showing an improvement on the jet tagging tasks and successfully used in several physical analyses [22, 136]. Its implementation follows the original paper with the exception of being adapted to a multimodality nature of the task. The modalities are embedded with the same Particle Embedding block as for TaT, with the only exception of removing the global token. The reason for that is an inherent limitation of the ParticleNet model: it assumes that the inputs to the first layer are placed in some coordinate system where the distance can be computed to define a k-nearest neighbor (kNN) graph. Since this is not applicable to the global token, it is concatenated with the learned embeddings in the encoder and further processed within the decoder. The encoder in ParticleNet consists of $N_l = 3$ EdgeConv layers as proposed in the original paper with a decreasing number of the nearest neighbours $k = 16 \rightarrow 12 \rightarrow 8$, the number of channels for each of the layers $C = (160, 128, 96)$ and feature aggregation via averaging. The kNN graph in the first EdgeConv layer is constructed in the $\eta - \phi$ plane around the $\tau_h$ direction of flight. After the encoder, the global pooling of embeddings is performed via summing across the constituents axis. It is followed by the decoder consisting $N_{ff} = 6$ of Dense layers with the decreasing dimensionality $192 \rightarrow 160 \rightarrow 128 \rightarrow 96 \rightarrow 64 \rightarrow 4$ and interleaved with the dropout layers ($p = 0.1$). Overall, TaT has 416k (embedding + encoder) + 50k (decoder) = 466k , while ParticleNet has 298k (embedding + encoder) + 98k (decoder) of trainable parameters. This is to be compared with the 1314k of trainable paramerers for DeepTau v2.5.

### Training setup

For training, 10% of the S&M data set (Sec. 4.3.2) is used, while another 10% of the S&M data set is used for validation. An extended set of features compared to DeepTau v2.5 (Sec. 4.3.2) is used to describe the input collections (see Appendix A.3). For the PF candidates, additional information about the number of hits and layers in the tracker system and HCAL energy deposits is included. For the RECO electrons, a set of variables is extended with those describing a shower shape in ECAL. For the global variables, information about a relative displacement of secondary vertex from a primary vertex is included if available. Furthermore, a collection of PF types is extended by adding the towers in the forward calorimeter reconstructed as either an hadron or an electromagnetic particle. Lastly, positional information about each of the particle is encoded

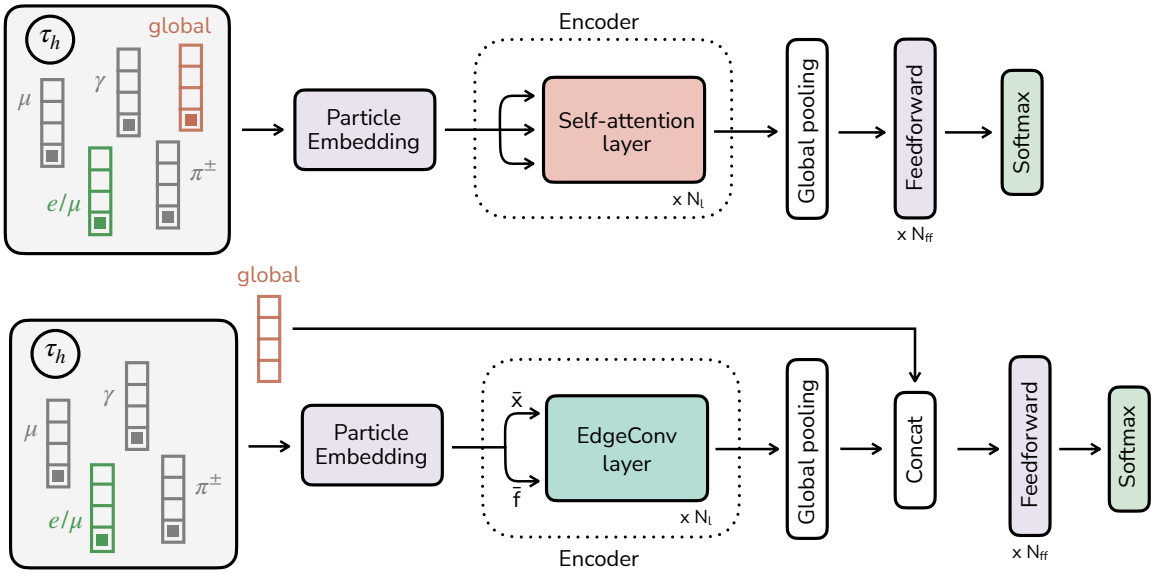


Figure 4.13: Tau Transformer (top) and ParticleNet (bottom) architectures as used for the $\tau_h$ identification task in this work. The hyperparameters and the structure of the layers are described in the text.

as two added features $(r, \theta)$, representing a radial and angular position in polar coordinates on $\eta - \phi$ plane centered around the HPS reconstructed $\tau_h$ direction of flight. No selection is applied on $r$ compared to the DeepTau case where the projection onto the grid naturally introduces a corresponding requirement. It is worth mentioning that this way of encoding relative positional information between constituents is implicit and might be suboptimal in performance. Studies of encoding it explicitly, for example, by injecting it into the attention matrices [137] – which also resembles the interaction terms from the Particle Transformer paper – could bring more insights into how Transformer-like models can profit from inductive biases of the particle physics domain.

The training is performed by minimising with an Adam optimiser (learning rate = $1e^{-4}$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 1e^{-7}$) a categorical cross-entropy loss function without any training weights and with the early stopping after 3 epochs. $\tau_h$ candidates are grouped into batches of 128 via a so-called uniform dynamic batching scheme. Since $\tau_h$ candidates have different number of PF and RECO constituents, a batch is padded with 0 (separately for PF and RECO modalities) for the tensors to have a regular shape. This effectively corresponds to adding for each $\tau_h$ candidate in the batch artificial "0" tokens until the maximum PF/RECO sequence length in the batch. Traditionally, $\tau_h$ candidates for batching are sampled randomly from the training

data set, which translates into a random sampling of particle sequences from the underlying distribution in the training data set (Fig. 4.14, left). Since the distribution is skewed towards $\tau_h$ candidates with a larger number of constituents, it results in a high probability for a given batch to have such a $\tau_h$ candidate, while the other $\tau_h$ candidates will have on average significantly lower number of constituents. After padding, it results in a large number of artificial tokens in a batch and on average larger batches if measured over the constituent dimension (Fig. 4.14, right, blue distribution). It in turn translates into inefficient computation of attention weights as it scales quadratically with a sequence length.

To mitigate this inefficiency and produce more compact batches, the training data set is divided into 30 equal bins from 0 to 300 over the number of PF constituents per $\tau_h$ candidate. Then, a single batch is allowed to be formed only from $\tau_h$ candidates sampled from a single bin. The batches are further shuffled in order to avoid bias in the training procedure. This procedure significantly reduces the number of padded tokens after batching and brings the distribution of the number of constituents per batch close to the original distribution in the training data set before padding (Fig. 4.14, right, orange distribution). Overall, a speed-up by a factor 2-3 in the training duration compared to the traditional batching is achieved without any difference in the performance.

**Experiments**

First, an impact of various modalities on the model performance is studied. Starting from the basic representation of a $\tau_h$ candidate as only a set of PF candidates, RECO electrons/muons and then global variables are added to the input representation with the model being trained for each of the three scenarios. To separate the impact of the multiclass setting, the training is performed separately for each of the three binary classification problems $\tau_h$ vs. $e/\mu$/jet. The model, with the parameters as described above, remains fixed in all of the experiments, as well as the training data set. In this particular study, the same set of input features for each of the modalities as for the DeepTau v2.5 training is used. An experiment corresponding to the model trained in a multiclass setup is additionally performed. The performance is evaluated with a ROC curve as described in Sec. 4.3.2, using the 2018 samples as in the DeepTau v2.5 case with genuine $\tau_h$ being sourced from the ggH sample, electrons and muons from the DY sample, and jets from the $t\bar{t}$ (semileptonic) sample.

Overall, significant gain in performance is observed from addition of both RECO electrons/muons and global features (Fig. 4.15, Appendix A.4). While addition of RECO electron/muons does not affect the performance against the jet scenario, it does improve it for the $\tau_h$ against the electron (in a high $\tau_h$ efficiency region, up to 10 times reduced fake rate) and against the muon (throughout the $\tau_h$ efficiency region of interest, up to 100 times reduced fake rate) scenarios. Addition of global variables closes the gap between TaT and DeepTau v2.5

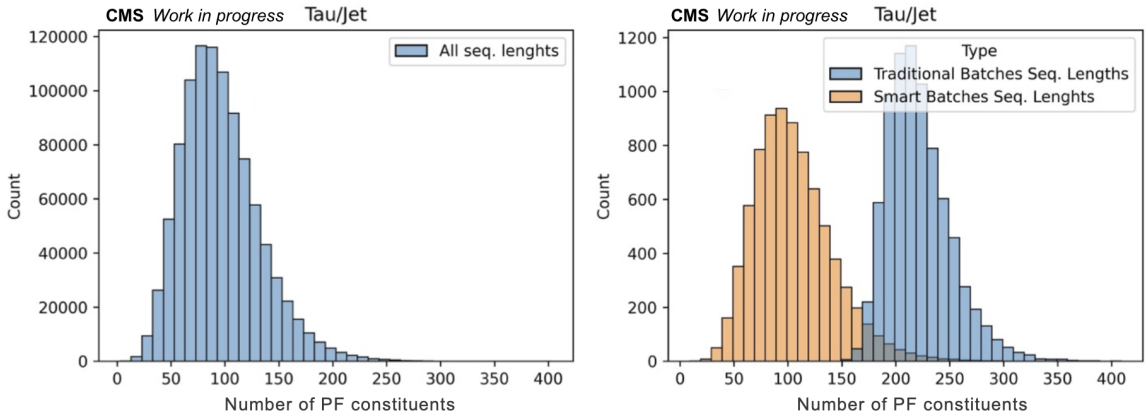


Figure 4.14: The distribution of the number of PF constituents per $\tau_h$ candidate in the training data set consisting of genuine $\tau_h$ and jet objects (left), after traditional batching (right, blue) and after uniform dynamic batching, also referred to as smart batching (right, orange). For the latter two, the distribution corresponds to the number of constituents per batch after padding.

performance for the scenario against electron and pushes the performance further by up to 30% for the scenarios against jet and muon. Lastly, switching from a binary classification to a multiclass model also improves the performance. This indicates that the model with a given set of hyperparameters profits from extending the effective size of the training data set and learning a joint representation to simultaneously discriminate between the four classes.

Second, a ParticleNet model with the parameters as described above (referred to as ParticleNet v0.1) is benchmarked against the TaT model corresponding to the multiclass scenario with all the modalities (referred to as TaT v0.2). It should be mentioned that both TaT and ParticleNet models do not have an adversarial fine-tuning step, which makes the comparison with DeepTau v2.5 more optimistic. However, the performance degradation due to this is not expeted to be large, as it was shown previosly during the DeepTau v2.5 (stage 2) training. An additional requirement on the z component of an impact parameter vector of $\tau_h$ w.r.t. a primary vertex $|d_z| < 0.2$ cm is applied during evaluation to be aligned with the recommended $\tau_h$ candidate selection (Sec. 5.3.3).

The corresponding ROC curves, evaluated with the same conditions as in the modality study,

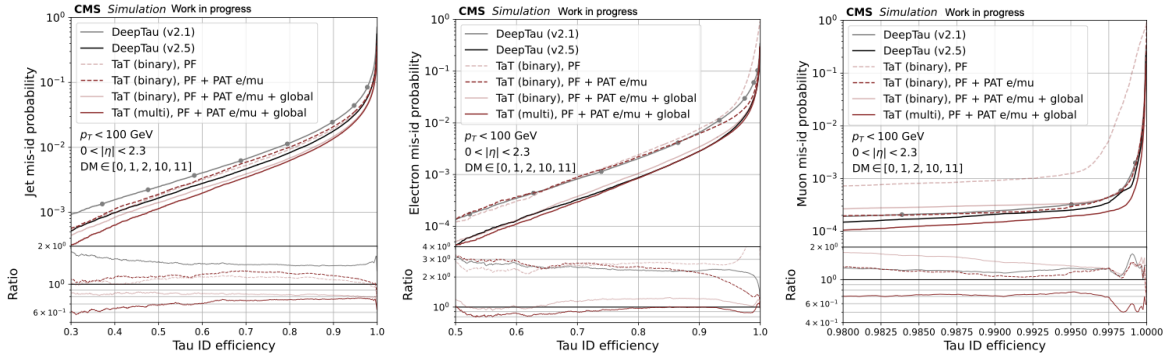


Figure 4.15: Efficiency for jets (left), electrons (middle), and muons (right) versus efficiency for genuine $\tau_h$ to pass the corresponding $D_\alpha$ discriminators for DeepTau v2.1, DeepTau v2.5 and TaT models. For the latter, several configurations are trained with various modalities used as an input: PF candidates only on a binary classification task (dashed red, pale), PF candidates with RECO electrons/muons (PAT e/mu in the legend) on a binary classification task (dashed red, dark), PF candidates with RECO electrons/muons and global features on a binary classification task (solid red, pale), PF candidates with RECO electrons/muons and global features on a multiclass classification task (solid red, dark). Working points (grey dots) for DeepTau v2.1 are also shown, as derived in the original paper. The panel at the bottom of each figure shows the ratio of each of the ROC curves with respect to the one of the DeepTau v2.5 model.

are shown on Fig. 4.16 and in Appendix A.5. Overall, the TaT architecture improves upon the current baseline of DeepTau v2.5. In the most populated region $p_T(\tau_h) \in [20, 100)$, $|\eta(\tau_h)| < 2.3$, $DM(\tau_h) \in \{0, 1, 10, 11\}$, TaT consistently reduces the misidentification rate against jet by up to 30% across the $\tau_h$ efficiency range. For the electrons and muons, the TaT performance is slightly better compared to DeepTau v2.5 by up to 10% in the misidentification rate at the fixed $\tau_h$ efficiency. However, the performance gain of TaT in the scenarios against electron/muon is more pronounced for the other regions of the phase space, in particular for $\tau_h$ candidates reconstructed in HPS decay modes 10 and 11, where it reaches up to 50% and 70% reduced misidentification rate against electrons and muons, respectively (Appendix A.5). Against jets, the ParticleNet model reaches the similar performance as TaT (DeepTau v2.5) in the low (high) $\tau_h$ efficiency region. While overall there is little difference in the performance against muons for all of the benchmarked models in the low-$p_T$ barrel region, for the scenario against electron ParticleNet shows a performance lower than both DeepTau v2.1 and DeepTau v2.5 by a significant margin.

In general, both models can profit from further hyperparameter tuning and an increase of the training data set size. On the Transformer side, as it is pointed out in [138] and also hinted in the Particle Transformer paper, attention-based models scale extremely well with the increase of training data set size. Therefore, future studies on the extended data set are needed to gauge the scalability of such models on the $\tau_h$ identification task.

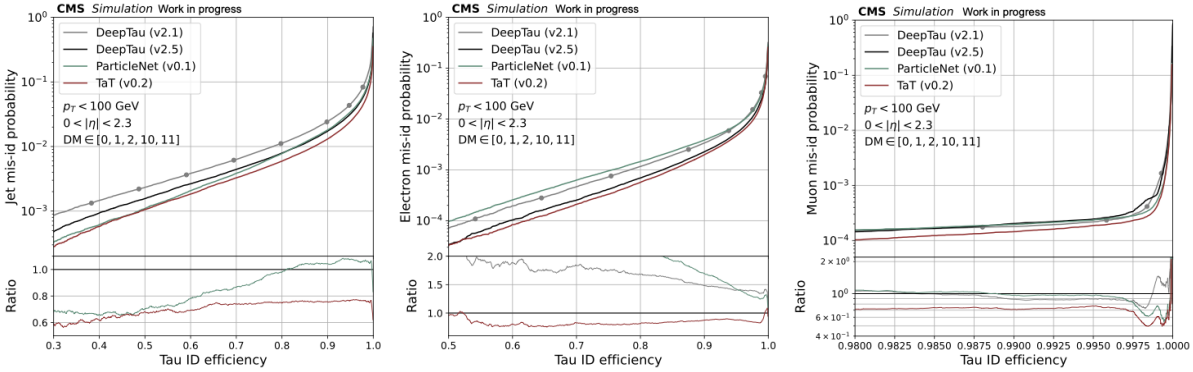


Figure 4.16: Efficiency for jets (left), electrons (middle), and muons (right) versus efficiency for genuine $\tau_h$ to pass the corresponding $D_\alpha$ discriminators for DeepTau v2.1 (grey), DeepTau v2.5 (black), ParticleNet v0.1 (dark cyan) and TaT v0.2 (red) models. Working points (grey dots) for DeepTau v2.1 are also shown, as derived in the original paper. The panel at the bottom of each figure shows the ratio of each of the ROC curves with respect to the one of the DeepTau v2.5 model.

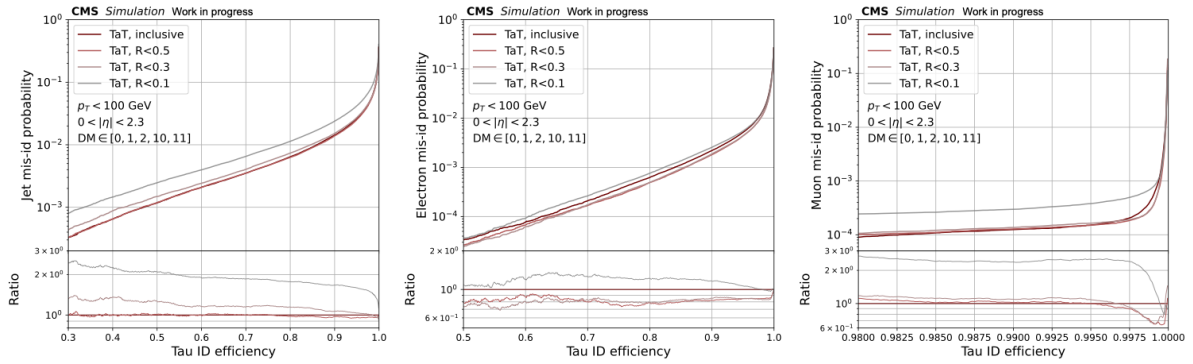


Figure 4.17: Efficiency for jets (left), electrons (middle), and muons (right) versus efficiency for genuine $\tau_h$ to pass the corresponding $D_\alpha$ discriminators for a TaT architecture with the various requirements on the cone distance between the directions of flight of constituents and $\tau_h$ candidate (R). The panel at the bottom of each figure shows the ratio of each of the ROC curves with respect to the model without any requirement on the cone distance (inclusive).

Last, an impact of the cone requirement on the constituents is studied for the TaT v0.2 architecture. In the original data set, used for the studies above, no selection is applied on $r$, corresponding to the radial (cone) distance in $\eta-\phi$ plane between the constituent (PF candidate or RECO electron/muon) and the HPS-reconstructed $\tau_h$ directions of flight. Additional trainings of the same architecture with the requirements on the constituents $r < 0.5$, $r < 0.3$ and $r < 0.1$ are performed. Since training instabilities were observed during the studies with the TaT v0.2 setup, the optimiser for the training was tuned. RAdam [139] optimiser with $\beta_1 = 0.9$, $\beta_2 = 0.999$,

$\epsilon = 1e^{-7}$ is used. Initial learning rate is set to $1e^{-4}$ and is reduced during the training by 10 after every 10 epochs. Also, the GeLU activation function [140] is introduced in all TaT layers instead of the previously used ReLU function.

The comparison of evaluated ROC curves for each of the trials is presented on Fig. 4.17 and in Appendix A.6. As expected, the performance degrades for the tightest $r < 0.1$ case, which approximately corresponds to using only the candidates from the inner grid of DeepTau (Sec. 4.3.1). The performance as measured by a misidentification rate against electrons and muons at a given $\tau_h$ efficiency improves by about 10-20% similarly for $r < 0.3$ and $r < 0.5$ cases. However, the $r < 0.3$ scenario is not favoured in the against jet task as being too narrow to capture the hadronisation patterns of QCD jets. Overall, the study indicates that the cone distance $r < 0.5$, also corresponding to the outer grid size of DeepTau, is the optimal option.

In general, the studies described in this work illustrate that the field of jet tagging and representation learning in HEP can largely profit from adaptation of attention-based architectures. On the side of performance, scalability, flexibility and multimodality treatment they offer a powerful alternative to graph-based architectures for analysts to model various physics objects in the detector. This motivates future studies to understand the scope of these models' performance and the broadness of their applicability.

# 5 | CP analysis in H $\rightarrow \tau_e \tau_h$ decays

## 5.1 Introduction

In order to probe the CP nature of the interaction between the Higgs boson and tau leptons one needs to modify the SM Lagrangian in a way to incorporate effects deviating from the Standard Model expectations. This can be done already at the tree level and without assuming any model dependence. For the tau lepton case, one can write the following extension of the Yukawa coupling term in the SM Lagrangian [141]:

$$\mathscr{L}_Y = -\frac{m_\tau}{v}(\kappa_\tau \bar{\tau}\tau + \tilde{\kappa}_\tau \bar{\tau} i\gamma_5 \tau)\mathrm{H}. \tag{5.1}$$

Here $m_\tau = 1.776(86)$ GeV is the mass of the tau lepton, $v = \left(\sqrt{2G_\mathrm{F}}\right)^{-1/2} \approx 246$ GeV is the vacuum expectation value of the Higgs field (Sec. 2.2.3), $G_\mathrm{F} = 1.1663787(6) \times 10^{-5}$ GeV$^{-2}$ is the Fermi constant, $\kappa_\tau$ and $\tilde{\kappa}_\tau$ are the coupling strength modifiers, H is the SM Higgs boson. While this term is written specifically for the SM scenario, it is generally applicable to any neutral spin-zero particle H of arbitrary CP nature with a flavor-diagonal Yukawa interaction with a fermion.

Few things should be noted regarding Eq. 5.1. First, while such parametrisation is model-independent *per se*, the couplings $\kappa_\tau$ and $\tilde{\kappa}_\tau$ can depend on the specific model under consideration and can be interpreted within the framework of these theories. For example, in the context of a two-Higgs doublet model, e.g. minimal supersymmetric SM extension (MSSM) [142], depending on the CP violation (CPV) scenario, the three mass eigenstates $h_i$ can be either two scalars (denoted as $h, H$) and one pseudoscalar (denoted as $A$) or states representing CP mixtures with both $\kappa$ and $\tilde{\kappa}$ couplings having non-zero values. In this work, no model-dependent interpretation of the couplings is made.

Second, an assumption of real-valued $\kappa$ and $\tilde{\kappa}$ is made. While this makes the Yukawa coupling term Hermitian similarly to the rest of the SM Lagrangian, it is arguably an imposed property

which might not hold true in nature [143]. Therefore, a generalisation of the study presented in this work towards a non-Hermitian Yukawa coupling [144, 145] is an interesting direction for future studies.

From the couplings formulation one can rewrite Eq. 5.1 as:

$$\mathscr{L}_Y = -g_\tau(\cos\alpha^{H\tau\tau}\bar{\tau}\tau + \sin\alpha^{H\tau\tau}\bar{\tau}i\gamma_5\tau)\mathrm{H}, \tag{5.2}$$

where $g_\tau$ is a generalised coupling modifier and an effective mixing angle is introduced:

$$\tan\left(\alpha^{H\tau\tau}\right) = \frac{\tilde{\kappa_\tau}}{\kappa_\tau}. \tag{5.3}$$

For the SM Higgs boson with the quantum numbers $J^{PC} = 0^{++}$ (pure scalar), $\tilde{\kappa_\tau} = 0$ and $\kappa_\tau = 1$, therefore $\alpha^{\mathrm{H}\tau\tau} = 0°$. The scenario of $J^{PC} = 0^{++}$ (pure pseudoscalar) corresponds to $\alpha^{\mathrm{H}\tau\tau} = 90°$. Any intermediate value corresponds to the mixture of the couplings between CP-even and CP-odd components.

Having the interaction defined in terms of the Lagrangian term, one can proceed to the derivation of the partial decay width of the SM Higgs boson into a pair of tau leptons. Using Eq. 5.2 with the approximation $\beta_\tau = \sqrt{1 - 4m_\tau^2/m_h^2} \approx 1$ one can obtain [146]:

$$d\Gamma_{\mathrm{H}\to\tau\tau} \sim 1 - s_z^+ s_z^- + \cos\left(2\alpha^{\mathrm{H}\tau\tau}\right)(\boldsymbol{s}_T^+ \cdot \boldsymbol{s}_T^-) + \sin\left(2\alpha^{\mathrm{H}\tau\tau}\right)\left[(\boldsymbol{s}_T^+ \times \boldsymbol{s}_T^-) \cdot \hat{\boldsymbol{k}}^-\right], \tag{5.4}$$

where $\hat{\boldsymbol{k}}^-$ is a normalised $\tau^-$ momentum in the Higgs rest frame which points towards a positive direction of the $z$ axis, and $\boldsymbol{s}_{\mathrm{T}}^\pm$ ($s_z^\pm$) is a projection of normalised spin vector of the tau lepton in its rest frame on the $xy$ plane ($z$ axis) in a right handed coordinate system. It can be seen that it is the spin correlation between transverse components of the tau leptons' spin vectors which is sensitive to the CP structure, parametrised by $\alpha^{\mathrm{H}\tau\tau}$. Introducing $\phi_s$ as an angle pointing from $\boldsymbol{s}_T^+$ to $\boldsymbol{s}_T^-$, one obtains:

$$d\Gamma_{\mathrm{H}\to\tau\tau} \sim 1 - s_z^+ s_z^- + |\boldsymbol{s}_T^+||\boldsymbol{s}_T^-|\cos\left(\phi_s - 2\alpha^{\mathrm{H}\tau\tau}\right). \tag{5.5}$$

Conceptually, Eq. 5.5 lays out the strategy to experimentally probe the CP structure of the Yukawa coupling between the Higgs boson and tau leptons. First, one needs to reconstruct the $\phi_s$ angle between the spin vectors $\boldsymbol{s}_T^\pm$ of the tau leptons. This can be achieved by studying the angular distributions of $\tau$ decay products, as described in Sec. 5.3.7. Second, in a simplified picture the differential distribution of the $\phi_s$ angle will allow to extract the phase shift $\alpha^{\mathrm{H}\tau\tau}$ from the fit with $a \cdot \cos\left(\phi - 2\alpha^{\mathrm{H}\tau\tau}\right) + b$ function. This in turn directly points to the CP nature (CP-even, CP-odd, or their mixture) of the SM Higgs boson via Eq. 5.3.

In the following sections of this chapter a step-by-step overview towards this goal is described. Starting from the description of the data sets used in the analysis (Sec. 5.2), an overview of physics objects and observable reconstruction is given in Sec. 5.3. In Sec. 5.4 a procedure to select H $\rightarrow \tau\tau$ candidates is described, followed by techniques to model background processes (Sec. 5.5). After the selection of the H candidates is performed, ML methods are used to categorise a given candidate as either originating from a signal or background processes (Sec. 5.6). Taking into account necessary systematic uncertainties (Sec. 5.7.3), a statistical inference procedure is performed (Sec. 5.7) to extract the effective mixing angle $\alpha^{H\tau\tau}$. Finally, results of the measurement and the corresponding conclusion are given in Sec. 5.8.

## 5.2 Data & Simulation

For this work, a data set of *pp* collisions collected by the CMS detector at $\sqrt{s}$ = 13 TeV in 2016, 2017, and 2018 years is used. The corresponding integrated luminosities are 35.9, 41.5, and 59.7 fb$^{-1}$.

Several Monte Carlo simulated data sets are produced in order to model signal and background processes. The signal processes consist of a Higgs boson being produced through the gluon-gluon fusion (ggH), vector boson fusion (VBF), or associated production with a W or Z boson (WH, ZH, or VH for their combination). These samples are generated at next-to-leading order (NLO) in QCD with the POWHEG 2.0 event generator [147–153]. The procedure is configured to produce a scalar Higgs boson. However, addition of CP mixing effects in the production mechanism, for example, by modifying the Higgs coupling to top and bottom quarks, can affect the distribution of physical observables (e.g. related to the accompanying jets), as well as the signal acceptance. It is studied that this contribution is negligible comparing to the theoretical uncertainties and therefore does not affect the CP measurement in the H $\rightarrow \tau\tau$ decay.

Reweighting is applied to distributions of the Higgs boson transverse momentum and the jet multiplicity to match with those of the samples produced at next-to-NLO with the POWHEG NNLOPS (version 1) generator [154,155]. The decay of the Higgs boson into a pair of tau leptons is described by the PYTHIA generator version 8.230 [156] without accounting for the $\tau$ spin correlations. These are included within the TAUSPINNER package [157], which reweights the signal samples according to predefined values of the mixing angle $\alpha^{H\tau\tau} = \{0°, 45°, 90°\}$ chosen to define the signal templates for the statistical inference (Sec. 5.7.2). For all 2016 samples NLO parton distribution functions (PDFs) are generated with the NNPDF3.0 [158]. For all 2017 and 2018 samples NNLO PDFs distributions are generated with the NNPDF3.1 [159].

Processes with a Z or W boson accompanied by up to four outgoing partons are generated with MADGRAPH5 aMC@NLO (version 2.6.0) [160]. W bosons originating from the top quark decay are generated at leading order with the MLM jet matching and merging approach [161],

as well as the diboson production at NLO. POWHEG 2.0 (1.0) is used for single top (ST) quark production (associated with a W boson) [162,163] and top quark-antiquark pair production [164]. For modelling of the parton showering, fragmentation, and the decay of the $\tau$ lepton the generators are interfaced with PYTHIA with its parameters set to the CUETP8M1 tune [165] , and CP5 tune [166] in 2017 and 2018.

The simulation of the CMS detector is based on GEANT 4 [167]. Additional *pp* interactions per bunch crossing (also referred to as pileup interactions) are generated with PYTHIA and reweighted to match the pileup distribution in data.

## 5.3 Event reconstruction

The particle-flow (PF) algorithm (Sec. 4.2.1) is at the core of the physics object reconstruction in CMS. It builds upon the idea of combining information from all the subsystems of the detector in order to improve the overall reconstruction efficiency. Using a hierarchical approach which starts from the construction of fundamental building blocks (e.g. tracks or clusters) it further combines them into high-level physics objects such as muons or charged hadrons. Furthermore, it serves as a basis for other algorithms building more complex objects, such as the jet clustering (Sec. 5.3.4) or the hadron-plus-strips algorithm (Sec. 4.2.2).

### 5.3.1 Electrons

Electron object reconstruction [82] also builds on top of the PF basic elements: GSF tracks and ECAL clusters (Sec. 4.2.1). Conceptually, these elements are further combined, refined and filtered to yield a final electron object in the following procedure:

1. ECAL clusters are combined into a supercluster (SC) with a so-called mustache algorithm (Fig. 5.1). The idea is to aggregate clusters coming from extensive bremsstrahlung and photon conversion within a "mustache" window in $\eta$-$\phi$ plane which accounts for the magnetic field of the CMS detector.

2. Association of SCs with GSF tracks is performed based on the output of a boosted decision tree (BDT) using as input SCs observables, track parameters and the SC-GSF matching variables.

3. Refinement of the mustache SCs is done, which leverages the information from subdetectors outside of ECAL. This step recovers additional bremsstrahlung and conversion clusters. Moreover, a conversion-finding algorithm [168] with a dedicated BDT are used to identify pairs of tracks compatible with a converted photon.

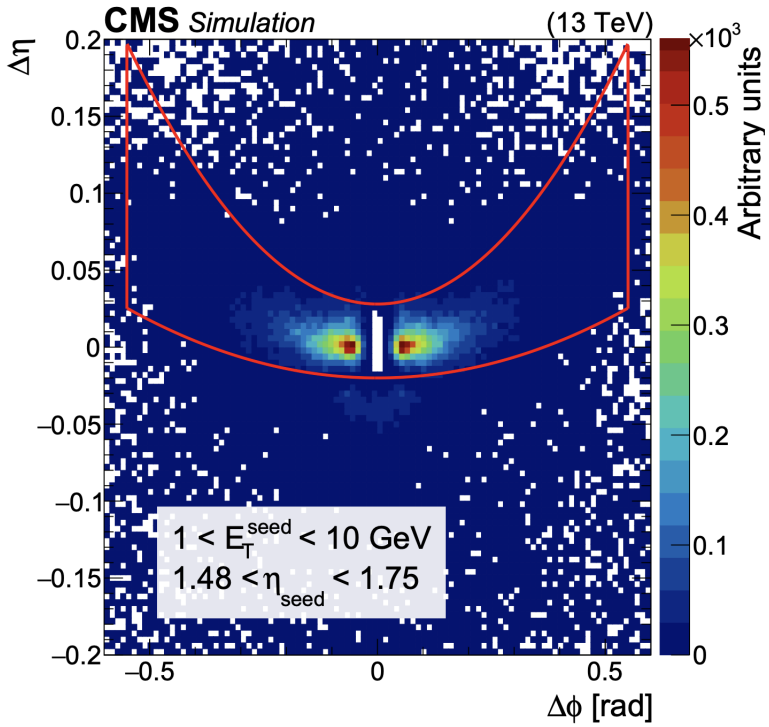


Figure 5.1: Distribution of PF clusters around the seed cluster for simulated electrons with $1 < E_T^{\text{seed}} < 10$ GeV and $1.48 < \eta^{\text{seed}} < 1.75$ [82]. The $z$ axis shows the number of PF clusters around the seed matched to simulation. The red line illustrates the region where the clusters are selected by the mustache algorithm.

4. All the input elements (ECAL clusters, mustache SCs, electron associated generic tracks, GSF tracks, conversion-identified tracks) are submitted to the PF algorithm to form electron candidates. After the linking, the final set of ECAL clusters for each candidate is promoted to a refined supercluster.

5. Final electron objects are formed from a refined SC with an associated GSF track based on the loose requirements on the BDT output. The BDT is trained using the shower-shape, isolation and track-related variables as input.

Overall, the procedure results in a good efficiency of electron reconstruction across $p_T$ and $\eta$ ranges (Fig. 5.2). However, it should be noted that a graph neural network (GNN) based algorithm to form supercluster has been recently proposed to recover for inefficiency of mustache energy aggregation and also to provide better robustness to pileup [169].

Since the energy of electrons is not fully reconstructed due to losses in the tracker or shower leakage in ECAL, corresponding corrections should be applied. This is achieved by firstly

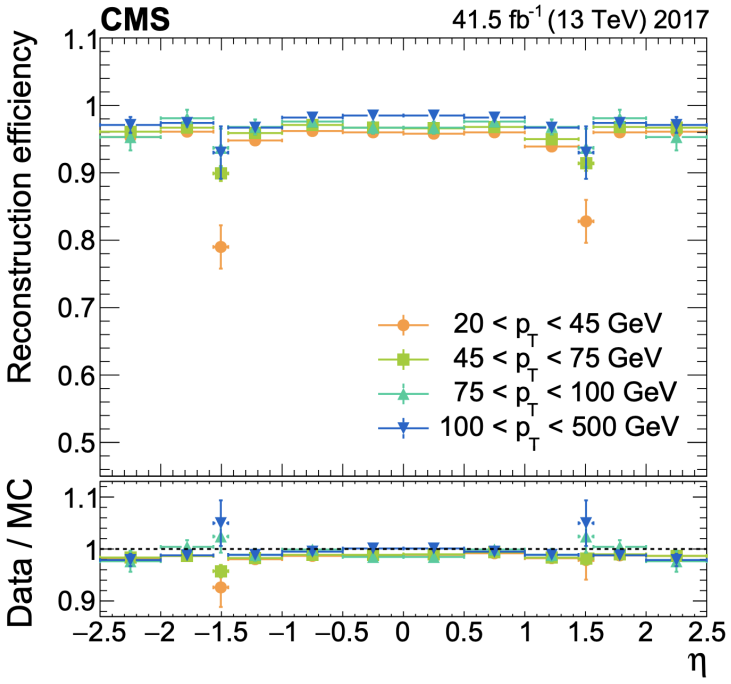


Figure 5.2: Electron reconstruction efficiency versus $\eta$ for various $p_T$ ranges (upper panel) and ratios of data and simulation efficiencies (lower panel) in 2017 data taking period [82].

performing correction of SC energy and resolution via a 3-step BDT regression. Second, residual discrepancies between data and simulation are taken into account with energy scale and smearing corrections derived from $Z \to ee$ events. These results in a significant improvement both on the side of energy resolution (Fig. 5.3, left) and physical observables (Fig. 5.3, right).

After the reconstruction of electron objects an identification step follows. Since the reconstruction algorithms are designed to be general-purpose and as inclusive as possible, it results in a sizeable fraction of background objects in the reconstructed electron collection. The identification step aims at the separation of prompt (created in the primary $pp$ interaction) genuine electrons from misidentified objects or non-prompt electrons (usually from heavy flavour jets). For that purpose, two methods are used. The first one is a cut-based discriminator based on the isolation variable:

$$I^e_{\text{rel}} = \frac{\sum p_T(h^{\pm}) + \max\left(\sum p_T(h^0) + p_T(\gamma) - \rho \cdot A_{\text{eff}}, 0\right)}{p^e_T}, \tag{5.6}$$

where $A_{\text{eff}}$ is the $\eta$-dependent isolation area [170], $E_T \equiv \sqrt{m^2 + p_T^2}$, $\rho$ is the average neutral component of the pileup energy density per unit area in the $\eta$-$\phi$ plane, and the sums are

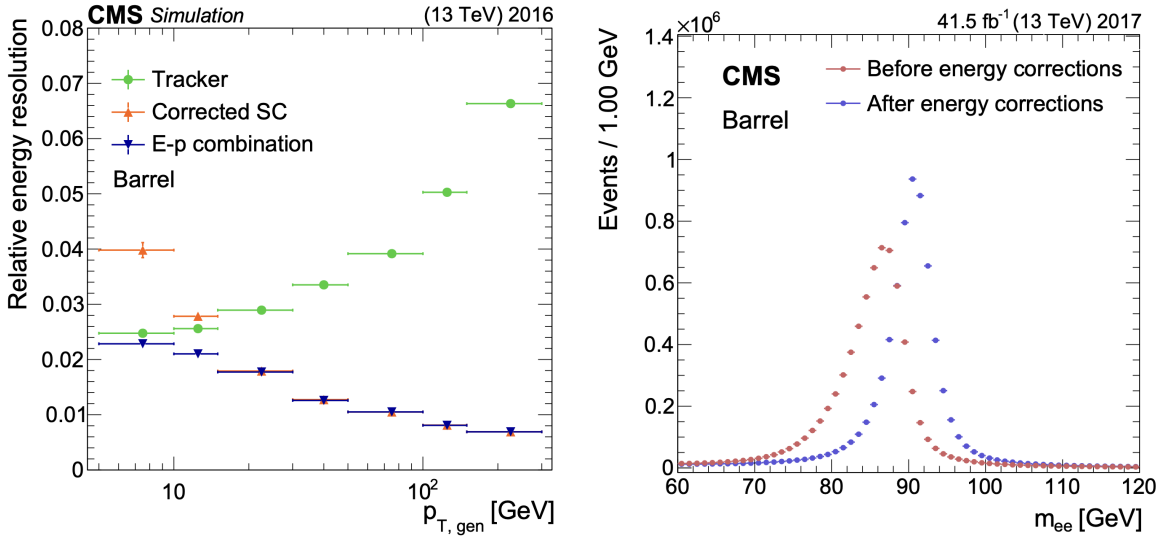


Figure 5.3: Left: relative energy resolution as a function of electron $p_T$, measured by the tracker, by ECAL ("corrected SC"), and after the third step of the energy regression ("E-p combination"). Right: invariant mass of an electron pair in the barrel region from $Z \to ee$ events in 2017 data before and after applying regression and scale corrections. [82].

computed across PF candidates of a given type within a cone $\Delta R \equiv \sqrt{\Delta\eta^2 + \Delta\phi^2} < 0.3$ around the reconstructed electron. The sum over the charged PF candidates runs over the candidates associated with PV, while the sum over the neutral PF candidates does not have this requirement. Thresholds on this discriminator are derived to target specific predefined selection efficiencies. In this work, the requirement $I^e_{\text{rel}} < 0.15$ is applied to the selected electron objects (Sec. 5.4).

The second method of electron identification is based on a boosted decision tree. It uses information about the track-cluster matching and energy deposits in HCAL/ECAL, as well as cluster-shape, track-quality variables and provides a score for a reconstructed electron object to be a genuine prompt electron. Working points are defined as thresholds on the score to target predefined electron selection efficiency. In this work, a working point corresponding to 90% efficiency is used.

Lastly, additional requirements are applied on the transverse ($|d_{xy}| < 0.045$ cm) and longitudinal ($|d_z| < 0.2$ cm) impact parameters of the selected electrons.

### 5.3.2 Muons

Muon reconstruction relies on the standalone algorithm as described in Sec. 4.2.1. An identification step for muons is based on a set of requirements aimed to provide a predefined selection

efficiency. In this work the muon object is required to be reconstructed as a tracker or global muon and pass the hit and segment compatibility quality selection. Same requirements on the impact parameters as in case of electrons are applied: $|d_{xy}| < 0.045$ cm and $|d_z| < 0.2$ cm. The isolation variable is also defined as:

$$I^{\mu}_{\text{rel}} = \frac{\sum p_{\text{T}}(\text{h}^{\pm}) + \max\left(\sum p_{\text{T}}(\text{h}^{0}) + p_{\text{T}}(\gamma) - \frac{1}{2}\sum p_{\text{T}}(\text{h}^{\pm}_{\text{PU}}), 0\right)}{p^{\mu}_{\text{T}}}, \tag{5.7}$$

where the sums are taken for the PF candidates in the isolation cone $\Delta R < 0.4$ centered around the reconstructed muon direction of flight. The sum $\sum p_{\text{T}}(\text{h}^{\pm}_{\text{PU}})$ is computed over the charged PF candidates originating from pileup vertices and scaled down by a factor 1/2 to approximate and subtract the pileup contribution from neutral particles. A requirement $I^{\mu}_{\text{rel}} < 0.15$ is also applied.

### 5.3.3 Tau leptons

Tau leptons decaying hadronically ($\tau_{\text{h}}$) are reconstructed with a dedicated hadron-plus-strips (HPS) algorithm as described in Sec. 4.2.2. First, it aims to reconstructs $\pi^0$ coming from the $\tau_{\text{h}}$ decays in a form of "strips". Second, it combines them with charged hadrons to form potential $\tau_{\text{h}}$ candidates according to the expected decay modes (DM) (Sec. 4.1).

For the identification step, a DeepTau model (Sec. 4.3.1) is used to separate $\tau_{\text{h}}$ candidates, reconstructed by the HPS algorithm, from jets, electrons, and muons. The model is built from 1D and 2D convolutional layers operating on a grid in the $\eta$-$\phi$ plane centered around the HPS-reconstructed $\tau_{\text{h}}$ candidate. It combines low-level information from PF candidates and RECO electrons/muons to separate between genuine $\tau_{\text{h}}$ and fakes. For this work, the $\tau_{\text{h}}$ candidate is required to pass the working points which correspond to the probability of 70%, 80%, and 99.95% (Medium, Tight, Very loose, respectively) for the genuine $\tau_{\text{h}}$ to pass DeepTau discriminators against jets, electrons, and muons, respectively. Furthermore, the $z$ component of the impact parameter of the leading charged track with respect to PV is required to be $|d_z| < 0.2$ cm.

Since there are undetectable neutrino(s) present in the $\tau$ decays, the full reconstruction of the $\tau\tau$ system is not possible by means of $\tau_{\text{h}}$ reconstruction algorithms only. A dedicated SVFit algorithm [171] is used to recover for this loss of information. It combines the missing transverse momentum vector $\vec{p}^{\,\text{miss}}_{\text{T}}$ and its uncertainty matrix with the reconstructed four-vectors of two tau leptons and uses a simplified matrix-element approach to reconstruct the invariant mass ($m_{\tau\tau}$) of the ditau system. This variable provides better separation between H $\rightarrow \tau\tau$ and $Z/\gamma^* \rightarrow \tau\tau$ events, compared to the visible mass ($m_{\text{vis}}$) of the ditau system (Fig. 5.4).

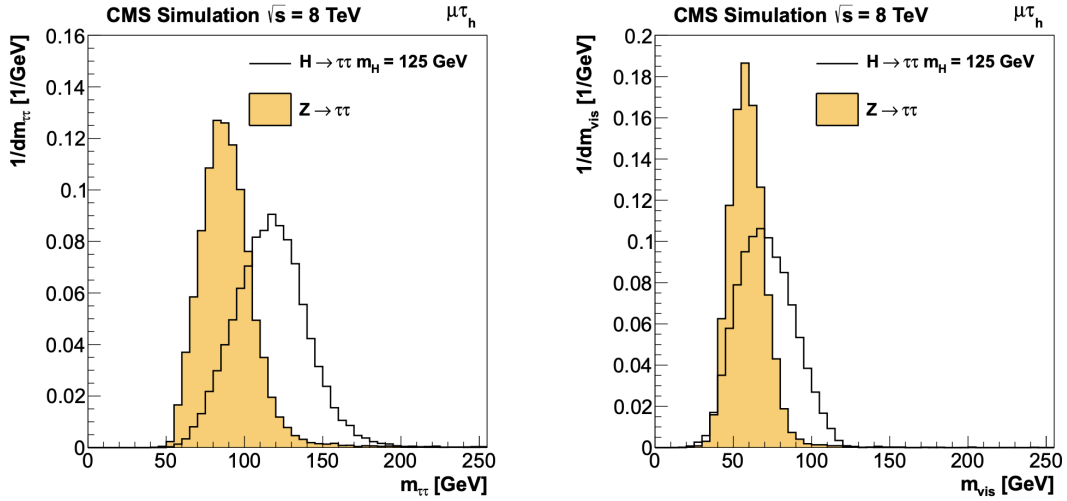


Figure 5.4: Distribution of $m_{\tau\tau}$ variable derived with the SVFit algorithm (left) and visible mass $m_{\text{vis}}$ of the ditau system (right) for the H → $\tau\tau$ (black line histogram) and $Z \rightarrow \tau\tau$ (yellow filled histogram) events in the $\tau_\mu\tau_h$ final state [171].

The analysis in this work, as it will be shown in Sec. 5.3.7, heavily relies on identification of $\tau_h$ decay modes. Several changes have been already introduced with the DeepTau algorithm to improve their purity and efficiency at the stage of the HPS algorithm. However, the migration from, for example, DM = 11 ($h^\pm h^\mp h^\pm h^0$) to DM = 10 ($h^\pm h^\mp h^\pm$) is still sizeable ($\sim$ 25%) and leads to the contamination of the latter DM category, which in turn affects the analysis sensitivity. Furthermore, DM = 2 ($h^\pm h^0 h^0$) is merged with DM = 1 ($h^\pm h^0$) which does not allow for their separate analysis.

To mitigate these limitations, two BDTs (referred to as MVA DM) are trained and applied on top of the HPS reconstructed $\tau_h$ candidates to predict their decay mode [172]. One BDT is designed to identify decay modes with one charged prong and the number of $\pi^0$ $n(\pi^0) = \{0, 1, 2\}$ (DM = 0, 1, 2), while the other targets $\tau_h$ candidates with three charged prongs and $n(\pi^0) = \{0, 1\}$ (DM = 10, 11). The input variables to the BDT describe the kinematics, invariant mass properties and angular information of the constituents of an HPS reconstructed $\tau_h$ candidate. H → $\tau\tau$ events in the $\tau_\mu\tau_h$ and $\tau_h\tau_h$ final states with the H produced via vector-boson or gluon-gluon fusion are used for the training. Overall, the BDTs provide the identification of $\tau_h$ candidates with DM = 2 and consistently improve the purity of DM selection by up to 25% without significant reduction in efficiency (Fig. 5.5).

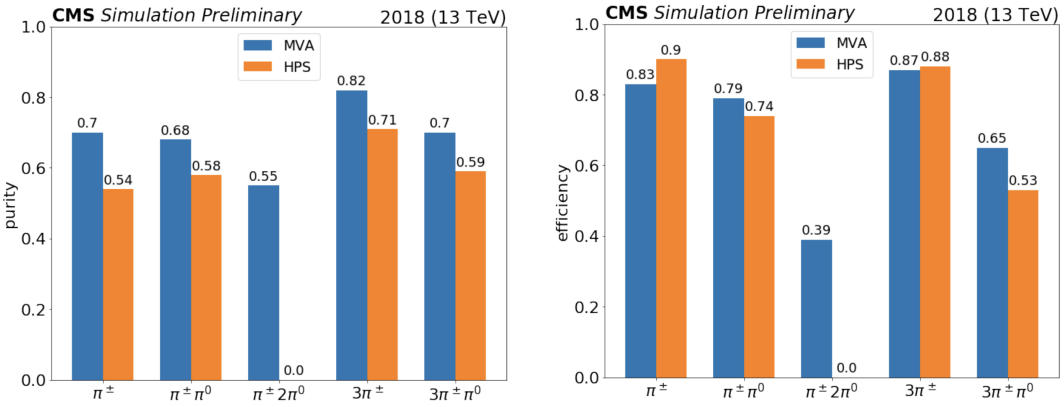


Figure 5.5: Comparison of purity and efficiency of the $\tau_h$ decay mode identification between the HPS (orange bars) and MVA DM algorithms (blue bars) [172].

### 5.3.4 Jets and missing transverse energy

An anti-$k_T$ algorithm [86] with the distance parameter $R = 0.4$ as implemented in the FastJet package [173] is used for the reconstruction of jets. Effectively, it is proposed as an extension of the $k_T$ [174] and Cambridge/Aachen [175] algorithms with redefining the distance measure as:

$$d_{ij} = \min(k_{T,i}^{2p}, k_{T,j}^{2p})\frac{\Delta_{ij}^2}{R^2} \tag{5.8}$$

$$d_{iB} = k_{T,i}^{2p}, \tag{5.9}$$

where $d_{ij}$ is the distance between entities $i$ and $j$ (either particles or "pseudojets"), $d_{iB}$ is the distance between the entity $i$ and the beam, $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ with $k_{T,i}$, $y_i$, $\phi_i$ being the transverse momentum, rapidity and azimuthal angle of the entity $i$, respectively. The parameter $p$ balances between the energy and geometrical scales. For $p = 0$ one obtains an inclusive Cambridge/Aachen algorithm, while the case $p = 1$ corresponds to the $k_T$ algorithm.

An inclusive anti-$k_T$ algorithm corresponding to the case $p = -1$ starts by combinatorically computing the distances $d_{ij}$ and $d_{iB}$ between input PF particles. If $d_{ij}$ is the smallest out of two, the particles/entities are merged together into a single entity (so-called "pseudojet"). Otherwise, the particle/entity $i$ is removed from the list and called a jet. Then the distances are recalculated until there are no entities left.

In order to correct for the impact of pileup interactions on the jet observables, a charge hadron subtraction (CHS) technique is used [176]. It identifies the PF candidates which originate from

pileup vertices and removes them from the collection used to cluster jets. Residual jet energy corrections are applied to correct for differences between data and simulation [177]. A large amount of noise in the ECAL endcaps during the 2017 data taking period caused a disagreement between the data and simulation. Therefore, jets with $p_T < 50$ GeV and $2.65 < |\eta| < 3.10$ are removed from the analysis of the 2017 data set.

Jets containing b-quarks are identified with a DeepCSV algorithm [178]. The Medium working point is used which corresponds to approximately 70% identification efficiency of b-jets with the misidentification of jets from light quarks/gluons at the level of 1%.

The missing transverse energy (MET) $\vec{p}_T^{\,\text{miss}}$ is reconstructed as the momentum imbalance in the transverse plane [179]. $\vec{p}_T^{\,\text{miss}}$ is calculated as a negative vectorial sum of the reconstructed PF candidates in the event with the jet energy corrections being taken into account. Pileup effects are mitigated with a pileup per particle identification (PUPPI) algorithm [180] which assigns a weight to each PF candidate that indicates the likelihood of the candidate to originate from a pileup interaction. These weights are further used to rescale the four-momentum of the PF candidates, which showed to improve both jet and MET observables comparing to the CHS method.

### 5.3.5 Primary vertex

Vertex corresponding to a primary $pp$ interaction is reconstructed in two steps:

1. An initial collection of primary vertex (PV) candidates is obtained by clustering tracks with a deterministic annealing algorithm [181]. A vertex having the largest value of $\sum p_T^2$ of the physics objects in the event (jets reconstructed from the tracks assigned to a candidate vertex and MET) is selected as a primary vertex.

2. A refitting procedure with an adaptive vertex fitter [182] is performed to improve the PV position resolution. Tracks originating from $\tau$ decay are removed from the fit in order to remove the bias arising from the displacement of the $\tau$ decay vertex. An additional constraint to the LHC beam spot – 3-D region where LHC beams collide in the CMS detector – is added.

The position and covariance matrix of the beam spot are precisely measured as an average over multiple collision events [76]. Therefore, using the beam spot information as an initial estimate of the PV position and uncertainty in the fit instead of a default fit configuration improves the fitting convergence and the PV position resolution in the transverse plane by a factor of 3 (Fig. 5.6).

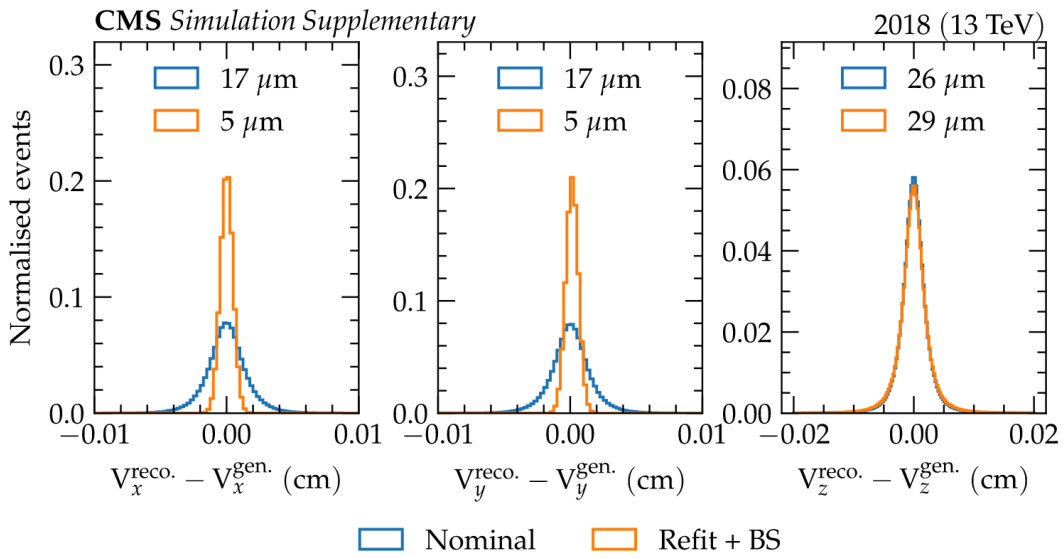


Figure 5.6: Difference between the generator-level and reconstructed primary vertex position for the $x$ (left), $y$ (middle), and $z$ (right) coordinates. Blue (orange) histogram corresponds to the nominal (refitted beamspot-corrected) PV reconstruction as described in step 1 (2) in Sec. 5.3.5

### 5.3.6 Impact parameter

In order to perform CP analysis in some of the $\tau\tau$ final states reconstruction of the impact parameter (IP) – a vector from the PV to the point of the closest approach of a charged particle track to PV – is needed for the charged prongs originating from $\tau$ decays (Fig. 5.7). To improve the IP resolution, a dedicated approach is developed. Contrary to another method using a tangent track extrapolation (Fig. 5.7), it parametrises the particle trajectory in the magnetic field as a helix $\vec{x}(t)$ and minimizes the distance between the trajectory and the primary vertex $\vec{d}(t) = |\vec{x}(t) - \overrightarrow{PV}|$. The resulting vector obtained after the minimisation IP $= \vec{x}(t_{\min}) - \overrightarrow{PV}$ is used as an impact parameter vector. It should be noted, that the constructed IP vector is used only for the $\phi_{CP}$ and IP significance (described below) computation. The selection requirements mentioned in Sec. 5.3.1, 5.3.2, and 5.3.3 are applied on the impact parameter vectors computed by the minimisation in the transverse plane in contrast to the 3D minimisation described in this section.

Furthermore, the minimisation procedure allows for a propagation of track parameter uncertainties to the impact parameter vector, therefore enabling the construction of an impact parameter significance variable:

$$S_{\mathrm{IP}} = \frac{|\mathrm{IP}|}{\sigma(\mathrm{IP})}, \tag{5.10}$$

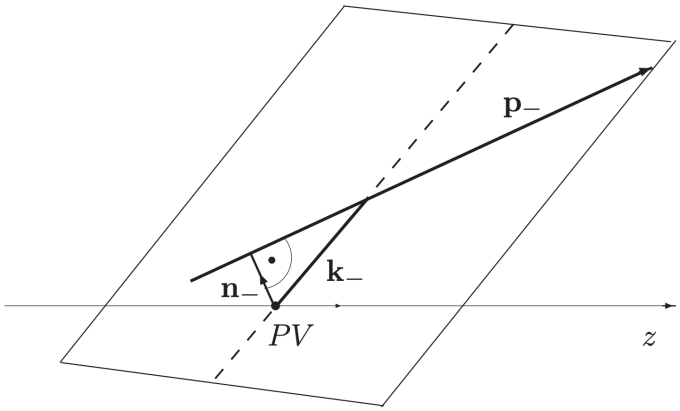


Figure 5.7: Illustration of the impact parameter vector $\boldsymbol{n}_-$ reconstruction via a tangent method for a decay of a tau lepton with momentum $\mathbf{k}_-$ to a one charged prong with momentum $\mathbf{p}_-$ in a laboratory frame [183]. The impact parameter vector is obtained by extrapolating $\mathbf{p}_-$ in the direction of the primary vertex (PV) from the first reference point on the track (the tracker hit nearest to PV, on the illustration corresponding to the intersection of the dashed and $\mathbf{p}_-$ lines). A vector pointing from PV to the point of the closest approach on the extrapolated tangent is an impact parameter vector.

where $\sigma(\mathrm{IP}) = \frac{\vec{\mathrm{IP}}^{\mathrm{T}}}{|\vec{\mathrm{IP}}|}\Sigma_{\vec{\mathrm{IP}}}\frac{\vec{\mathrm{IP}}}{|\vec{\mathrm{IP}}|}$ and $\Sigma_{\vec{\mathrm{IP}}}$ is the covariance matrix for the impact parameter vector derived with the error propagation. $S_{\mathrm{IP}}$ variable is further used in the event categorisation step described in Sec. 5.6.

### 5.3.7 $\phi_{CP}$ observable

#### Introduction

As described in Sec. 5.1 and specifically in Eq. 5.5, the CP nature of the Higgs boson coupling with tau leptons can be accessed through the spin correlations of the tau leptons resulting from its decay. However, it is not straight-forward *a priori* how to analyse this correlations experimentally. Furthermore, the situation is also complicated by the necessity to reconstruct the Higgs rest frame, which is not available in $pp$ collision at LHC.

The following approach is proposed by Berge et al. [146, 184, 185]. Firstly, considering the general form of the tau lepton decay via a charged prong $\tau^{\pm} \rightarrow a^{\pm} + X$ with $a^{\pm} \in$

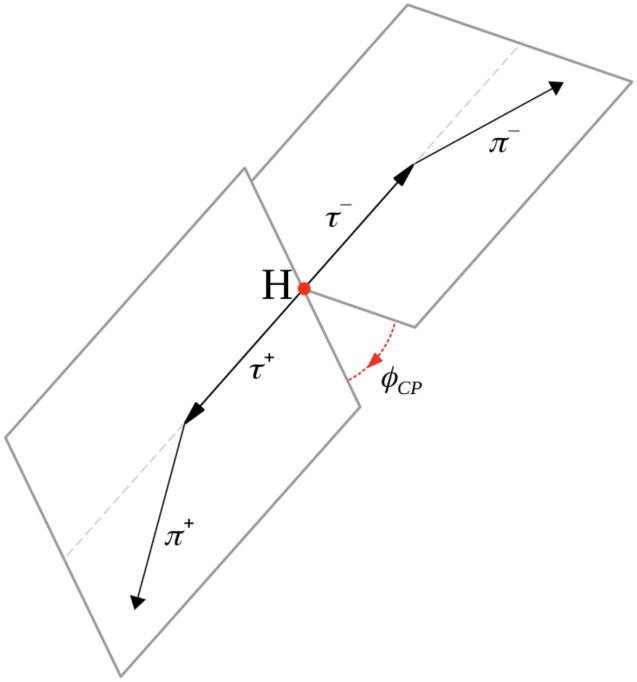


Figure 5.8: Illustration of the H → $\tau\tau$ in its rest frame where each $\tau$ decays into a single charged pion [186]. The $\phi_{CP}$ angle between the tau lepton decay planes is shown as a red arrow.

$\{e^{\pm}, \mu^{\pm}, \pi^{\pm}, \rho^{\pm}, a_1^{L,T,\pm}\}$, one obtains the partial decay width of the tau lepton:

$$\Gamma_a d\Gamma(\tau^{\pm}(\boldsymbol{s}^{\pm}) \to a^{\pm}(q^{\pm}) + X) = n(E_{\pm}) \cdot [1 \pm b(E_{\pm})\boldsymbol{s}^{\pm} \cdot \boldsymbol{q}^{\pm}] \cdot dE_{\pm} \frac{d\Omega\pm}{4\pi}, \tag{5.11}$$

where $\boldsymbol{s}^{\pm}$ is a normalised spin vector of the tau lepton in its rest frame, $E_{\pm}$ and $\boldsymbol{q}^{\pm}$ are the energy and the direction of flight of $a^{\pm}$ in the tau rest frame. $n(E_{\pm})$ and $b(E_{\pm})$ are referred to as spectral functions [184].

Using Eq. 5.2 and 5.11 one obtains for the partial decay width of H → $\tau\tau$:

$$\frac{d\Gamma}{d\phi_{CP}}(\mathrm{H} \to \tau\tau) \sim 1 - \frac{\pi^2}{16} b(E^+)b(E^-) \cos\Big(\phi_{CP} - 2\alpha^{\mathrm{H}\tau\tau}\Big), \tag{5.12}$$

where a $\phi_{CP}$ observable is introduced as the angle between the tau lepton decay planes in the Higgs rest frame (Fig. 5.8). However, since the latter cannot be reconstructed in $pp$ collisions, a zero-momentum frame (ZMF) using the charged decay products of the tau leptons is used in this work as an approximation of the Higgs rest frame. This might potentially reduce the

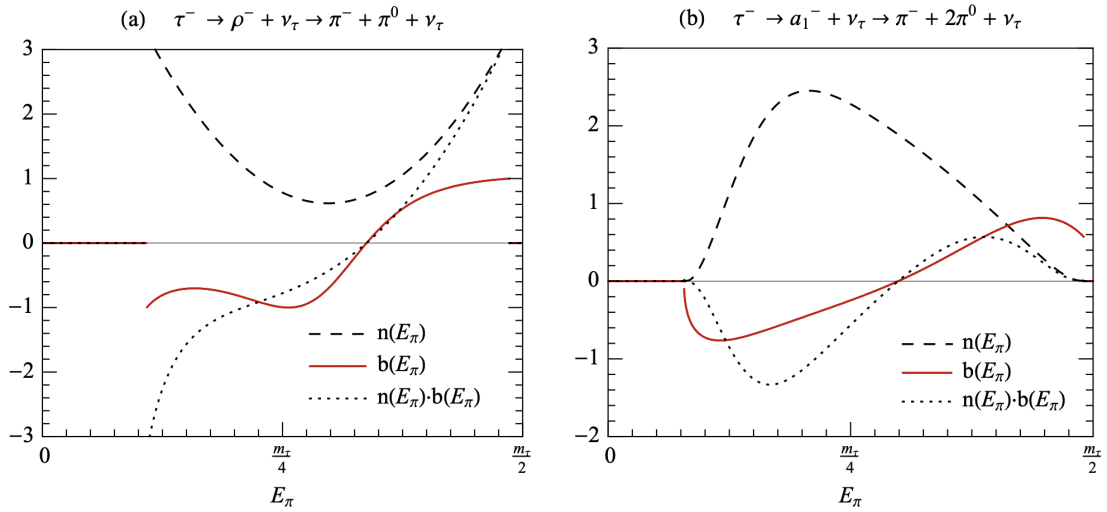


Figure 5.9: Spectral functions $n(E_\pi)$ and $b(E_\pi)$ for the charged pion in $\tau^- \rightarrow \rho^-\nu_\tau \rightarrow \pi^-\pi^0\nu_\tau$ (left) and $\tau^\pm \rightarrow a_1^-\nu_\tau \rightarrow \pi^-\pi^0\pi^0\nu_\tau$ (right) decays as a function of the charged pion energy ($E_\pi$) in the tau rest frame [184]. The values of $n(E_\pi)$ and $b(E_\pi)$ are given in units of GeV$^{-1}$.

overall sensitivity of the analysis, therefore hinting towards further studies of the ditau system reconstruction in $pp$ collisions.

One can contrast Eq. 5.12 with Eq. 5.5 and observe that $\phi_{CP}$ angle effectively resembles the angle between the transverse spin vectors of the tau leptons. The $\tau$ decay product topology can therefore be viewed as having a spin analysing power which allows to access the spin information experimentally. However, this power is dependent on the $\tau$ decay mode and on the properties of the charged prong as encoded with the spectral functions. The functions show complex behaviour (Fig. 5.9) and for some scenarios can change their sign therefore affecting the separation between pure scalar and pseudoscalar hypotheses. No dedicated optimisation of CP sensitivity is carried out in this work as the analysis is largely limited by the available statistics. This strongly affects the room for optimisation of the event selection with respect to the spectral functions as it will further reduce the amount of signal candidates. However, as more data will be available in the future, such optimisation can be carried out in the context of, for example, the differential measurement of CP H $\rightarrow \tau\tau$ properties.

Similarly to Eq. 5.5, a CP mixing angle $\alpha^{\mathrm{H}\tau\tau}$ enters in Eq. 5.12 as a phase shift of the cosine distribution. Therefore, given enough sensitivity one would be able to gauge the CP nature of the H$\tau\tau$ coupling by the shift of the modulation from the expected SM (CP-even) scenario (Fig. 5.10). It should also be noted that for the $Z/\gamma^* \rightarrow \tau\tau$ process, which constitute one of the major background in this work, the distribution of $\phi_{CP}$ observable is uniform at the generator level.

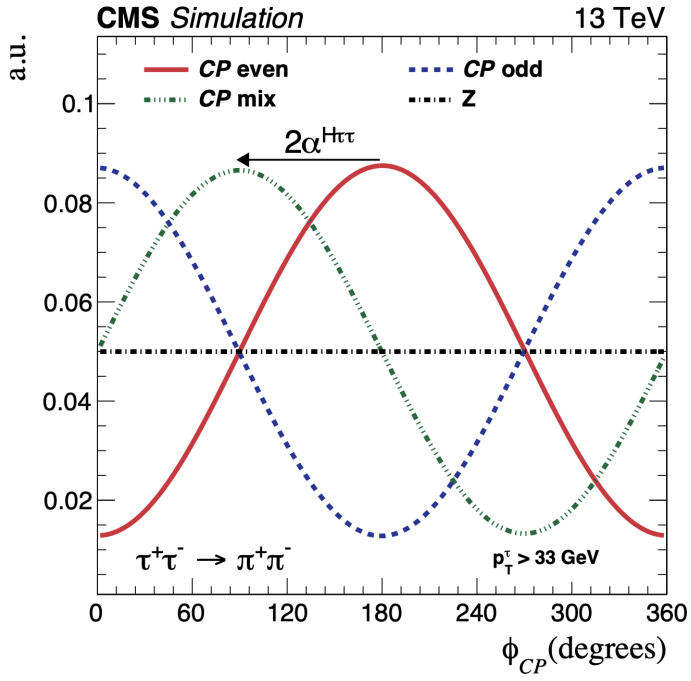


Figure 5.10: The distribution of $\phi_{CP}$ angle in the Higgs rest frame at the generator level for the H → $\tau\tau$ events where both tau leptons decay into a charged pion and a neutrino [186]. The hypotheses of a scalar (solid red), pseudoscalar (dashed blue), and a CP mixture with $\alpha^{H\tau\tau} = 45°$ (dash-do-dot green) Higgs boson as well as a Z vector boson (dash-dot black) are shown.

However, at the detector level the limited track and PV resolution leads to the distortion of the $\phi_{CP}$ distribution both for the signal and background processes [185]. For example, for the $\tau_e\tau_h$ analysed in this work, for the $e\pi$ final state (one tau decays into electron and neutrinos, the other into a charged pion and neutrinos) it is visible on Fig. 5.11 (left) that the distribution of the simulated Drell-Yan events is not uniform and peaks towards 0 and $2\pi$ values of $\phi_{CP}$. This is due to the the PV misreconstruction effects that on average pull the reconstructed IP vectors towards smaller values, which consequently translates to $\phi_{CP}$ values. This effect is pronounced only for the final states where IP vectors are used for the reconstruction of $\phi_{CP}$ for both tau leptons, as described below in this section. Despite these effects destruct the uniformity of the Drell-Yan background events, some symmetries can still be used in the construction of the templates for the statistical inference, as described in Sec. 5.7. For the H → $\tau\tau$ events, the modulations are clearly visible at the reconstruction level for all the final states being considered, and the pure CP even and CP odd hypotheses are separable.

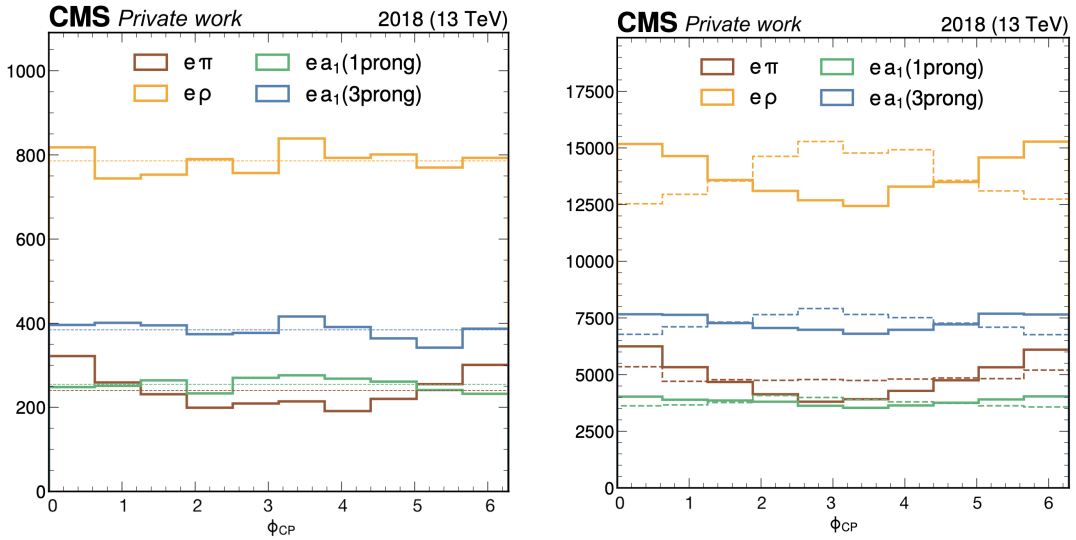


Figure 5.11: Unnormalised $\phi_{CP}$ distributions for the simulated Drell-Yan (left) and H → $\tau\tau$ (right) events on the reconstructed level for four final states analysed in this work. Dashed lines on the left figure show the mean of the corresponding histogram counts. On the right figure, solid (dashed) lines represent the distribution of the scalar (pseudoscalar) H → $\tau\tau$ hypothesis. Histogram counts are in auxiliary units.

### Methods

By definition, $\phi_{CP}$ is the angle between the tau lepton decay planes in the Higgs rest frame. Intuitively, this can be reconstructed for the tau leptons both decaying into at least two reconstructable objects by using the momenta vectors of the latter. However, if one of the tau leptons decays into a single charged prong and neutrino, it is no longer possible to define in this way its decay plane because of the neutrino escaping detection. This is particularly the case for the $\tau_e\tau_h$ final state analysed in this work, where one of the tau leptons decays into an electron and two neutrinos.

Since there are two decay planes involved in the $\phi_{CP}$ computation, the problem factorises into the problem of reconstructing separately a decay plane for each of the tau leptons followed by computing the angle between them. Furthermore, there are generally two four-vectors in the laboratory frame needed to construct the plane in the ZMF. To unify the notation across various decay modes, these are further referred to as $\lambda^{\pm}$ ($\lambda^{ZMF^{\pm}}$) and $q^{\pm}$ ($q^{ZMF^{\pm}}$) in the laboratory (zero-momentum) frame, where ± refers to the charge of the tau lepton. Vectors in the ZMF are obtained by the Lorentz boost from the laboratory frame. Depending on the decay mode, the four-vectors are constructed using various approaches, as described further.

A transverse component of $\lambda^{ZMF^{\pm}}$ with respect to $q^{ZMF^{\pm}}$ is derived and the corresponding

normalised unit vector is denoted as $\hat{\lambda}_{\perp}^{ZMF^{\pm}}$. Then, variables $\phi^{ZMF}$ and $O^{ZMF}$ are defined as:

$$\phi^{ZMF} = \arccos\left(\hat{\lambda}_{\perp}^{ZMF^{+}} \cdot \hat{\lambda}_{\perp}^{ZMF^{-}}\right), \tag{5.13}$$

$$O^{ZMF} = \hat{q}_{\perp}^{ZMF^{-}} \cdot (\hat{\lambda}_{\perp}^{ZMF^{+}} \times \hat{\lambda}_{\perp}^{ZMF^{-}}). \tag{5.14}$$

Finally, one obtains $\phi_{CP} \in [0°, 360°]$ angle as:

$$\phi_{CP} = \begin{cases} \phi^{ZMF} & \text{if } O^{ZMF} \geq 0 \\ 360° - \phi^{ZMF} & \text{if } O^{ZMF} < 0 \end{cases} \tag{5.15}$$

For the decays into one charged prong $\tau^- \to e^- \bar{\nu}_e \nu_\tau$ and $\tau^- \to \pi^- \nu_\tau$, an **impact parameter (IP) method** is used to construct the spanning four-vectors (Fig. 5.12, left). In these decay modes there is only one momentum vector available as $q^{\pm}$. Therefore, the impact parameter vector (Sec. 5.3.6) of a charged prong is used as $\lambda^{\pm}$ to be able to define a decay plane. It should be noted that in this case the resulting decay plane is not a "genuine" one, but rather a necessary approximation.

For the decays $\tau^- \to \rho^- \nu_\tau$, $\tau^- \to a_1^-(1\text{pr})\nu_\tau \to \pi^{\mp}\pi^0\pi^0\nu_\tau$, and $\tau^- \to a_1^-(3\text{pr})\nu_\tau \to \pi^{\mp}\pi^{\pm}\pi^{\mp}\nu_\tau$, a **neutral-pion (NP) method** is used (Fig. 5.12, right). For the $\tau \to \rho$ case, a four vector of the neutral pion resulting from the $\rho$ meson decays is taken as $\lambda^{\pm}$. For this four-vector the energy is set to the sum of energies of electrons and photons collected by the HPS algorithm, the momentum direction is taken as the direction of the leading electron/photon, and the mass is set to the $\pi^0$ mass.

For the $\tau \to a_1(1\text{pr})$ case, all the electromagnetic constituents from the $a_1^-(1\text{pr})$ decay are combined together and treated analogously to the $\tau \to \rho$ case. Additionally, in order to avoid destructive interference between longitudinal and transverse polarised components of the intermediate mesons, in all the decay modes where the NP method is used the components are separated by the following variable:

$$y^{\tau^{\pm}} = \frac{E_{\pi^{\pm}} - E_{\pi^0}}{E_{\pi^{\pm}} + E_{\pi^0}}, \; y^{\tau} = y^{\tau^+} y^{\tau^-}, \tag{5.16}$$

where $E_\pi$ is the energy of the pion in the laboratory frame. If $y^\tau < 0$, $\phi_{CP}$ is recomputed with a shift as $\phi_{CP} \to \phi_{CP} - 360°$.

For the $\tau \to a_1(3\text{pr})$ case, a pair of oppositely charged pions with the mass closest to the $\rho^0$ meson mass is chosen. Out of these two pions, the one with the charge of the tau lepton is used for the definition of the ZMF and the $q^{\pm}$ vector. The pion with the charge opposite to the one of the tau lepton is treated like a neutral pion. Then, the neutral-pion is applied as described for the $\tau \to \rho$ case.

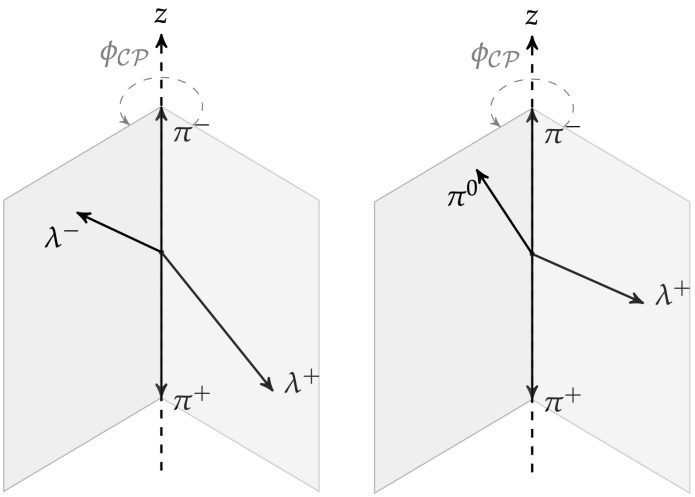


Figure 5.12: Illustration of the $\phi_{CP}$ angle construction for two decay topologies [186]. Left: decay planes for both tau leptons are defined with the impact parameter method as spanned by the momentum and IP vector of the corresponding charged prong. Right: for one tau lepton the impact parameter method is used and the neutral-pion method for the other with the plane spanned by the momenta of the charged and the neutral pions. Vectors are shown in the zero-momentum frame constructed from the momenta of the charged constituents in the H $\rightarrow \tau\tau$ decay.

To summarise, the final states to be analysed with the corresponding methods to reconstruct decay planes are:

- $e\pi$ ($\tau\tau \rightarrow e^{\pm}\pi^{\mp} + 3\nu$) $\longrightarrow$ IP + IP $\longrightarrow$ Fig. 5.12, left,
- $e\rho$ ($\tau\tau \rightarrow e^{\pm}\rho^{\mp} + 3\nu$) $\longrightarrow$ IP + NP $\longrightarrow$ Fig. 5.12, right,
- $ea_1$(1pr, 3pr) ($\tau\tau \rightarrow e^{\pm}a_1^{\mp}$(1pr, 3pr) + $3\nu$) $\longrightarrow$ IP + NP $\longrightarrow$ Fig. 5.12, right.

## 5.4 Event selection

The following procedure to select events for the analysis of the $\tau_e\tau_h$ final state, where one tau lepton decays into an electron plus neutrinos ($\tau_e$) and the other into hadrons plus neutrino ($\tau_h$), is followed:

1. Events are selected online by the CMS trigger system (Sec. 3.2.6). Either a cross trigger

($e$ & $\tau_h$) or a single electron trigger ($e$) should be fired depending on the data taking year (online $p_T$ thresholds for the corresponding objects are shown in brackets in GeV):

- 2016: $e(25)$,
- 2017: $e(27)$ OR $e(24)$ & $\tau_h(30)$,
- 2018: $e(32)$ OR $e(24)$ & $\tau_h(30)$.

2. For each event, pairs of oppositely charged electron and hadronically decaying tau lepton reconstructed offline (Sec. 5.3.1 and Sec. 5.3.3) are selected with the requirement of being sufficiently separated (cone distance $\Delta R > 0.5$). These offline objects are required to match the corresponding online objects within the cone distance $\Delta R < 0.5$. Offline electron ($\tau_h$) objects should have $p_T$ at least 1(5) GeV higher when the online $p_T$ thresholds for the corresponding trigger legs.

3. Events with an additional loosely identified electron or muon as well as a pair of electrons are vetoed.

4. Events containing a single jet (jets) with $p_T > 25$ GeV and $|\eta| < 2.4$ passing the Medium (Loose) working point of the DeepCSV classifier (Sec. 5.3.4) are vetoed.

5. The reconstructed electron candidate is selected with $p_T > 25$, $|\eta| < 2.1$ as well as IP, identification, and isolation requirements described in Sec. 5.3.1.

6. The reconstructed $\tau_h$ candidate is selected with $p_T > 20$, $|\eta| < 2.3$ ($|\eta| < 2.1$ for the cross trigger) as well as IP and identification requirements described in Sec. 5.3.3.

7. The transverse mass of the electron candidate and the missing transverse energy $\vec{p}_T^{\,miss}$ is defined as:

$$m_T \equiv \sqrt{2p_T^e p_T^{miss}[1 - \cos(\Delta\phi)]}, \tag{5.17}$$

with $\Delta\phi$ denoting the azimuthal angle between the vector of electron transverse momentum $\vec{p}_T^{\,e}$ and $\vec{p}_T^{\,miss}$. The requirement $m_T < 50$ GeV is applied in order to reject the background from the W+jets process.

8. In case there are multiple $\tau_e\tau_h$ candidates in the event, the pairs are ranked firstly with the highest priority given to the pairs with the most isolated electron, then with the electron with the highest $p_T$, then with the $\tau_h$ candidate with the highest DeepTau against jet score, then with the $\tau_h$ candidate with the highest $p_T$.

## 5.5 Background estimation

The main background sources in the $\tau_e \tau_h$ final state can be roughly classified as those involving genuine tau leptons, jets faking $\tau_h$ (jet $\rightarrow \tau_h$), and prompt/non-prompt leptons faking $\tau_h$ ($l \rightarrow \tau_h$). In terms of physical processes, the expected contributions are from the Drell-Yan, QCD multijet, top quark-antiquark pair production (bar), single top quark production (ST), W+jets, and diboson production processes. In this work, the backgrounds are largely modelled with data-driven methods: a $\tau$-embedding technique [187] is used to model background with two genuine tau leptons, and a "fake factor" ($F_F$) method [188] is used to model jet $\rightarrow \tau_h$ background. These two methods together account for approximately 90% of all background processes. Other minor background processes are modelled from the simulation where events with a pair of genuine tau leptons or with a jet faking $\tau_h$ are removed to avoid double-counting.

### 5.5.1 $\tau$-embedding method

The most challenging part in the simulation of the $Z/\gamma^* \rightarrow ll$ process is to model jets originating from partons emitted in the initial state radiation. This translates, after hadronisation, in the multiplicity of jets in the final state and the corresponding hadronic activity in the detector. Its modelling would require resource-demanding simulation of samples at NLO and higher orders, which in practise still does not guarantee adequate description of event observables in data. Therefore, finding a way to model the $Z$+jets process without relying only on simulation would be beneficial.

A $\tau$-embedding method is designed with that purpose and builds up on an idea of using the lepton universality to extract and transfer the detector activity from data to simulation across leptonic final states in $Z/\gamma^* \rightarrow ll$. It proceeds with the following steps (Fig. 5.13):

1. A sample of $Z \rightarrow \mu^+\mu^-$ events is recorded in data with a dedicated dimuon trigger. This sample of muon pairs is of the highest purity thanks to the excellent muon reconstruction in CMS. However, an excessively tight selection aimed at high purity might introduce a bias to the selected events as having, for example, little hadronic activity in the detector. Therefore, only a loose kinematic selection without any isolation requirement is applied. The final dimuon sample has the $Z \rightarrow \mu^+\mu^-$ purity of 99.11% for $m_{\mu\mu} > 70$ GeV. Remaining contributions come from $t\bar{t}$ (0.55%), diboson and single top (0.17%), QCD (0.10%), $Z \rightarrow \tau\tau$ (0.05%), and W+jets (0.02%) processes.

2. All traces in the detector which are associated with the muons are removed from the event. This includes hits in the tracking system and muon chambers, plus energy deposits in the calorimeter which are compatible with the fitted global-muon track.

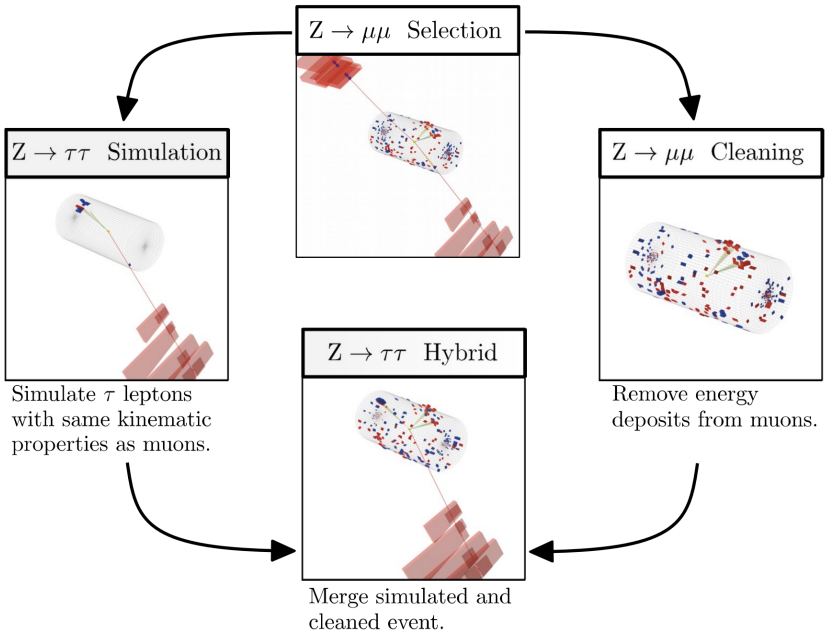


Figure 5.13: Illustration of the $\tau$-embedding steps as described in Sec. 5.5.1 [187].

3. A pair of tau leptons with the kinematic properties of the two muons in data is simulated with PYTHIA and passed through the empty detector environment (no other particles, no pileup). The tau leptons are forced to decay into a predefined $\tau_e \tau_h$ final state with a branching fraction 100%. However, electrons and muons can also in principle be simulated and injected for validation purposes.

4. A hybrid event is created as an overlay of the event with the removed muon traces and the event with the simulated pair of tau leptons. The combination is performed at the reconstruction level of physics objects (tracks, $e/\mu$, calorimeter clusters). The resulting sample of hybrid events can be used in the analysis to model the backgrounds with two genuine $\tau$ leptons.

Overall, the $\tau$-embedding method provides a fully data-driven description of detector activity in $Z/\gamma^* \rightarrow ll$ events. This saves both computational resources for simulation of the highly dense pileup environment and provides an excellent description of jet-related physical observables. Furthermore, only systematical uncertainties related to the simulated pair of tau leptons and the selection efficiency of $\mu^+\mu^-$ pairs in data have to be introduced, therefore improving the sensitivity of the analysis.

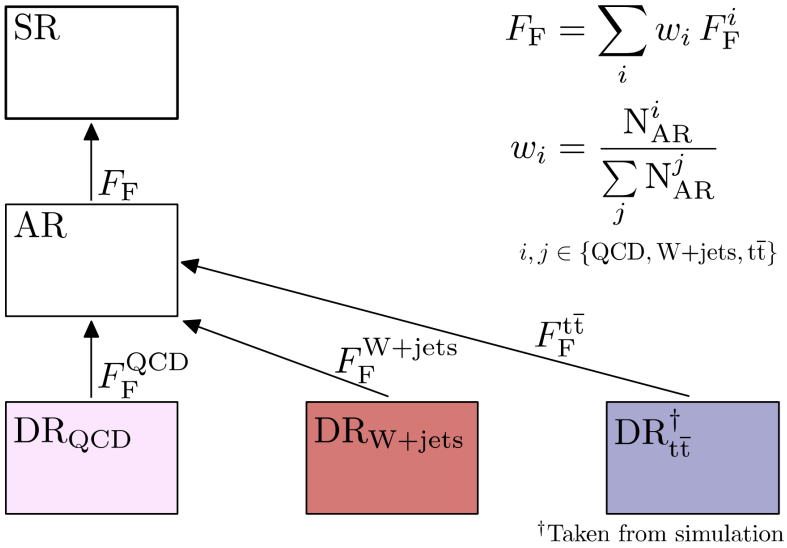


Figure 5.14: Illustration of the $F_F$ method steps as described in Sec. 5.5.2 [189].

### 5.5.2 $F_F$ method

The jet $\rightarrow \tau_h$ background constitutes another important background in the analysis. It is driven by the presence of QCD jets which fake rate to $\tau_h$ is hard to model and requires computationally intensive simulation to reach the desired level of statistics. Therefore, a data-driven approach is also desired in this case.

The fake factor method is proposed to tackle the modelling of this background. It proceeds with the following steps (Fig. 5.14):

1. A signal region (SR) is defined in data as described in Sec. 5.4. This includes the nominal Medium working point of the DeepTau ID discriminator against jets applied to the $\tau_h$ candidate (Sec. 5.4).

2. An application region is defined in data with all the selection requirement as in SR, except that the WP selection against jets for the $\tau_h$ candidate is inverted: the candidate is required to pass the loosest VVVLoose WP (nominal $\tau_h$ efficiency 98%) and fail the Medium WP. Events in the AR will serve as an estimate of the jet $\rightarrow \tau_h$ background in SR once assigned a fake factor weight $F_F$ on the event-by-event basis. The latter is derived as follows:

$$F_F = \sum_i w_i \cdot F_F^i, \tag{5.18}$$

$$w_i = \frac{N_{AR}^i}{\sum_j N_{AR}^j}, \tag{5.19}$$

$$F_F^i = \frac{N_{DR}^i(\text{Medium})}{N_{DR}^i(\text{VVVLoose \& !Medium})}. \tag{5.20}$$

Here the final FF weight is obtained as a weighted sum across the jet $\rightarrow \tau_h$ background processes $i \in \{\text{QCD, W+jets, } t\bar{t}\}$. The weights $w_i$ correspond to the fraction of the background process $i$ in all jet $\rightarrow \tau_h$ events in AR. $F_F^i$ for each process is obtained as the number of events passing the Medium WP against jets divided by the number of events passing the VVVLoose WP and failing the Medium one in a so-called determination region (DR). Both $w_i$ and $F_F^i$ are parametrised functions of several variables as described below.

3. The determination region (DR) is constructed depending on the background process $i$. For $t\bar{t}$, it is not straight-forward to find sufficiently pure region in data, therefore the fake factor is taken from the simulation of the $t\bar{t}$ process. For W+jets, DR region is defined in data with inverting the transverse mass requirement (Sec. 5.4) $m_T > 70$ GeV with all the other selection remaining the same as in SR. For QCD, DR region is defined in data with inverting the opposite-sign requirement, i.e. electron and $\tau_h$ candidates are required to have the same charge. Additionally, $I_{rel}^e > 0.05$ requirement is applied to remove events with genuine tau leptons. All the other selection criteria are the same as in SR.

4. For QCD and W+jets, $F_F^i$ are measured in bins of $N_{jets}$, MVA DM (Sec. 5.3.3), where the MVA DM equal to 0 is further split into two bins on the IP significance (Sec. 5.3.6) $S_{IP} < 1.5, S_{IP} \geq 1.5$. $F_F^i$ are measured separately for events passing the single $e$ and $e$ & $\tau_h$ cross triggers. For $t\bar{t}$, $F_F^i$ are measured only in bins of MVA DM and $S_{IP}$. The contribution of other processes to each of DRs is subtracted using simulation. For each bin, $F_F^i$ is parametrised as a function of $\tau_h$ $p_T$ as obtained from the polynomial fit to the $F_F^i$ distribution.

5. Fractions of the processes in AR $w_i$ are parametrised in bins of the dedicated BDT score. This is motivated by the fact that it is difficult to find a small set of variables providing a good differentiation between the jet $\rightarrow \tau_h$ processes. Therefore, a summary statistics is constructed as a BDT output which is trained to distinguish between the three background processes QCD, W+jets, and $t\bar{t}$ using kinematic information about the ditau system, $p_T^{miss}$, and $N_{jets}$ as input variables. Fractions are then computed according to Eq. 5.20 in bins of W+jets and QCD BDT scores for W+jets, $t\bar{t}$ (both taken from simulation) and QCD (taken from data with all other processes subtracted with simulation) processes.

6. Fake factor weights are computed according to Eq. 5.19 in AR, where the contribution of processes with genuine tau leptons and $l \rightarrow \tau_h$ fakes is subtracted using simulated events.

7. Corrections are applied to account for discrepancies in the closure tests when $F_F^i$ are applied to events in the corresponding DR. Differences between DR and AR resulting in a different jet $\rightarrow \tau_h$ rates between these two regions are also accounted for in the corresponding corrections.

### 5.5.3 Corrections

While most of the background is estimated from data, it still relies to a certain extent on simulation which requires dedicated corrections to be applied in order to refine its modelling of data.

The corrections applied to the embedded samples are:

- Electron tracking/ID/isolation/trigger scale factors (SFs),
- Electron energy scale (ES) and resolution smearing corrections,
- Hadronic tau ID/trigger SFs,
- Hadronic tau ES corrections.

The scale factors are derived with a tag-and-probe method [190] generally as a function of e/$\tau_h$ $p_T$, $\eta$, and MVA DM using $Z/\gamma^* \rightarrow ll$ events. These are aimed to account for the mismatch in the corresponding ID/trigger/isolation selection efficiencies between data and simulation. The energy scale corrections are derived by varying the lepton energy scale in simulation and performing the maximum likelihood fit to physical observables in data (e.g. $\tau_h$ and $\mu\tau_h$ invariant mass ($m_{vis}$) for $\tau_h$ case) in order to find the most optimal value. The value of the energy scale shift corresponding to the minimum of the negative log-likelihood is taken as the correction factor.

It should be noted that while tau leptons in the embedded samples are still simulated, the corresponding corrections are derived specifically for the embedded samples and differ from the same corrections applied to the simulated samples.

The corrections applied to the simulated samples include those applied to the embedded samples, plus the following ones:

- Pileup reweighting. The distribution of PU interactions in simulated samples is reweighted to match the one observed in data.

- $e \rightarrow \tau_h$ fake rate and ES corrections. The corrections are obtained for the DeepTau discriminator against electrons similarly to e.g. $\tau_h$ ID SFs using the tag-and-probe method with $Z \rightarrow ee$ events.

- MET recoil corrections. The corrections are applied to Drell-Yan, W+jets, and Higgs simulated samples and aim to correct for the mismodelling of $\vec{p}_T^{\,miss}$. The corrections are derived with $Z \rightarrow \mu\mu$ events on the variable defined as the vectorial difference between measured $\vec{p}_T^{\,miss}$ and the sum of the transverse momenta of neutrinos from a boson decay.

- b-tagging efficiency corrections. Since veto is applied on events where the jets pass certain working points of the DeepCSV classifier (Sec. 5.4), one needs to ensure that the mismodelling of the DeepCSV score is not propagated to the analysis. A so-called "promote-demote" technique is used which randomly assign/remove a given jet to/from b-tagged category in order to match WP selection efficiency in simulation to the one measured in data.

- $Z$ mass and $p_T$ reweighting. The corresponding spectra are corrected in simulation to better match those obtained in data for $Z \rightarrow \mu\mu$ events.

- Top quark $p_T$ reweighting. The distribution of the top quark transverse momentum is reweighted in the NLO simulated samples to match the distribution obtained from NNLO.

- Prefiring. During the 2016 and 2017 data taking periods it was observed that the L1 trigger system would sometimes "prefire", i.e. record an event corresponding to the previous bunch crossing. The issue was related to the shift in ECAL pulses and therefore a corresponding weight is introduced to recover for this effect [191].

Lastly, corrections to the impact parameter significance variable are applied both to embedded and simulated samples. As it is discussed in Sec. 5.6, a selection is applied based on this variable to further improve the sensitivity to the CP mixing angle. Therefore, it is important to ensure good modelling of data in this observable. A quantile mapping method is used to correct the IP vector coordinates and covariance matrix based on their cumulative distributions in data and simulation using $Z \rightarrow ee$ ($Z \rightarrow \mu\mu$) events for electron (pion) legs.

After estimating the backgrounds as described above and applying all the necessary corrections good agreement between data and simulation is observed for the key physical observables for all years (Fig. 5.15, Fig. 5.16, and Appendix A.7).

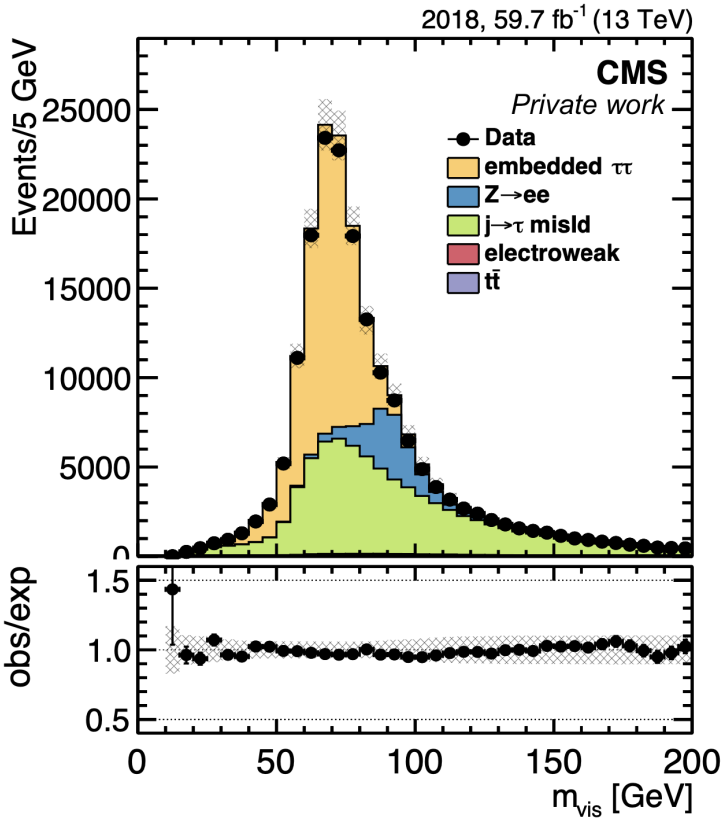


Figure 5.15: Comparison of data with simulation for the 2018 data-taking period for the visible mass of the $\tau_e\tau_h$ system.

## 5.6 Event categorisation

After the selection of $\tau_e\tau_h$ candidate pairs (Sec. 5.4), one has to perform their categorisation. Fundamentally, that means that for each candidate pair one wants to understand which physical process it originates from. Since the selection criteria do not yield the sample with only H → $\tau\tau$ events but the one which is contaminated by several background processes, one is interested in purifying this sample. This is hardly possible to perform manually due to an extremely large number of events to be categorised. Therefore, some automated refining procedure is needed.

One possible solution would be to follow a rule-based approach and define a set of criteria on custom variables constructed using expert knowledge. This is however a difficult task due to the multidimensionality of the problem, which moreover might not give an optimal result. Therefore, a Machine Learning (ML) approach is taken to classify events into predefined categories. These are defined as follows, together with the corresponding samples used for the composition of a training data set:

- **Signal:** to target signal H events originating from the ggH and VBF production processes. Events for the training are taken from the corresponding simulated samples.
- **Genuine** $\tau$: to target background events with two genuine tau leptons. These include $Z \rightarrow \tau\tau$, $t\bar{t}$, and diboson processes. Events for the training are taken from the corresponding simulated samples.
- **Fakes:** to target background events with jets or leptons faking $\tau_h$. These include $Z \rightarrow ll$, $Z \rightarrow$ jets, $t\bar{t}$, diboson, W+jets, and QCD processes. Events for the training are taken from the corresponding simulated samples except for QCD, which is estimated from data by inverting the opposite-sign requirement for the electron-$\tau_h$ pair.

A neural network (NN) is trained to leverage the multidimensionality of the problem and to construct an optimal classifier in an automated way from the following high-level input variables:

- $p_T(e)$, electron $p_T$,
- $p_T(\tau)$, $\tau_h$ $p_T$,
- $p_{T,\tau\tau} \equiv |\vec{p}_T(e)+\vec{p}_T(\tau)+\vec{p}_T^{\,miss}|$, vectorially combined $p_T$ of electron, tau lepton and missing transverse energy,
- $m_{vis}$, visible invariant mass of the electron and $\tau_h$ decay products,
- $m_{\tau\tau}$, invariant ditau mass obtained with the SVFit algorithm (Sec. 5.3.3),
- $E_T^{miss}$, missing transverse energy obtained with the PUPPI algorithm (Sec. 5.3.4),
- $m_T$, transverse mass of the electron and MET (Sec. 5.4),
- $N_{jets}$, number of jets in the event,
- $p_T$(jet1), $p_T$ of the jet with the highest $p_T$, referred to as the "leading" jet (in events with at least one jet),
- $p_T$(jet2), $p_T$ of the jet with the second highest $p_T$, referred to as the "trailing" jet (in events wit hat least two jets),
- $p_{T,jj}$, combined $p_T$ of the two leading jets,
- $m_{jj}$, invariant mass of the two leading jets,
- $\Delta\eta_{jj}$, difference in pseudorapidity between the two leading jets.

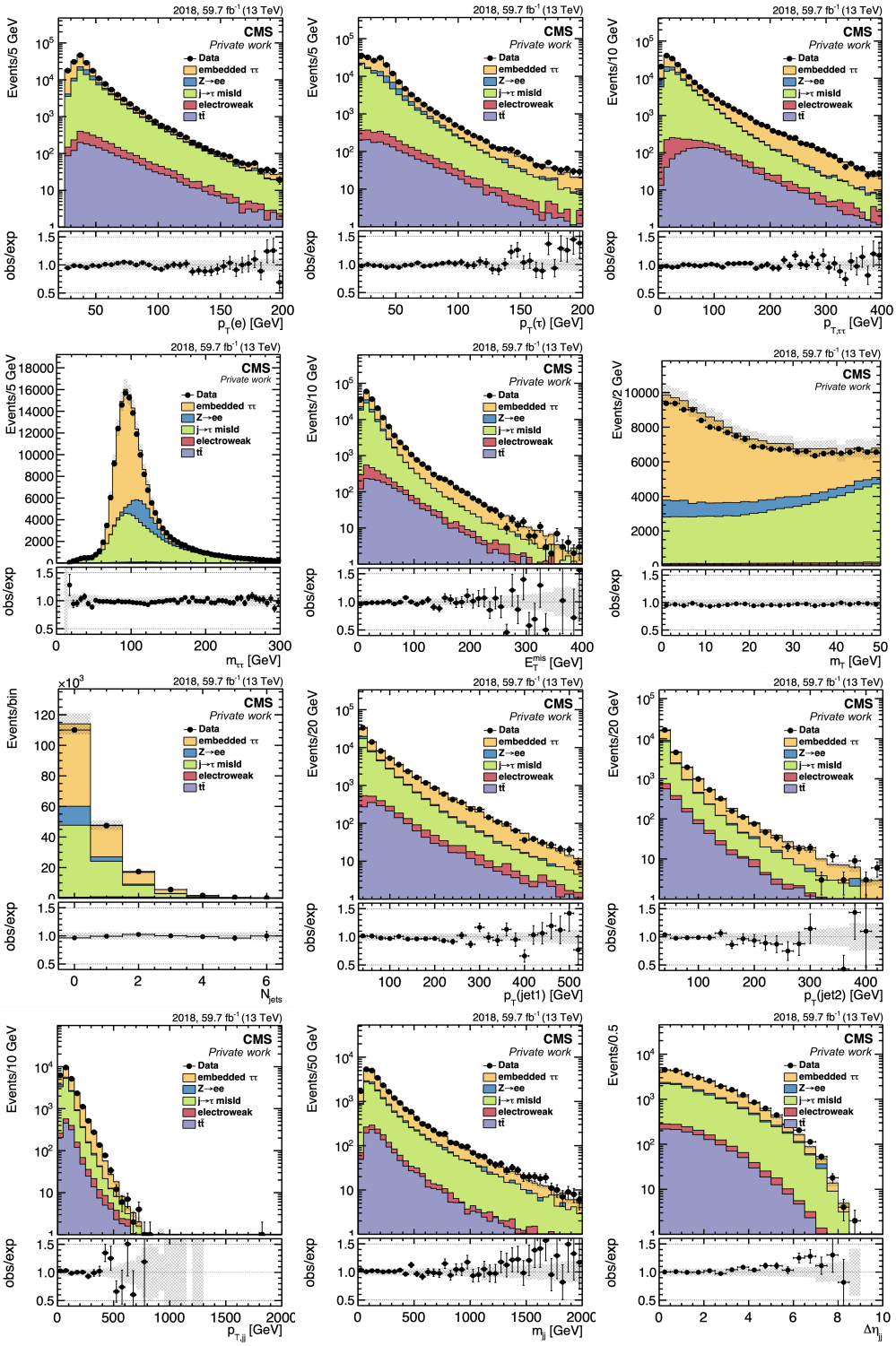


Figure 5.16: Comparison of data with simulation for the 2018 data-taking period for the variables used in the neural network training, as described in Sec. 5.6.

This set of variables is chosen as the ones which are known *a priori* to provide discriminating power between the classes and to be well-modelled in data (Fig. 5.15, Fig. 5.16, and Appendix A.7). Each event is therefore represented as a vector of length 13, which defines the input to the model.

The model architecture consists of three consecutive blocks. Each block has the same structure and is constructed from a feed-forward layer with 100 nodes, followed by a batch normalisation layer, a ReLU activation function, and a dropout layer with probability $p = 0.5$. The output layer is a feed-forward layer with three nodes normalised to sum up to 1 with a softmax function. It therefore defines the probability of an event to belong to either of the three classes, as defined above.

The batch size equals to 1000 and the training is performed using the TensorFlow library [112] until convergence with an early stopping in case of no validation loss improvement for 20 consecutive iterations. Three separate trainings are done with the same architecture used for each of the data-taking years (2016, 2017, 2018). Each training in fact corresponds to the training of two models in a "two-fold" manner. The training data set, consisting of the mixture of various data samples as defined above, is split into two parts based on an ID number which is unique for each event. Then, one network is trained on a half of the data set with even event IDs, while the other on the odd ones. At the prediction step, the networks change halves and the even network is applied to the odd half of the data set, and vice versa. This procedure allows to use all available simulated samples to produce templates for the statistical inference. Furthermore, no bias is introduced due to the usage of the same events for both training and template composition with the same model.

The loss function for each of the models is a categorical cross-entropy which is minimised with an Adam optimiser [114] with the learning rate $10^{-4}$. The training weights are added to the loss function to balance the difference in expected number of events for across physical processes in the corresponding data-taking period. 90% of the even/odd halves which are provided to each of the models for the training is used for the actual training, while the other 10% is used for validation.

After the training, each event for both data and simulated samples is classified to one of the three categories where the corresponding NN score is the highest. This score is also used further at the statistical inference step (Sec. 5.7).

Lastly, a requirement on the impact parameter significance $S_{IP} > 1.5$ is applied for the electron in events which are classified into a signal category. The same requirement $S_{IP} > 1.5$ is applied for the single charged pion from the $\tau_h$ candidate with MVA DM equal to 0 (Sec. 5.3.3) for events both in the signal and background categories. This selection requirement removes events with poorly reconstructed IP vectors. Furthermore, it showed to have slightly better separation between CP-even and CP-odd hypotheses at the reconstructed level without

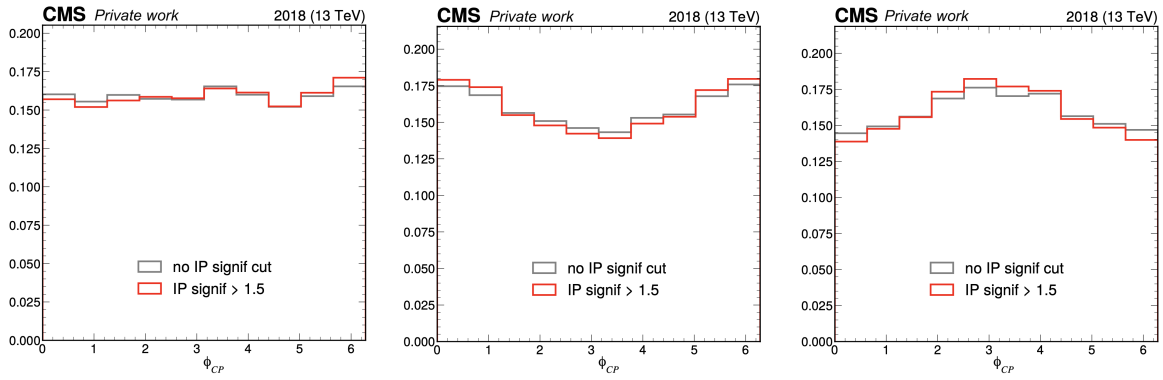


Figure 5.17: Normalised distributions of the $\phi_{CP}$ observable for events passing the selection requirements in Sec. 5.4 with the $\tau_h$ candidate identified with MVA DM equal to 1 ($e\rho$ final state) from the embedded samples (left), ggH sample under the CP-even hypothesis (middle), and ggH sample under the CP-odd hypothesis (right). The histogram in gray (red) corresponds to no ($S_{IP} > 1.5$) selection applied to the IP significance of the electron.

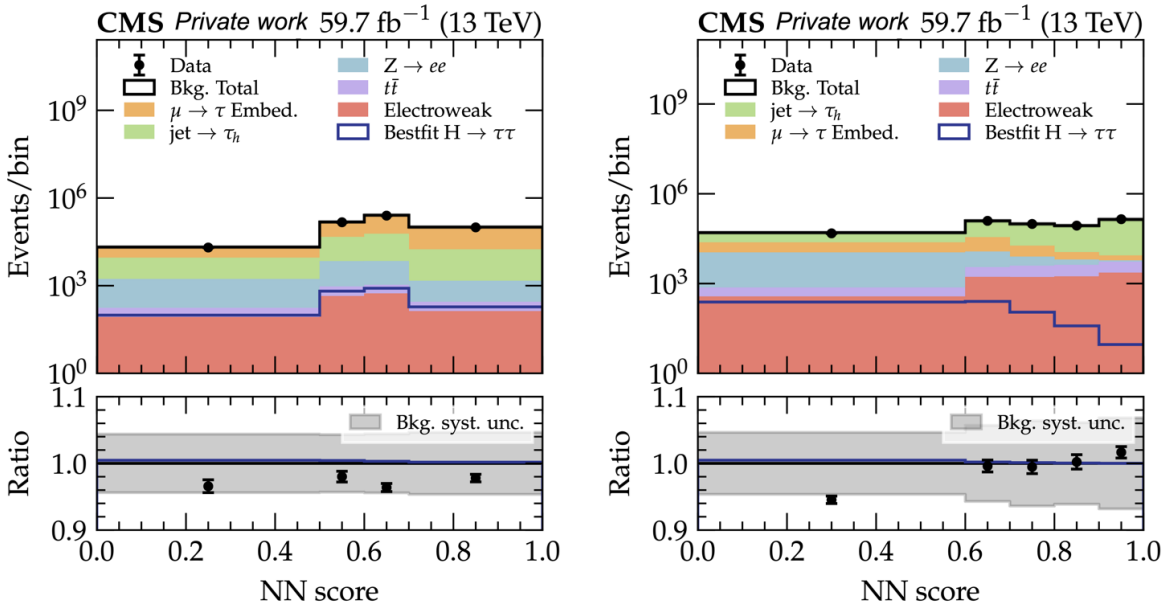


Figure 5.18: Pre-fit distribution of the NN score in the genuine $\tau$ (left) and fakes (right) background categories for 2018 data-taking period.

introducing significant deviations to the $\phi_{CP}$ distribution in the embedded samples (Fig. 5.17).

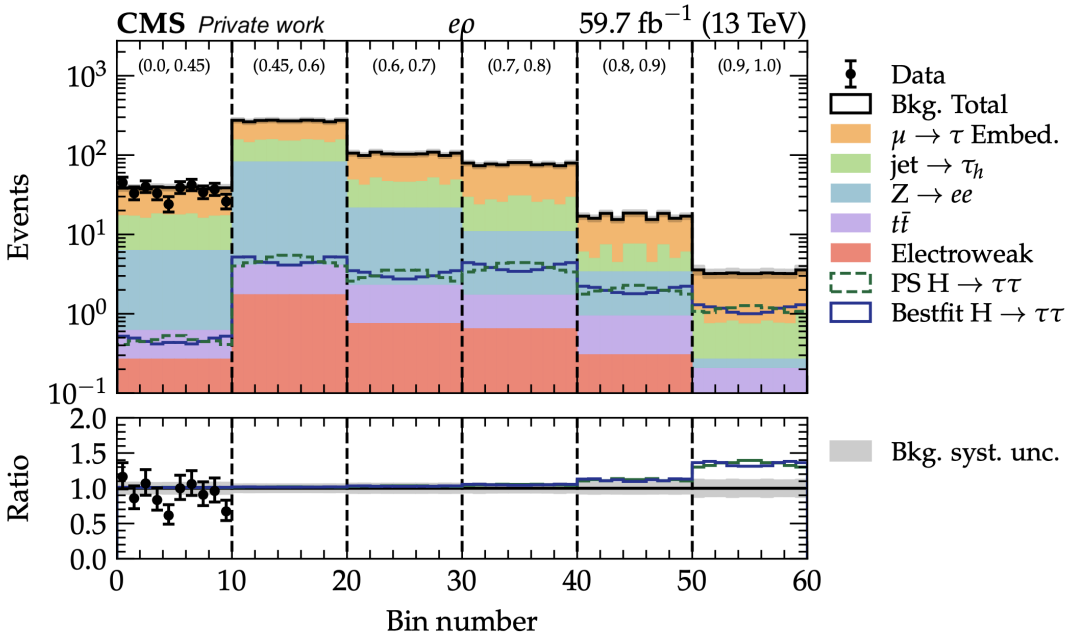


Figure 5.19: Pre-fit blinded distribution of unrolled in bins of the NN score $\phi_{CP}$ observable in $e\rho$ signal category for 2018 data-taking period.

Overall, good agreement between data and simulation is observed in background categories before performing the statistical fit (Fig. 5.18, Appendix A.8). For the blinded signal categories, the unrolled distribution of the $\phi_{CP}$ observable shows increasing signal over background ratio from the first to the last bins of NN score, with a clear separation between the pure CP-even and CP-odd hypotheses in the most sensitive $e\rho$ category (Fig. 5.19 and Appendix A.8).

## 5.7 Statistical inference

### 5.7.1 Framework

The strategy to extract the value of the main parameter of interest (POI) $\alpha^{H\tau\tau}$ follows the likelihood formalism [192–194]. The likelihood function is parametrised by several POIs and nuisance parameters as follows:

$$L(\alpha^{H\tau\tau}, \vec{\mu}, \vec{\theta}) = \prod_{j}^{N_{categories}} \prod_{i}^{N_{bins}} P\left(n_{ij} | \mathscr{L} \cdot \vec{\mu} \cdot \vec{A}_{ij}(\vec{\theta}, \alpha^{H\tau\tau}) + B_{ij}(\vec{\theta})\right) \prod_{m}^{N_{nuisance}} C_m(\theta_m). \quad (5.21)$$

It is parametrised by:

- $\alpha^{H\tau\tau}$: the mixing angle between SM and anomalous couplings as defined in Eq. 5.3.
- $\vec{\mu} \equiv (\mu_{ggH}, \mu_{qqH})$: a vector of the Higgs boson signal strength modifiers which are defined as the ratio of the corresponding cross section times the H $\rightarrow \tau\tau$ branching ratio with respect to the SM value. The ggH and qqH processes are considered in the statistical inference, where the latter scales the combined VBF and VH production modes.
- $\vec{\theta}$: a vector of nuisance parameters corresponding to the systematic uncertainties (Sec. 5.7.3).

The likelihood function is computed as a product over categories $j$ and bins $i$. In the analysis of the $\tau_e\tau_h$ final state the former product is taken over the categories for each of the three (2016, 2017, 2018) data-taking periods as defined by the corresponding neural networks (Sec. 5.6). In the combination with the other channels (Sec. 6), the corresponding categories as defined by the analysis of these channels are additionally included in the product. The signal (Higgs) category is further split into four categories based on the MVA DM predictions (Sec. 5.3.3) for the $\tau_h$ candidate with the decay modes MVA DM = {0, 1, 2, 10} being considered. Therefore, the final set of categories in the $\tau_e\tau_h$ analysis is:

- $e\pi$ (signal),
- $e\rho$ (signal),
- $ea_1^{1pr}$ (signal),
- $ea_1^{3pr}$ (signal),
- Genuine $\tau$ (background),
- Fakes (background).

The bins in Eq. 5.21 corresponds to the bins in the unrolled 2D histogram ($\phi_{CP}$, NN score) for the signal categories and the bins in 1D histogram of NN score for the background categories. The unrolled histogram is constructed by firstly binning the NN score distribution and then plotting for the events in each of the bins the histograms of the $\phi_{CP}$ distribution. The following NN score bin edges are used for all the data-taking periods:

- Signal categories: [0, 0.45], [0.45, 0.6], [0.6, 0.7], [0.7, 0.8], [0.8, 0.9], [0.9, 1.0].
- Genuine $\tau$ category: [0, 0.5], [0.5, 0.6], [0.6, 0.7], [0.7, 1.0].

- Fakes category: [0, 0.6], [0.6, 0.7], [0.7, 0.8], [0.8, 0.9], [0.9, 1.0].

In the signal categories, for each of the NN bins as defined above, the $\phi_{CP}$ distribution has 10, 8, and 4 equally sized bins in the range $[0°, 360°]$ for $e\rho$, $e\pi$, and $ea_1^{1pr}/ea_1^{3pr}$ categories, respectively.

Counts in each of the bins are modelled in Eq. 5.21 with a Poisson distribution $P(n_{ij}|n_{ij}^{exp})$ where $n_{ij}$ is the observed number of events in data and expected counts $n_{ij}^{exp}$ are modelled as a sum of the signal $\mathscr{L} \cdot \vec{\mu} \cdot \vec{A}_{ij}(\vec{\theta}, \alpha^{H\tau\tau})$ and background $B_{ij}(\vec{\theta})$ contributions. Here $\mathscr{L}$ is the integrated luminosity, $\vec{A}_{ij}(\vec{\theta}, \alpha^{H\tau\tau})$) is a vector of signal acceptances for each of the H production modes (ggH and qqH). $\vec{A}_{ij}$ and $B_{ij}(\vec{\theta})$ are produced in a form of templates as described in Sec. 5.7.2. Constraints on the systematic uncertainties are incorporated as prior probability density functions $C_m(\theta_m)$. For uncertainties altering only the normalisation of the counts with the same rate across all the bins (referred to as normalisation uncertainties) these are taken as log-normal distribution. For the uncertainties producing asymmetric count variation across the bins and therefore altering the shape of the templates (referred to as shape uncertainties) these are implemented in the likelihood minimisation as the continuous morphing with a Gaussian prior probability density function. Parameters of the probability density functions are further described in Sec. 5.7.3.

The statistical inference is performed using the Combine statistical toolkit [195]. The main parameter of interest (POI) is the CP mixing angle $\alpha^{H\tau\tau}$. In order to extract its value from data, a test statistic is constructed as a log-likelihood ratio:

$$q(\alpha^{H\tau\tau}) \equiv -2 \ln \left( \frac{L(\alpha^{H\tau\tau}, \hat{\hat{\vec{\mu}}}, \hat{\hat{\vec{\theta}}})}{L(\hat{\alpha}^{H\tau\tau}, \hat{\hat{\vec{\mu}}}, \hat{\hat{\vec{\theta}}})} \right), \tag{5.22}$$

where the denominator is the best fit value of the likelihood function with respect to all the POIs ($\alpha^{H\tau\tau}$, $\vec{\mu}$, and $\vec{\theta}$), and the numerator corresponds the likelihood function where all the POIs but the main one are profiled. The value $\hat{\alpha}^{H\tau\tau}$ which corresponds to the minimum of $q(\alpha^{H\tau\tau})$ is quoted as the best-fit value with the 68.3, 95.5, and 99.7% confidence intervals obtained using asymptotic approximation as the values of the mixing angle $\alpha^{H\tau\tau}$ where $q(\alpha^{H\tau\tau})$ equals to 1.00, 4.02, and 8.81 [196].

### 5.7.2 Template composition

As it was previously mentioned in Sec. 5.3.7, one expects certain symmetries to be preserved in the $\phi_{CP}$ distribution for the background processes: for example, the fact that genuine tau backgrounds are uniformly distributed at the generator level. It was also mentioned that smearing

and resolution effects come into play when one moves from the generator to the reconstructed level. The distributions are therefore distorted and the original symmetries are no longer applicable. However, some symmetries still remain, and they can be exploited as described below.

The motivation to impose symmetries comes from the observation that the statistical fluctuations for the simulated samples in the last bins of the unrolled $\phi_{CP}$ distribution are sizeable (Fig. 5.23). Therefore, in order to constrain the associated statistical bin-by-bin uncertainties it is beneficial to correct the signal and background templates and associated statistical uncertainties to have the expected symmetry properties. This also removes potential bias on the statistical inference which might appear due to the statistical fluctuation in the template bins.

For the background templates the following modifications are applied depending on the background process and the method used to reconstruct the $\phi_{CP}$ observable (Fig. 5.20):

- Genuine $\tau_h$ (IP+NP): flattening.
- Genuine $\tau_h$ (IP+IP): symmetrisation around $\phi_{CP} = 180°$.
- jet $\rightarrow \tau_h$ fakes (IP+NP, IP+IP): symmetrisation around $\phi_{CP} = 180°$.
- $l \rightarrow \tau_h$ fakes (IP+NP): flattening.
- $l \rightarrow \tau_h$ fakes (IP+IP): symmetrisation around $\phi_{CP} = 180°$.

In general, the flattenning is performed by setting the value of all the $\phi_{CP}$ bins in a single NN score bin to their average and introducing a single statistical uncertainty as a fully-correlated quadratic sum of uncertainties of the original bins. The symmetrisation is performed by setting the value of each of the two bins which are symmetric with respect to 180° to their average value. Each pair is attributed a single common nuisance parameter as in the flattening case. The effective number of associated nuisance parameters is thus reduced by 1/2 ($1/N_{bins}$) for the symmetrisation (flattening) procedures.

As it can be seen, the templates are kept uniform only for the background processes involving genuine $\tau$ leptons and in the signal categories where the neutral-pion method is used for the $\tau_h$ side ($e\rho, ea_1^{1pr}, ea_1^{3pr}$). The usage of the impact parameter vector for both of the prongs is sensitive to the smearing effects affecting the primary vertex reconstruction. These effects introduce a correlated behaviour in the $\phi_{CP}$ reconstruction where the $\phi_{CP}$ values of 0° and 360° are more favoured. However, the symmetry around the $\phi_{CP}$ value of 180° still holds and can be used. The same applies also to the jet $\rightarrow \tau_h$ fake events due to the kinematic properties of these events.

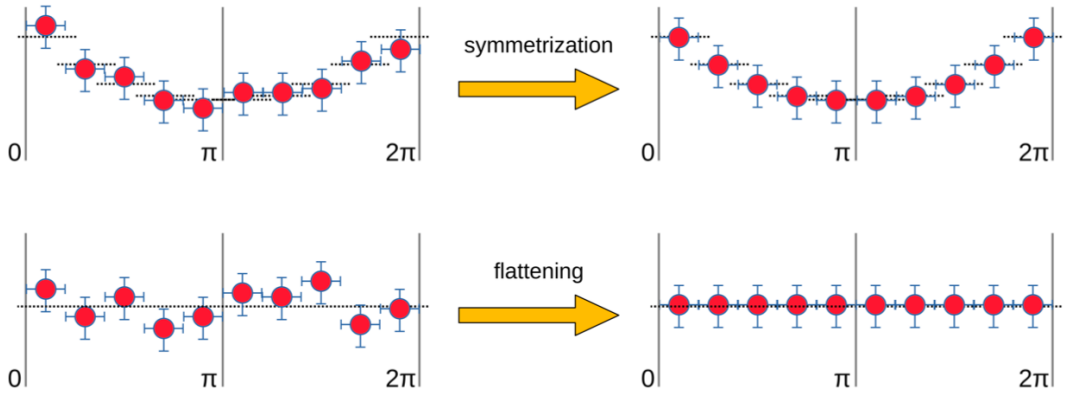


Figure 5.20: Illustration of the symmetrisation (top) and flattening (bottom) procedures applied to the signal and background templates [197]. Red circles represent the bin content in the histogram of the $\phi_{CP}$ observable, which originally is distributed randomly around the expected values (dashed lines). After the symmetrisation/flattening, the bin content is adjusted to the value estimated from the averaging across the symmetric bins as described in the text.

As it was mentioned in Sec. 5.2, the Higgs signal samples are generated according to the three discrete values of $\alpha^{H\tau\tau} = \{0°, 45°, 90°\}$, which corresponds to the CP-even, CP-odd, and CP-mix scenarios, respectively. Since in the statistical inference procedure a continuous range of $\alpha^{H\tau\tau}$ values is assumed, the general template $T(\alpha^{H\tau\tau})$ for any auxiliary value of the mixing angle is constructed separately for ggH and qqH production from the basis three templates ($T_{\text{even}}, T_{\text{odd}}, T_{\text{mix}}$) as follows:

$$T(\alpha^{H\tau\tau}) = (\cos^2\alpha^{H\tau\tau} - \cos\alpha^{H\tau\tau}\sin\alpha^{H\tau\tau})\cdot T_{\text{even}} + (\sin^2\alpha^{H\tau\tau} - \cos\alpha^{H\tau\tau}\sin\alpha^{H\tau\tau})\cdot T_{\text{odd}} + \\ + 2\cos\alpha^{H\tau\tau}\sin\alpha^{H\tau\tau}\cdot T_{\text{mix}}.$$

The signal templates for the $\phi_{CP}$ observable can also be affected by statistical fluctuations in the highest NN score bins. Therefore, similarly to the background templates, $T_{\text{even}}$ and $T_{\text{odd}}$ are symmetrised around $\phi_{CP} = 180°$. The symmetrisation of the $T_{\text{mix}}$ template is performed by the generation of an additional signal sample (separately for ggH and qqH templates) corresponding to $\alpha^{H\tau\tau} = -45°$ and its averaging with the $T_{\text{mix}}$ template shifted by 180°.

### 5.7.3 Systematic uncertainties

As it was mentioned earlier, there are two distinct types of systematic uncertainties: normalisation and shape uncertainties.

Normalisation uncertainties shift the normalisation of the templates without affecting the shape. Their parameters are constrained by adding a log-normal prior with the mean parameter corresponding to the nominal case of no systematic variation. The sigma of the distribution depends on the source of uncertainty and is provided below as a percentage of the nominal value.

Shape uncertainties are modelled with a continuous morphing procedure with a Gaussian prior on a morphing parameter. The parameter interpolates between two discrete up/down variations of the template corresponding to $\pm 1\sigma$ variations of the Gaussian prior. The mean of the prior distribution is set to 0, corresponding to the nominal template shape. Magnitudes of shape variations are provided below as a percentage of the systematic source variation resulting in $1\sigma$ up/down template variations.

Since in the statistical inference procedure categories for the three data-taking periods are analysed jointly, some sources of systematic uncertainties are (partially) correlated across the years, resulting in shared nuisance parameters in the fit. These cases are marked in Table 5.1, which also summarises all the sources of uncertainties incorporated into the statistical inference procedure together with the samples they are applied to. Lastly, uncertainties related to electron/$\tau_h$ identification and energy scale are treated as 50% correlated between the simulated and embedded samples. All the other common uncertainties are taken to be uncorrelated.

**Normalisation**

The following sources of normalisation uncertainties are considered in the analysis:

- Electron reconstruction (tracking/ID/isolation) efficiency: 2%.
- Electron trigger efficiency: 2%.
- $\tau_h$ ID (against $e, \mu$): 3%.
- b-tagging scale factors: 1-9%.
- Integrated luminosity: 2.5, 2.3, and 2.5% for 2016, 2017, and 2018 respectively [198–200].
- Embedded yield: 4%.
- Cross section uncertainties:
  - W+jets: 4%,
  - Drell-Yan: 2%,
  - Diboson: 5% [201],
  - Single top: 5% [202],

  - $t\bar{t}$: 4.2%,
  - H: 2-5% [18],
  - H $\rightarrow \tau\tau$ branching fraction: 2% [18].

- $e \rightarrow \tau_h$ misidentification rate: up to 10%, decay mode dependent.
- Impact parameter significance: the $S_{IP}$ correction is varied by ±25% (±40%) for a single pion (electron) and the variation is converted into a normalisation uncertainty in the range of 1-5%.

**Shape**

The following sources of shape uncertainties are considered in the analysis:

- $\tau_h$ reconstruction & ID: up to 3%, $p_T$ and decay mode dependent.
- $\tau_h$ trigger: $p_T$/decay-mode dependent.
- $\tau_h$ energy scale: 0.8-1.1 (0.2-0.5)% for simulated (embedded) samples, $p_T$/decay-mode dependent.
- Electron energy scale: $< 1$%, $p_T$ and $\eta$ dependent.
- $e \rightarrow \tau_h$ energy scale: 0.5-6.5%.
- Jet energy scale: event-by-event depending on the jet topology and kinematics. The uncertainties are also propagated to $\vec{p}_T^{\,miss}$ and observables which are dependent on $\vec{p}_T^{\,miss}$ for the simulated samples where no recoil corrections is applied (single top quark, $t\bar{t}$, and diboson production).
- Jet energy resolution: event-by-event depending on the jet topology and kinematics. The uncertainties are also propagated to $\vec{p}_T^{\,miss}$ and observables which are dependent on $\vec{p}_T^{\,miss}$ for the simulated samples where no recoil corrections is applied (single top quark, $t\bar{t}$, and diboson production).
- $\vec{p}_T^{\,miss}$ unclustered scale: event-dependent, used for the samples where recoil corrections are not applied (single top quark, $t\bar{t}$, and diboson production). The uncertainties are also propagated to $\vec{p}_T^{\,miss}$ and observables which are dependent on $\vec{p}_T^{\,miss}$.
- $\vec{p}_T^{\,miss}$ recoil corrections: event-dependent, applied for the Z+jets, W+jets and signal samples. The uncertainties are also propagated to $\vec{p}_T^{\,miss}$ and observables which are dependent on $\vec{p}_T^{\,miss}$.

Table 5.1: Summary of systematic uncertainties included into the statistical inference as described in Sec. 5.7.3. The first column describes the source of uncertainty. The second column describes the magnitude of the systematic variations and its dependency on observables. The third column describes the samples to which the uncertainty is applied (where "MC" corresponds to simulated samples). The fourth column describes if the uncertainty is correlated across the data-taking periods. The fifth column describes the type of uncertainty, where *ln*N corresponds to normalisation uncertainty.

| Uncertainty | Magnitude | Samples | Correlation | Type |
|---|---|---|---|---|
| Electron reconstruction | 2% | MC | Yes | ln N |
| Electron trigger | 2% | MC | No | ln N |
| $\tau_h$ ID (against $e, \mu$) | 3% | MC, embedded | No | ln N |
| b-tagging scale factors | 1-9% | $t\bar{t}$, single top | No | ln N |
| Integrated luminosity | 2.3-2.5% | MC | Partial | ln N |
| Embedded yield | 4% | Emb. | No | ln N |
| W+jets cross section | 4% | W+jets MC | Yes | ln N |
| DY cross section | 2% | DY MC | Yes | ln N |
| Diboson cross section | 5% | Diboson MC | Yes | ln N |
| Single top cross section | 5% | Single top MC | Yes | ln N |
| $t\bar{t}$ cross section | 4.2% | $t\bar{t}$ MC | Yes | ln N |
| H cross sections | 2-5% | Signal MC | Yes | ln N |
| H $\rightarrow \tau\tau$ branching fraction | 2% | Signal MC | Yes | ln N |
| $e \rightarrow \tau_h$ rate | 10% | MC with $e \rightarrow \tau_h$ | No | ln N |
| $S_{IP}(e, \pi)$ | 1-5% | MC | No | ln N |
| $\tau_h$ reconstruction | $p_T$/DM dep. | MC, embedded | Partial | Shape |
| $\tau_h$ trigger | $p_T$/DM dep. | MC | No | Shape |
| $\tau_h$ energy scale | $p_T$/DM dep. | MC, embedded | No | Shape |
| Electron energy scale | $p_T/\eta$ dep. | MC, embedded | No | Shape |
| $e \rightarrow \tau_h$ energy scale | 0.5-6.5% | MC with $e \rightarrow \tau_h$ | No | Shape |
| Jet energy scale | Event-dep. | MC | Partial | Shape |
| Jet energy resolution | Event-dep. | MC | No | Shape |
| $\vec{p}_T^{\,miss}$ unclustered scale | Event-dep. | ST, $t\bar{t}$, diboson MC | No | Shape |
| $\vec{p}_T^{\,miss}$ recoil corrections | Event-dep. | Z/W+jets, signal MC | No | Shape |
| $t\bar{t}$/diboson in embedded | 10% | embedded | Yes | Shape |
| Top quark $p_T$ reweighting | top $p_T$ dep. | ST, $t\bar{t}$ | Yes | Shape |
| Z mass and $p_T$ reweighting | Z $p_T$/mass dep. | DY MC | Partial | Shape |
| $F_F$ | Described in text | jet $\rightarrow \tau_h$ fakes | Partial | Shape |
| Prefiring | Event-dep. | MC | Yes | Shape |
| Theoretical uncertainties | Event-dep. | Signal MC | Yes | Shape |

- $t\bar{t}$/diboson in the embedded samples: 10% of $t\bar{t}$ and diboson contribution as estimated from the simulation is added/subtracted in the embedded templates. This is aimed to reduce a potential bias introduced by the embedding procedure to the selection of genuine tau lepton pairs originating from these processes.

- Top quark $p_T$ reweighting: $p_T(t)$ dependent, defined with up (down) variation corresponding to twice (no) correction size.

- Z mass and $p_T$ reweighting: $p_T(Z), m(Z)$ dependent, defined with up/down variation corresponding to ±10% of the correction size.

- $F_F$: uncertainties associated with $\vec{p}_T^{\,miss}$/electron $p_T$ non-closure corrections and extrapolation to same-sign/high-$m_T$ regions corrections are applied for each of the fake factors (W+jets, QCD, $t\bar{t}$). For $t\bar{t}$ $F_F$ (derived using simulated samples) an uncertainty is added to account for the differences between data and simulation. Additional uncertainty is assigned due to the subtraction of background processes without jet $\rightarrow \tau_h$ fakes.

- Prefiring: 0-4%, dependent on the process and category.

- Theoretical uncertainties: event-dependent, applied to the signal samples these include renormalisation and factorisation scales and parton showering uncertainties.

## 5.8 Results

After the likelihood minimisation is performed (Sec. 5.7), one can firstly investigate if there is no significant discrepancies between data and simulation in the post-fit distributions in the background categories. Overall, good description of data with the fitted templates is observed for all the categories and data-taking periods (Fig. 5.21 and Appendix A.9).

In a more formalised way, one can perform a goodness-of-fit (GoF) test to estimate if there is a statistically significant difference between data and fitted templates. Results of the saturated model GoF test [203] performed for the combination of all the categories (signal and background) and all the data-taking years (2016, 2017, 2018) show the $p$-value of 0.22, which also indicates a good quality of the fit.

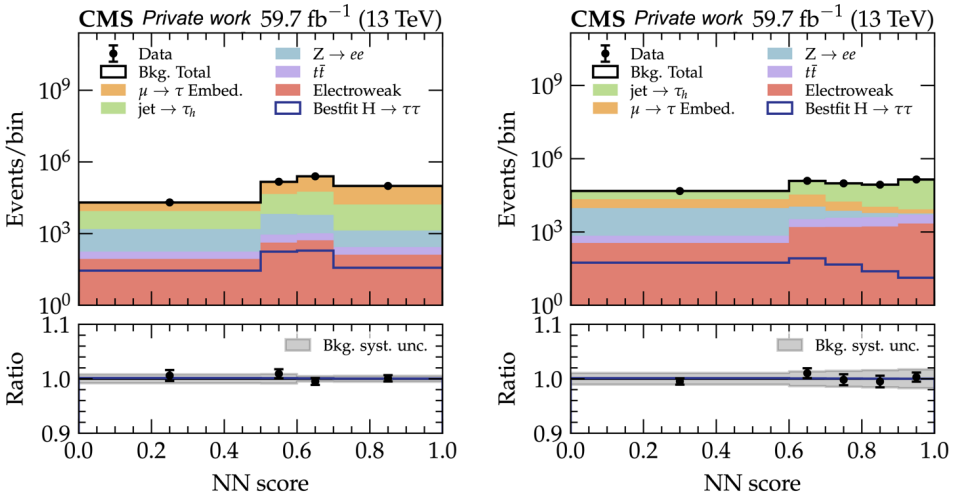


Figure 5.21: Post-fit distribution of the NN score in the genuine $\tau$ (left) and fakes (right) background categories for 2018 data-taking period.

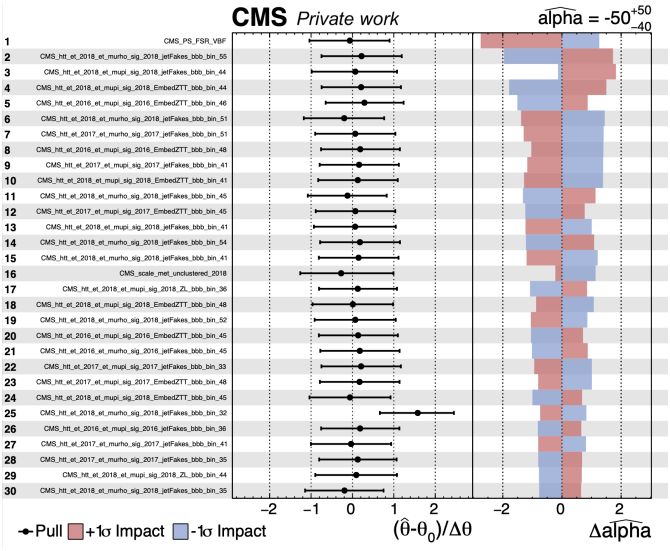


Figure 5.22: Summary of the post-fit analysis of the 30 leading nuisance parameters (left panel). In the middle panel a value of $(\hat{\theta} - \theta_0)/\Delta\theta$ is shown for each nuisance parameter, where $\hat{\theta}$ is a post-fit value, $\theta_0$ is a pre-fit nominal value, and $\Delta\theta$ is a nominal variance. The error bars correspond to the 68.3% ($1\sigma$) confidence level as obtained from the profiled likelihood scan. In the right panel, an impact distribution is shown, where each nuisance parameter is varied by $\pm 1\sigma$ and the corresponding variation of the main POI ($\alpha^{H\tau\tau}$) from its best fit value is shown as a red/blue bar.

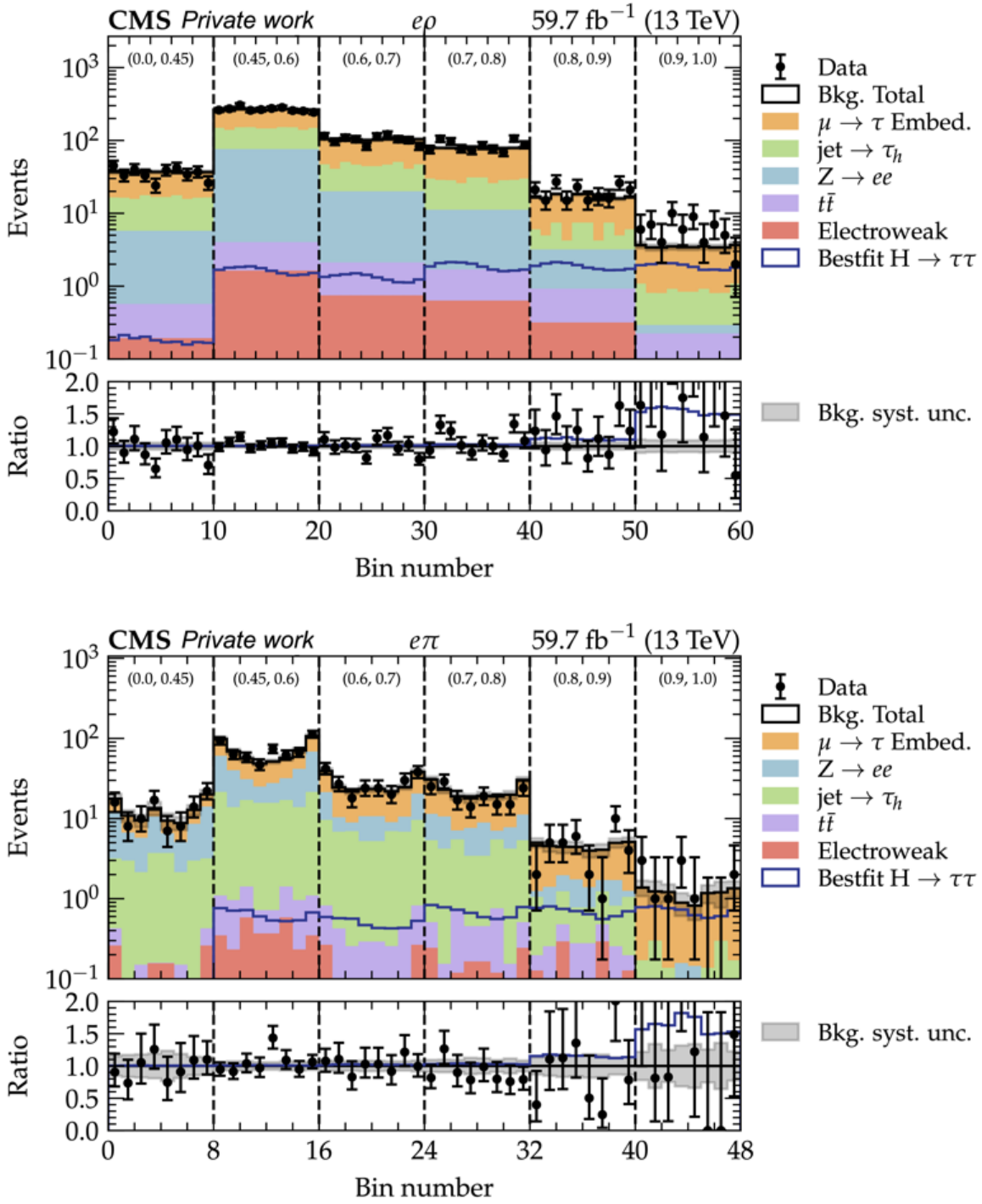


Figure 5.23: Post-fit distribution of unrolled in bins of the NN score $\phi_{CP}$ observable in the two most sensitive $e\rho$ (top) and $e\pi$ (bottom) signal categories for 2018 data-taking period.

In order to gauge the behaviour of the systematic uncertainties in the fit, a scan of the likelihood function Eq. 5.21 is performed for each of the nuisance parameter with all the POIs except for the nuisance parameter being profiled. Results are shown on Fig. 5.22 for the first 30 leading nuisance parameters as well as the impact of the each nuisance parameter variation on the main POI $\alpha^{H\tau\tau}$. Overall, no anomalous behaviour is observed in the nuisance parameter diagnostics.

The signal categories are then unblinded and the resulting unrolled distributions of $\phi_{CP}$ observable are shown in Fig. 5.23 and Appendix A.9. A slight presence of the H $\rightarrow \tau\tau$ signal is visible in the bins of the NN score, albeit diluted by statistical uncertainties.

Before proceeding to the extraction of the observed $\alpha^{H\tau\tau}$ value from the fit to data it is good to have an estimate of what one would expect under the null hypothesis, which is the Standard Model. The expected values of $\alpha^{H\tau\tau}$ are obtained with the same procedure as described in Sec. 5.7 but with the template fit being performed to the Azimov data set. The latter is obtained with fixing the cross sections of all the physical processes to their SM values and the $\phi_{CP}$ distribution to the prediction of the pure CP-even hypothesis.

The obtained expected value of the CP mixing angle is $\alpha^{H\tau\tau}_{exp} = 0^\circ \pm 90^\circ$ (Fig. 5.24, left), which corresponds to the expected CP sensitivity of $0.99\sigma$. The latter describes the statistical significance to exclude the pure CP-odd hypothesis if taken as a null hypothesis. Expected sensitivities split by the final states are:

- $e\tau_h$ (total): $0.99\sigma$,
- $e\rho$: $0.57\sigma$,
- $e\pi$: $0.54\sigma$,
- $ea_1^{3pr}$: $0.38\sigma$,
- $ea_1^{1pr}$: $0.17\sigma$.

One can further introduce $\mu$ as an inclusive signal strength modifier which scales the cross section times the H $\rightarrow \tau\tau$ branching fraction of all the three production modes altogether (opposite to the two separate $\mu_{ggH}$ and $\mu_{qqH}$ used to obtained the final result). The likelihood scan on the Azimov data set gives its expected value $\mu_{exp} = 1.00^{+0.26}_{-0.24}$. Overall, the expected values of both CP sensitivity and the signal strength show that the $\tau_e\tau_h$ channel is not sufficient on its own to provide significant information about the CP structure of the H$\tau\tau$ interaction. However, as it is shown in Sec. 6 a combination with the other $\tau_\mu\tau_h$ and $\tau_h\tau_h$ leads to conclusive results.

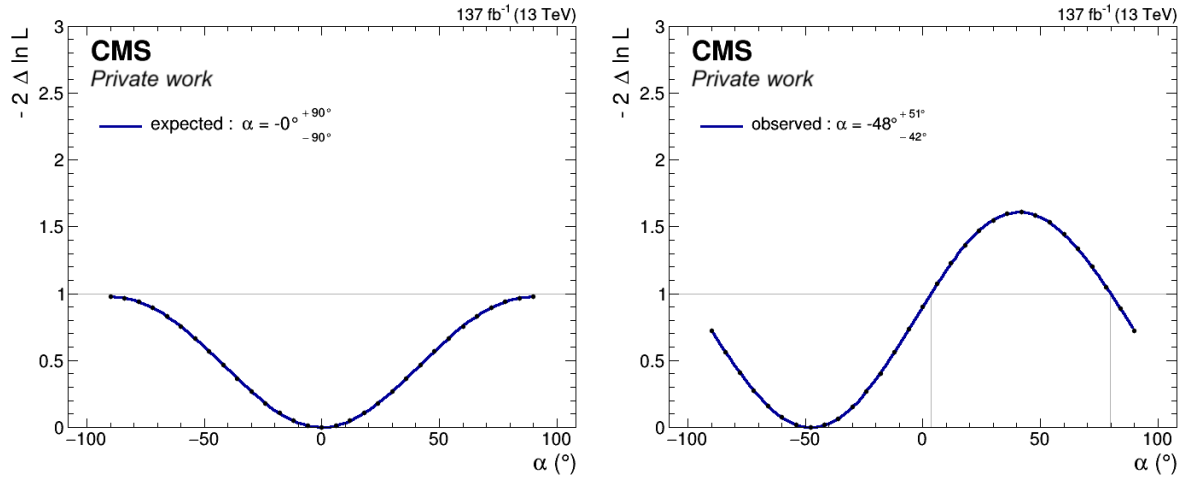


Figure 5.24: Left: the profiled likelihood scan for the $\alpha^{H\tau\tau}$ parameter of interest on the Azimov data set. Right: the profiled likelihood scan for the $\alpha^{H\tau\tau}$ parameter of interest on the observed data set.

Substituting the Azimov data set with the observed data, one can obtain with the same statistical inference procedure the observed values of the CP mixing angle $\alpha^{H\tau\tau}_{obs} = -48^{+51^\circ}_{-42^\circ}$ (Fig. 5.24, right). The observed value of the inclusive signal strength modifier is $\mu_{obs} = 1.14^{+0.27}_{-0.25}$, which is in agreement with the expectation.

To summarise, the $\tau_e\tau_h$ final state is *a priori* expected to provide subleading contribution to the analysis in terms of the CP sensitivity. This is due to the challenges in the electron reconstruction where the bremsstrahlung significantly impacts the resolution of the impact parameter vector, which in turn decreases the separation between CP-even and CP-odd hypothesis with the $\phi_{CP}$ observable. Moreover, a larger number of jets misidentified as electrons leads to higher $p_T$ thresholds at the trigger level, which further reduces the electron selection efficiency. Quantitatively, the observed (expected) value of the CP mixing angle is obtained to be $\alpha^{H\tau\tau} = -48^{+51^\circ}_{-42^\circ}(0^\circ \pm 90^\circ)$ which does not hint to any preferable CP hypothesis with the main limiting factor being the lack of statistics in this final state. Nevertheless, the expected CP sensitivity in the $\tau_e\tau_h$ channel is comparable with the expected contribution of the most sensitive $\mu\rho, \rho\rho, \rho\pi$ final states ($1.16\sigma$, $1.10\sigma$, and $1.08\sigma$) respectively. Therefore, it plays an important role in the final combination of all the final states considered in the CP analysis of the $H\tau\tau$ coupling, as described in Sec. 6.

# 6 | Combination of $\tau_e\tau_h$, $\tau_\mu\tau_h$, and $\tau_h\tau_h$ channels

In Sec.5 an overview of the analysis of the Yukawa coupling between the Higgs boson and tau leptons in the $\tau_e\tau_h$ channel has been presented. However, the first channels to be analysed and released as preliminary were $\tau_\mu\tau_h$ and $\tau_h\tau_h$ [197,204]. These two channels are expected to be the most sensitive to the CP measurement due to the larger branching fraction and/or more precise reconstruction of the final state and therefore of the $\phi_{CP}$ observable.

In addition, in the fully-hadronic $\tau_h\tau_h$ channel and specifically in the $a_1^{3pr}a_1^{3pr}$ final state, one can fully reconstruct the Higgs rest frame, which is not straight-away possible in the other final states (hence an approximation with the zero-momentum frame is used). This makes it a unique playground also to probe the *polarimetric vector method* to reconstruct the $\phi_{CP}$ observable.

Originally, the $a_1^{3pr}a_1^{3pr}$ final state was analysed with the NP+NP method for the $\phi_{CP}$ observable reconstruction. However, the final publication [186] introduces the polarimetric vector method to the analysis and adds $\tau_e\tau_h$ channel, as described in this work, to the combination with the $\tau_\mu\tau_h$ and $\tau_h\tau_h$ channels. This chapter therefore provides a brief overview of the analysis in the $\tau_\mu\tau_h$ (Sec. 6.1) and $\tau_h\tau_h$ (Sec. 6.2) channels. The results of the overall combination are presented in Sec. 6.3.

## 6.1 Overview of $\tau_\mu\tau_h$ channel

Generally, there is little difference between the analysis of $\tau_\mu\tau_h$ and $\tau_e\tau_h$ channels in terms of strategy. It proceeds with the selection of pairs of muon (Sec. 5.3.2) and $\tau_h$ (Sec. 5.3.3) physics objects. The requirements for the $\tau_h$ candidates are the same as in the $\tau_e\tau_h$ channel except for the DeepTau working points, which are required to be Medium against jets, VVLoose against electrons, and Tight against muons. One more difference with respect to the $\tau_e\tau_h$ channel is related to the trigger (single muon or cross-trigger) and the corresponding online $p_T$ requirements

(in brackets in GeV):

- 2016: $\mu(22)$ OR $\mu(19)$ & $\tau_h(20)$.
- 2017/2018: $\mu(24)$ OR $\mu(20)$ & $\tau_h(27)$.

The offline $p_T$ requirements are also chosen to be 1(5) GeV higher compared to the online $p_T$ thresholds for the muon ($\tau_h$) legs. All the other selection requirements are identical to those described in Sec. 5.4 when replacing the electron with the muon.

The main background sources in the $\tau_\mu\tau_h$ are genuine tau lepton pairs, jets faking $\tau_h$ and less pronounced contribution from prompt/non-prompt leptons faking $\tau_h$ compared to the $\tau_e\tau_h$ channel. The former background is modelled with the embedding technique (Sec. 5.5.1) where the simulated tau leptons are forced to decay into the $\mu\tau_h$ final state with a 100% branching fraction. The jet $\rightarrow \tau_h$ background is modelled with the $F_F$ method (Sec. 5.5.2) following the same procedure as in the $\tau_e\tau_h$ channel including the same corrections for non-closure tests and the difference between DR and SR/AR. All the other minor backgrounds ($\mu \rightarrow \tau_h$, jet $\rightarrow \tau_\mu$) are taken from the simulation. Corrections to simulated/embedded samples which differ from those described in Sec. 5.5.3 are related to the muon leg and include muon tracking/ID/isolation/trigger scale factors, $\mu \rightarrow \tau_h$ fake rate and energy scale corrections as derived with the tag-and-probe method. A dedicated calibration of the muon impact parameter is derived with $Z \rightarrow \mu\mu$ events in data.

Event categorisation is performed with a set of neural networks following the strategy outlined in Sec. 5.6. For each data-taking year two neural networks are trained in a two-fold manner. The set of input features (with the change of electron features to those of a muon), NN hyperparameters, training data set composition, and the training setup are identical to those used in the analysis of the $\tau_e\tau_h$ channel.

Three output categories (signal, genuine $\tau$, fakes) are defined and the signal category is further split depending on the $\tau_h$ MVA DM identification with the $\tau_h \rightarrow \pi/\rho/a_1^{1pr}/a_1^{3pr}$ decay modes being considered. A requirement on the impact parameter significance of the muon and pion $S_{IP}(\mu/\pi) > 1.5$ is applied in the signal categories to the $\tau$ candidates which decay into a single muon or into a single charged pion. In the background categories, the requirement is applied only to the latter case. The $\phi_{CP}$ observable is used to extract the CP mixing angle $\alpha^{H\tau\tau}$ as outlined in Sec. 5.3.7 with the impact parameter (IP) and neutral pion (NP) methods being used depending on the $\tau_h$ decay mode. To summarise, the following categories and $\phi_{CP}$ reconstruction methods on the side of the $\tau_\mu\tau_h$ channel are used in the final combination:

- $\mu\pi$ (signal) $\longrightarrow$ IP + IP,
- $\mu\rho$ (signal) $\longrightarrow$ IP + NP,

- $\mu a_1^{1pr}$ (signal) $\longrightarrow$ IP + NP,
- $\mu a_1^{3pr}$ (signal) $\longrightarrow$ IP + NP,
- Genuine $\tau$ (background),
- Fakes (background).

## 6.2 Overview of $\tau_h\tau_h$ channel

While there is little difference between the analysis of $\tau_e\tau_h$ and $\tau_\mu\tau_h$ channels due to the similarity of the physical processes, the $\tau_h\tau_h$ channel has some specific properties slightly altering the analysis strategy.

Events are triggered with a ditau trigger for all three data-taking years. The online $p_T$ thresholds for the candidates is 35 GeV and therefore the requirement $p_T(\tau_h) > 40$ GeV is applied offline as well as $|\eta(\tau_h)| < 2.1$. A pair of $\tau_h$ candidates matched to the trigger objects with the opposite charge and cone distance $\Delta R > 0.5$ is selected. Both candidates are required to pass the DeepTau Medium working point against jets, VVLoose working point against electrons, and VLoose working point against muons. No requirement on the transverse mass $m_T$ is applied since the W+jets background is not sizeable in this case. For the $a_1^{3pr}a_1^{3pr}$ final state, described in more details below, a presence of a reconstructed secondary vertex is required for each of the $\tau_h$ candidates.

The fact that the $\tau_h\tau_h$ final state is fully-hadronic, opposite to the $\tau_l\tau_h$ where at least one light lepton is required, results in the background composition dominated by the QCD jets faking $\tau_h$. This affects the $F_F$ derivation which now consists only in the fake factor responsible for the QCD process. It is derived as in the $\tau_e\tau_h$ case with the determination region defined by inverting the opposite-sign requirement for the tau lepton pair. The fake factor is measured for the leading $\tau_h$ candidate as a function of its $p_T$ in bins of the number of jets and MVA decay mode, with the case of MVA DM = 0 split into bins of $S_{IP}$ ($S_{IP} < 1.5$, $S_{IP} \geq 1.5$). The application region, where $F_F$ is applied as a weight on event-by-event basis, is defined from inverting the nominal DeepTau working points in the signal region for the leading $\tau_h$ candidate. Corrections are derived to account for non-closure differences in the $\vec{p}_T^{miss}$ variable and for the difference between the opposite-sign and same-sign regions.

Background processes involving two genuine $\tau_h$ are modelled with the embedded samples. It is also possible that the leading $\tau_h$ candidates is a genuine tau and the subleading $\tau_h$ candidate is lepton/jet fake. These events are modelled with the simulation, as well as the ones where the leading $\tau_h$ candidate is a lepton faking $\tau_h$. All the corrections described in Sec. 5.5.3 which are related to the $\tau_h$ objects are applied.

The $\tau_h\tau_h$ candidates are also divided into the signal, genuine $\tau$, and fakes categories. The former inclusively captures the ggH, VBF, and VH production modes. The genuine $\tau$ category is aimed at the background processes involving two genuine tau leptons and is represented in the training data set by events from embedded samples. The fake categories captures events where at least one $\tau_h$ candidates is a misidentified jet or lepton. In the training data set this category is represented by events taken from simulated samples (application region weighted with $F_F$) for the cases where a lepton (jet) fakes $\tau_h$. The categorisation is performed with a BDT trained in a two-fold manner separately for each data-taking year. The input features describe the kinematics of the two $\tau_h$ candidates, jets, MET, and the ditau system.

The signal category is further split into categories depending on the MVA DM classification of both the $\tau_h$ candidates. The final states classified as $\rho a_1^{1pr}$ and $a_1^{1pr}a_1^{1pr}$ are merged together because the MVA DM classification is not able to disentangle efficiently decay modes with one and two neutral pions. A requirement on the impact parameter significance of the pion $S_{IP}(\pi) > 1.5$ is applied in the signal and background categories to the $\tau_h$ candidates which decay into a single charged pion. The $\phi_{CP}$ observable is reconstructed depending on the decay mode of each $\tau_h$ candidate with the methods described in Sec. 5.3.7 with the only exception of the $a_1^{3pr}a_1^{3pr}$ final state described below. Overall, the following nine signal and two background categories with the corresponding decay plane reconstruction methods are used in the statistical inference procedure:

- $\pi\pi$ (signal) $\longrightarrow$ IP + IP,
- $\pi\rho$, $\pi a_1^{1pr}$, $\pi a_1^{3pr}$ (signal) $\longrightarrow$ IP + NP,
- $\rho\rho$, $\rho a_1^{1pr}$, $\rho a_1^{3pr}$, $a_1^{1pr}a_1^{3pr}$ (signal) $\longrightarrow$ NP + NP,
- $a_1^{3pr}a_1^{3pr}$ (signal) $\longrightarrow$ Polarimetric vector,
- Genuine $\tau$ (background),
- Fakes (background).

As it was mentioned in introduction to this Chapter, the $a_1^{3pr}a_1^{3pr}$ final state where both the $a_1$ mesons decay into three charged pions and one tau neutrino provides a unique opportunity of reconstructing the Higgs decay frame. It starts from reconstructing for each $\tau_h$ candidate its decay vertex by refitting together three charged tracks resulting from the tau decay. Given this secondary vertex (SV) one can obtain the direction of the tau lepton momentum in the laboratory frame as the vector connecting the refitted primary vertex (Sec. 5.3.5) with the secondary one.

The Gottfried-Jackson angle $\theta_{GJ}$ is defined as the angle between the tau lepton (as a vectorial distance between PV and SV) and the $a_1$ momentum (as a vectorial sum of its charged prong

momenta) vectors in the laboratory frame. Then one can obtain a magnitude of the tau lepton momentum as [205]:

$$|\vec{p}_\tau| = \frac{(m_{a_1}^2 + m_\tau^2)|\vec{p}_{a_1}|\cos\theta_{GJ} \pm \sqrt{(m_{a_1}^2 + |\vec{p}_{a_1}|^2)((m_{a_1}^2 - m_\tau^2)^2 - 4m_\tau^2|\vec{p}_{a_1}|^2\sin^2\theta_{GJ})}}{2(m_{a_1}^2 + |\vec{p}_{a_1}|^2\sin^2\theta_{GJ})}. \quad (6.1)$$

Here, $m_{a_1}$ is the reconstructed invariant mass of the intermediate $a_1$ meson, $m_\tau$ is the mass of the tau lepton, $\vec{p}_{a_1}$ is the momentum vector of the $a_1$ meson reconstructed from its decay products. The maximum possible value of the Gottfried-Jackson angle is:

$$\theta_{GJ}^{max} = \arcsin\left(\frac{m_\tau^2 - m_{a_1}^2}{2m_\tau|\vec{p}_{a_1}|}\right). \quad (6.2)$$

In cases where the reconstructed $\theta_{GJ}$ exceeds $\theta_{GJ}^{max}$, its value is set to $\theta_{GJ}^{max}$. From Eq. 6.1 one can observe that there are two possible solutions to the equation. This ambiguity results from two possible ways the $a_1$ meson can decay in the tau rest frame: opposite to and in the direction of the tau lepton momentum. The ambiguity for the $a_1^{3pr}a_1^{3pr}$ pair is resolved by selecting one of the four solutions which results in the invariant mass closest to the Higgs boson mass. Eq. 6.1 together with the tau lepton direction defined by the displacement of SV from PV reconstructs the full four-momentum of each of the tau lepton candidates in the laboratory frame. By combining them together one obtains the four-momentum of the H in the laboratory frame and therefore can perform a Lorentz boost into its rest frame.

The full reconstruction of the $a_1$ and tau momentum also enables the computation of the polarimetric vector $h_\mu$. It is a four-vector which enters the most general parametrisation of $\tau \to X + \nu_\tau$ decay width [206, 207]:

$$d\Gamma = \frac{|M|^2}{2m_\tau}(1 + h_\mu s^\mu)dLips, \quad (6.3)$$

where $|M|^2$ is the spin averaged squared matrix element of the reaction, $dLips$ is the Lorentz invariant phase space element, $s^\mu$ is the tau spin four-vector in the tau rest frame. For the decay into a single charged pion and a neutrino $\tau \to \pi + \nu_\tau$ the spacial part of the polarimetric vector takes its simplest form $\vec{h} = -\vec{n}_\pi$, i.e. equals to the negative unit vector of the pion momentum in the tau rest frame. However, for the three prong tau decays the expression cannot be derived analytically and quantitative methods are used within the TAUOLA library [208–210].

Using Eq. 6.3 one can derive that the $\phi_{CP}$ angle can be computed as the angle between the planes spanned for each of the tau leptons by the polarimetric and tau momentum vectors with Eq. 5.12 for the partial decay width holding true. Therefore, for the $a_1^{3pr}a_1^{3pr}$ final state one uses

polarimetric vectors $\vec{h}^{\pm}$ for each of the tau leptons together with their momentum vectors $\vec{q}^{\pm}$ in the laboratory frame to construct unit vectors $\hat{\lambda}^{\pm}$ (Sec. 5.3.7) as:

$$\hat{\lambda}^{\pm} = \frac{\vec{h}^{\pm} \times \vec{q}^{\pm}}{|\vec{h}^{\pm} \times \vec{q}^{\pm}|}. \tag{6.4}$$

By making a Lorentz-boost of $\hat{\lambda}^{\pm}$ into the Higgs rest frame, derived as described above, one obtains $\hat{\lambda}^{HMF^{\pm}}$ and uses them to define the $\phi^{HMF}$ and $O^{HMF}$ variables:

$$\phi^{HMF} = \arccos\left(\hat{\lambda}^{HMF^{+}} \cdot \hat{\lambda}^{HMF^{-}}\right), \tag{6.5}$$

$$O^{HMF} = -(\hat{\lambda}^{HMF^{+}} \times \hat{\lambda}^{HMF^{-}}) \cdot \hat{q}^{HMF^{+}}. \tag{6.6}$$

Finally, the $\phi_{CP}$ angle is constructed as in Eq. 5.15:

$$\phi_{CP} = \begin{cases} \phi^{HMF} & \text{if } O^{HMF} \geq 0 \\ 360^\circ - \phi^{HMF} & \text{if } O^{HMF} < 0 \end{cases} \tag{6.7}$$

In theory, the polarimetric vector method for the $\phi_{CP}$ observable computation can be used in all tau decay modes provided that the tau momentum can be reconstructed. This is however not possible straight-away because of neutrinos escaping detection and carrying away a fraction of momentum. The $a_1^{3pr}a_1^{3pr}$ final state provides a unique opportunity to infer the momentum of both tau leptons from the information about the tau decay vertex and further uses it to derive polarimetric vectors to construct the $\phi_{CP}$ observable. It was studied that the polarimetric vector method improves the expected CP sensitivity in the $a_1^{3pr}a_1^{3pr}$ final states by approximately 80% from $0.14\sigma$ to $0.25\sigma$. This improved sensitivity hints towards further studies in the direction of using the method in the other final states.

## 6.3 Results

The statistical inference for the combination of the $\tau_e\tau_h$, $\tau_\mu\tau_h$, and $\tau_h\tau_h$ channels is performed following the procedure described in Sec. 5.7. The likelihood function is constructed with Eq. 5.21 where the product over categories is extended to include the categories described in Sec. 6.1 and 6.2 for the $\tau_\mu\tau_h$ and $\tau_h\tau_h$ channels respectively with the corresponding nuisance parameters. Most of the systematic uncertainties are common across all the three channels and therefore are formalised with common nuisance parameters in the fit. Uncertainties which are

Table 6.1: Summary of systematic uncertainties additionally included in the combination of $\tau_e\tau_h$, $\tau_\mu\tau_h$, and $\tau_h\tau_h$ channels with respect to ones described in Table 5.1.

| Uncertainty | Magnitude | Samples | Correlation | Type |
|---|---|---|---|---|
| Muon reconstruction ($\tau_\mu\tau_h$) | 1% | MC | Yes | ln N |
| SV reconstruction eff. ($a_1^{3pr}a_1^{3pr}$) | 2% | MC | No | ln N |
| Muon trigger ($\tau_\mu\tau_h$) | 2% | MC | No | ln N |
| Tau trigger ($\tau_h\tau_h$) | $p_T$/Decay-mode dep. | MC | No | Shape |
| $e \rightarrow \tau_h$ fake rate ($\tau_h\tau_h$) | 2-10% | MC | No | ln N |
| $\mu \rightarrow \tau_h$ fake rate ($\tau_\mu\tau_h$) | up to 40% | MC | No | ln N |
| Muon energy scale ($\tau_\mu\tau_h$) | 0.4-2.7% | MC | Yes | Shape |
| $\mu \rightarrow \tau_h$ energy scale ($\tau_\mu\tau_h$) | 1% | MC | No | Shape |

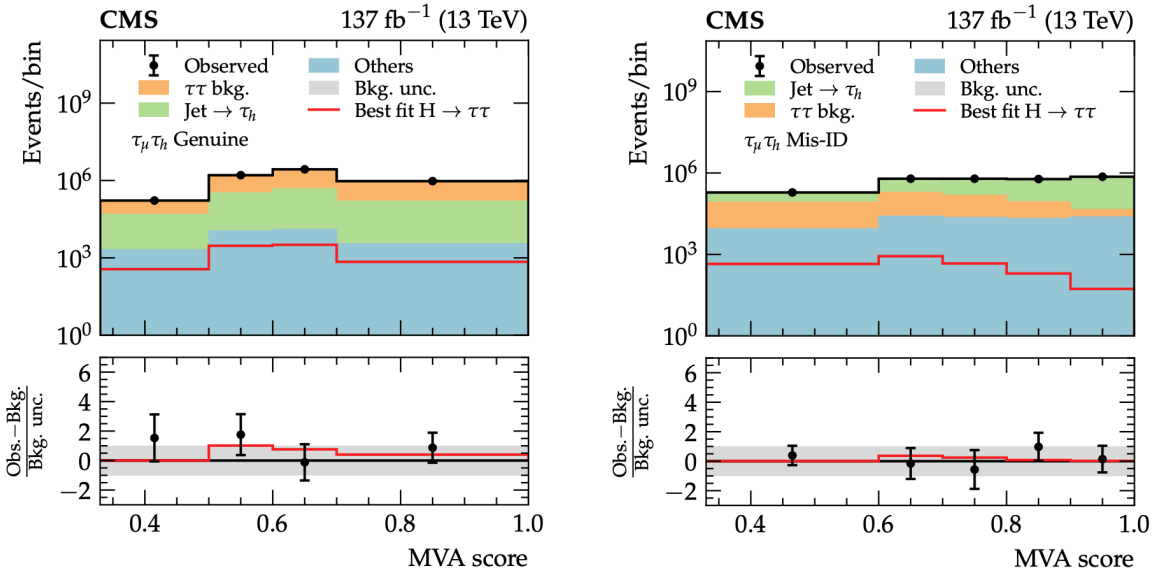


Figure 6.1: Post-fit distribution of the NN score in the genuine $\tau$ (left) and fakes (right) background categories combined for all the data-taking periods in the $\tau_\mu\tau_h$ channel.

additionally included to those described in Sec. 5.7.3 are related to the trigger selection and corrections applied specifically in $\tau_\mu\tau_h$ and $\tau_h\tau_h$ channels (Table 6.1).

After the likelihood function is maximised, a GoF test with the saturated model is performed. The obtained $p$-value of 0.51 combined for all the categories and all the data-taking years indicates a good agreement of the fitted statistical model with data. This is also verified by the post-fit data distributions in the background categories for $\tau_\mu\tau_h$ (Fig. 6.1) and $\tau_h\tau_h$ (Fig. 6.2) channels. The unrolled distributions of the $\phi_{CP}$ observable in bins of the BDT/NN score in the signal categories for the most sensitive $\mu\rho$ (Fig. 6.3, top) and $\rho\rho$ (Fig. 6.3, bottom) final states

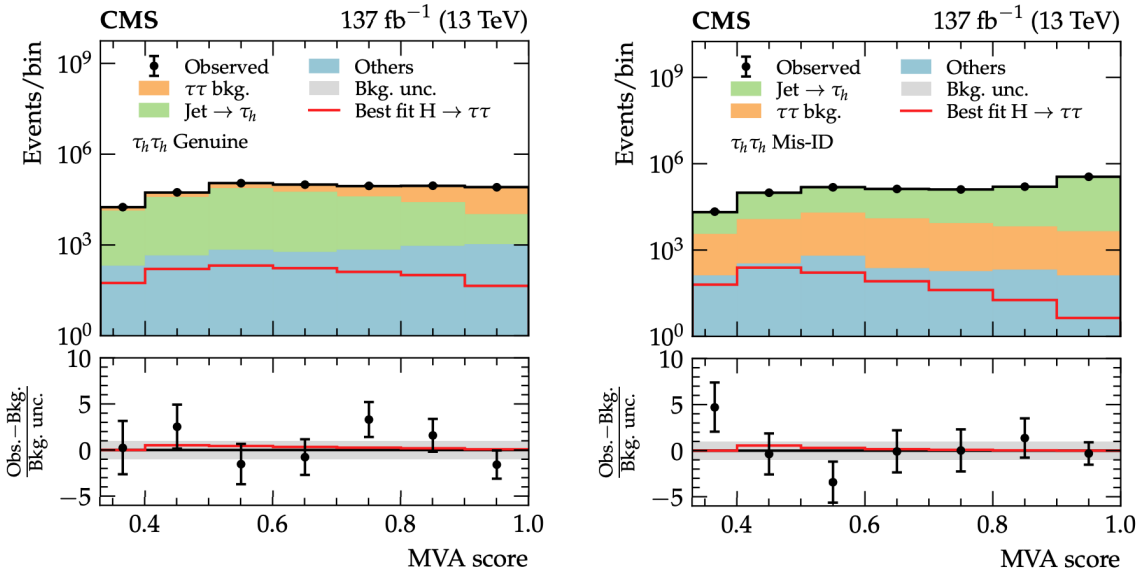


Figure 6.2: Post-fit distribution of the NN score in the genuine $\tau$ (left) and fakes (right) background categories combined for all the data-taking periods in the $\tau_h\tau_h$ channel.

also show good level of agreement between data and simulation. The presence of the H signal is visible in the last bins of the NN score as well as the modulations in the distribution, however still dominated by the statistical fluctuations.

Results of the statistical fit to the Asimov data set yield the following expected sensitivity to reject the pure CP-odd hypothesis under the CP-even null hypothesis:

- $\tau_\mu\tau_h + \tau_e\tau_h + \tau_h\tau_h$: $2.6\sigma$.
- $\tau_\mu\tau_h$: $1.47\sigma$.
  - $\mu\rho$: $1.16\sigma$,
  - $\mu\pi$: $0.71\sigma$,
  - $\mu a_1^{3pr}$: $0.51\sigma$,
  - $\mu a_1^{1pr}$: $0.24\sigma$.
- $\tau_e\tau_h$: $0.99\sigma$.
  - $e\rho$: $0.57\sigma$,
  - $e\pi$: $0.54\sigma$,
  - $e a_1^{3pr}$: $0.38\sigma$,

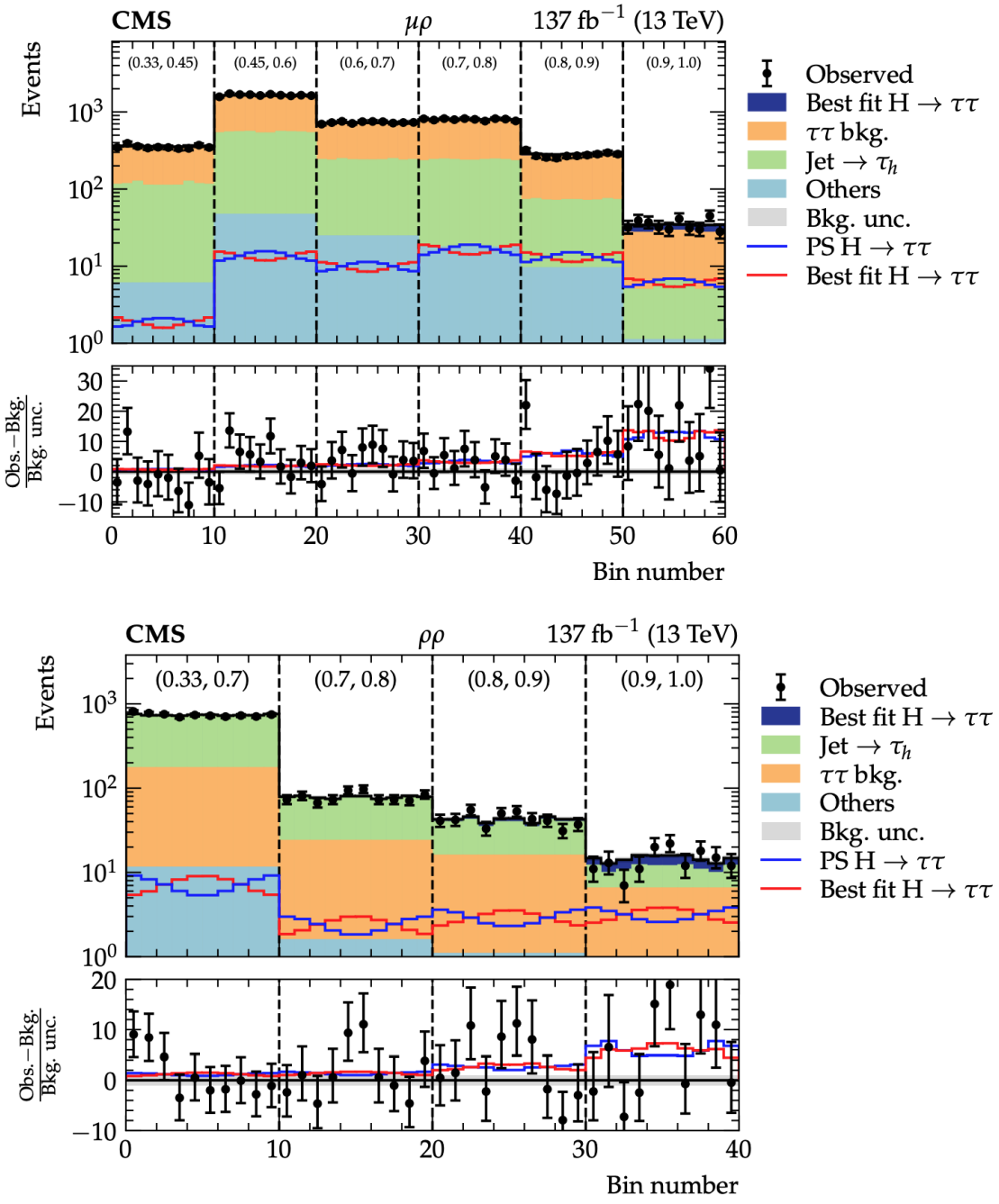


Figure 6.3: Post-fit distribution of unrolled in bins of the NN score $\phi_{CP}$ observable in the two most sensitive $\mu\rho$ (top) and $\rho\rho$ (bottom) signal categories combined for all data-taking periods.

– $ea_1^{1pr}$: $0.17\sigma$.

- $\tau_h\tau_h$: $1.85\sigma$.
  - $\rho\rho$: $1.10\sigma$,
  - $\rho\pi$: $1.08\sigma$,
  - $\rho a_1^{3pr}$: $0.65\sigma$,
  - $\pi\pi$: $0.39\sigma$,
  - $\pi a_1^{3pr}$: $0.48\sigma$,
  - $a_1^{1pr}\rho + a_1^{1pr}a_1^{1pr}$: $0.30\sigma$,
  - $\pi a_1^{1pr}$: $0.23\sigma$,
  - $a_1^{3pr}a_1^{3pr}$: $0.28\sigma$,
  - $a_1^{3pr}a_1^{1pr}$: $0.13\sigma$.

Out of the three channels the $\tau_h\tau_h$ channel provides the largest sensitivity, with the second and third contributions coming from the $\tau_\mu\tau_h$ and $\tau_e\tau_h$ channels respectively. The most significant final states are those including at least one tau lepton decaying into a $\rho$ meson and a tau neutrino. However, contributions from the single-prong final states (e.g. $\mu\pi$) are also sizeable. As it was mentioned earlier, despite the $\tau_e\tau_h$ channel provides a subleading contribution to the expected CP sensitivity comparing to the $\tau_\mu\tau_h$ and $\tau_h\tau_h$ channels, it is comparable to the most leading final states and therefore serves as an important addition to the overall combination. The relative contribution of the $a_1^{3pr}a_1^{3pr}$ final state is even smaller comparing to the other final states. However, almost twice the relative gain of the polarimetric vector method in expected sensitivity with respect to the neutral pion method is an important step further in developing analysis techniques for CP- and tau-related analyses.

One of the parameters of interest in the fit is the vector of signal strength modifiers $\vec{\mu} = (\mu_{ggH}, \mu_{qqH})$. The corresponding observed values after the fit $\mu_{ggH} = 0.59^{+0.28}_{-0.32}$ and $\mu_{qqH} = 1.39^{+0.56}_{-0.47}$ are strongly anticorrelated (correlation coefficient $\rho = -0.76$). This result is expected since the analysis does not differentiate between various H production modes by design, e.g. at the categorisation level where all the production modes are treated inclusively in the signal category. One can further introduce instead of $\vec{\mu}$ a single $\mu$ POI which would scale simultaneously contributions from ggH and VBF processes times the $H \to \tau\tau$ branching fraction. Observed (expected) values of this parameter are $0.82 \pm 0.15$ ($1.00 \pm 0.16$) which is compatible with the dedicated $H \to \tau\tau$ coupling analysis [211].

The likelihood scan of the main parameter of interest $\alpha^{H\tau\tau}$ for the final combination on both the Azimov and collected data sets is shown in Fig. 6.4. Observed (expected) value of the CP mixing angle is measured to be $-1 \pm 19^\circ$ ($0 \pm 21^\circ$) at 68.3% CL. This result is compatible

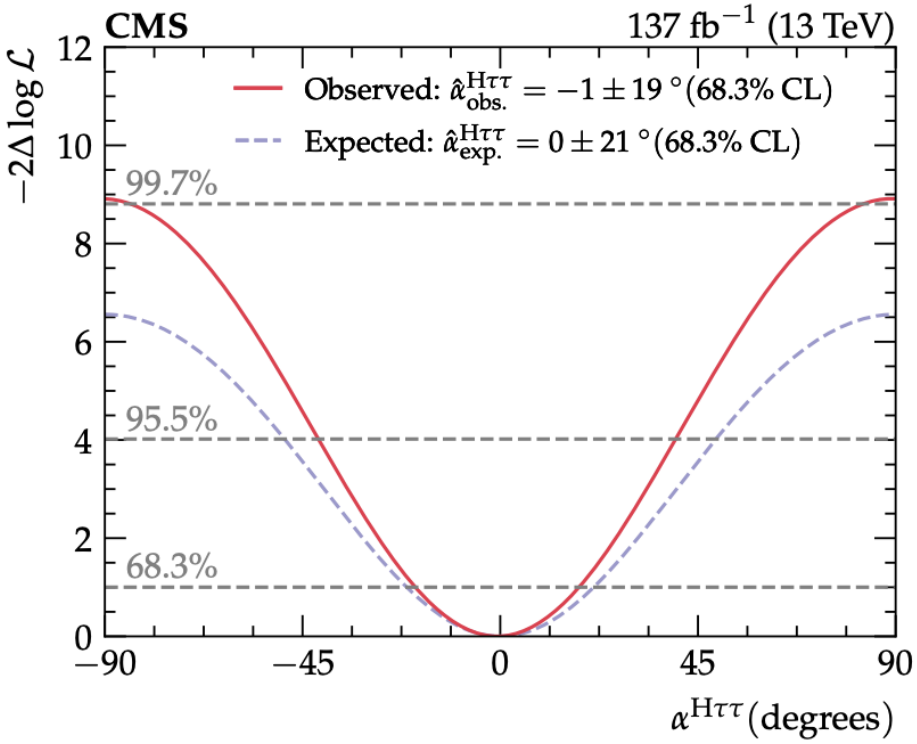


Figure 6.4: Negative log-likelihood scan of the CP mixing angle $\alpha^{H\tau\tau}$ for the combination of $\tau_\mu\tau_h$, $\tau_e\tau_h$, and $\tau_h\tau_h$ channels for all the data-taking years on the Azimov (dashed blue line) and collected (red solid line) data sets. Horizontal dashed grey lines indicate the thresholds for the 68.3%, 95.5%, and 99.7% CLs.

with the Standard Model expectation of the $H \rightarrow \tau\tau$ coupling to correspond to the pure CP-even hypothesis. It also corresponds to the observed (expected) rejection of the pure CP-odd hypothesis under assumption of the pure CP-even hypothesis at the level of $3.0\sigma$ ($2.6\sigma$).

The observed value of $\alpha^{H\tau\tau}$ with the uncertainty value decomposed into different sources (statistical uncertainties, experimental systematic uncertainties, statistical uncertainties in the signal and background templates, theoretical uncertainties) is:

$$\alpha^{H\tau\tau} = -1 \pm 19\ (\text{stat}) \pm 1\ (\text{syst}) \pm 2\ (\text{bin-by-bin}) \pm 1\ (\text{theo})^\circ. \tag{6.8}$$

The result is therefore largely dominated by the statistical uncertainties with the second leading uncertainty coming from the bin-by-bin uncertainties in the signal and background templates. The latter is directly linked to the former since background contribution is largely estimated from data itself.

One can also investigate if there is a correlation between the inclusive signal strength mod-

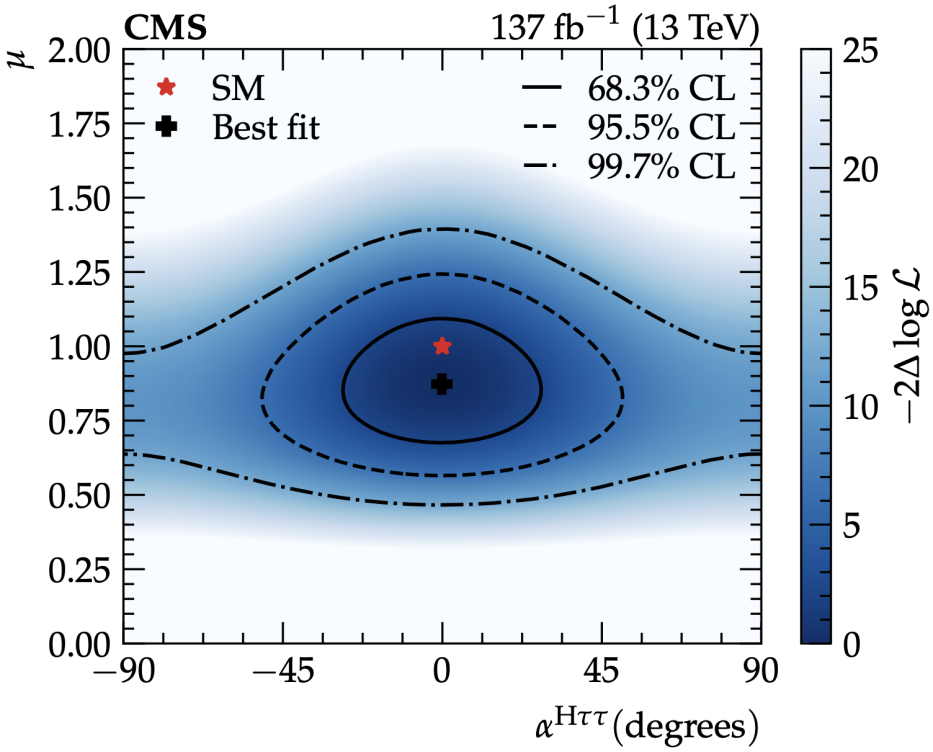


Figure 6.5: The 2D negative log-likelihood scan of the inclusive signal strength modifier $\mu$ versus the CP mixing angle $\alpha^{H\tau\tau}$.

ifier $\mu$, scaling the total H production cross section times H $\rightarrow \tau\tau$ branching fraction, and the CP mixing angle $\alpha^{H\tau\tau}$. The 2D negative log-likelihood scan is shown in Fig. 6.5 with the corresponding 68.3%, 95.5%, and 99.7% CL contours being derived as the values where $-2\Delta\log\mathscr{L}$ equals 2.30, 6.20, and 11.62, respectively. No significant correlation between the two parameters is observed.

An additional scan is performed for the likelihood function parametrised in terms of the couplings $\kappa_\tau$ and $\tilde{\kappa}_\tau$ (Eq. 5.1). All the other H couplings (both affecting the H production and decay into $\tau\tau$) are fixed to their SM values. The 2D scan is shown in Fig. 6.6, where one should note that the fit cannot distinguish the absolute sign of $\kappa_\tau$ and $\tilde{\kappa}_\tau$, which results in the symmetrical shape of the negative log-likelihood distribution.

Lastly, one can combine the most sensitive signal categories together into a single representative plot to visualise which CP hypothesis is overall more favoured by data. The corresponding distribution is shown in Fig. 6.7 for the $\rho\rho$, $\pi\rho$, $\mu\rho$, and $e\rho$ final states. It is obtained by the following procedure: firstly, the bins of the unrolled $\phi_{CP}$ distribution for all the signal categories corresponding to these final states are taken and the background contribution is sub-

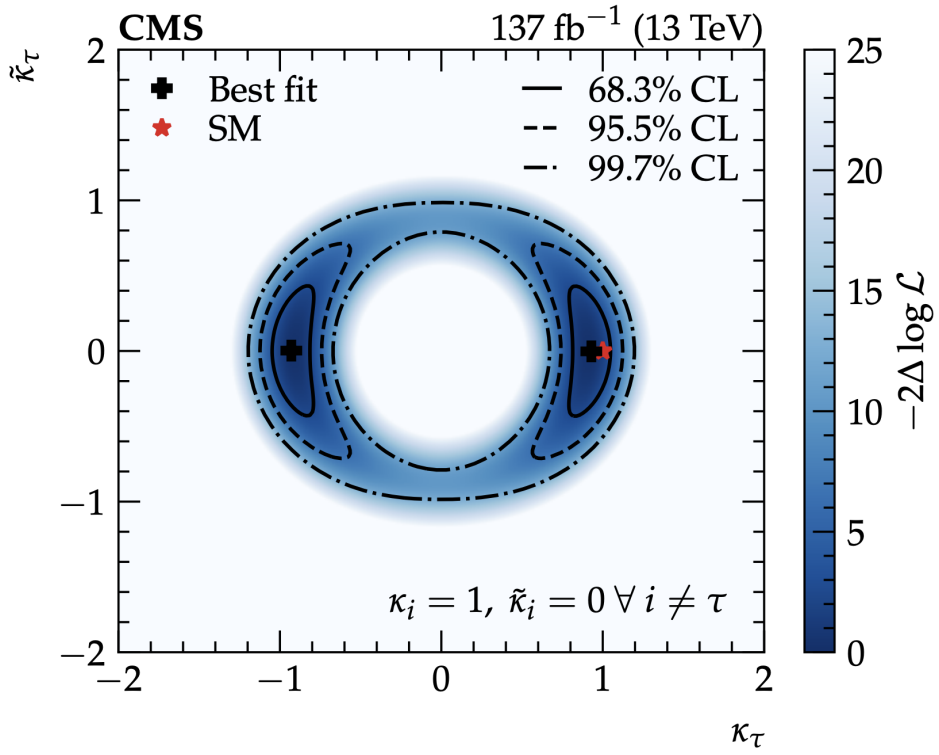


Figure 6.6: The 2D negative log-likelihood scan of the CP-even ($\kappa_\tau$) and CP-odd ($\tilde{\kappa}_\tau$) Yukawa couplings between the SM Higgs boson and tau leptons (Eq. 5.1).

tracted from data. Then, for each bin a weight is computed as $w_{\text{bin}} = A \cdot S/(S + B)$ where $A = |\text{CP}^{\text{even}} - \text{CP}^{\text{odd}}|/(\text{CP}^{\text{even}} + \text{CP}^{\text{odd}})$ is the separation between the pure CP-even and CP-odd contributions in this bin normalised to the total number of bins; $S$ and $B$ are expected signal and background contributions in the bin respectively. This weight is applied for each of the bins both to the background-subtracted data and CP-even/CP-odd signal templates. The weighted sum of contributions across all the signal categories per $\phi_{CP}$ bin is finally plotted with corresponding uncertainties. Additionally, since there is a phase shift of 180° between the $\tau_l\tau_h$ and $\tau_h\tau_h$ channels due to the different sign of spectral functions for the electron/muon, the corresponding shift is applied in the weighted combination of bins. Overall, one can observe that the data favours the pure CP-even hypothesis.

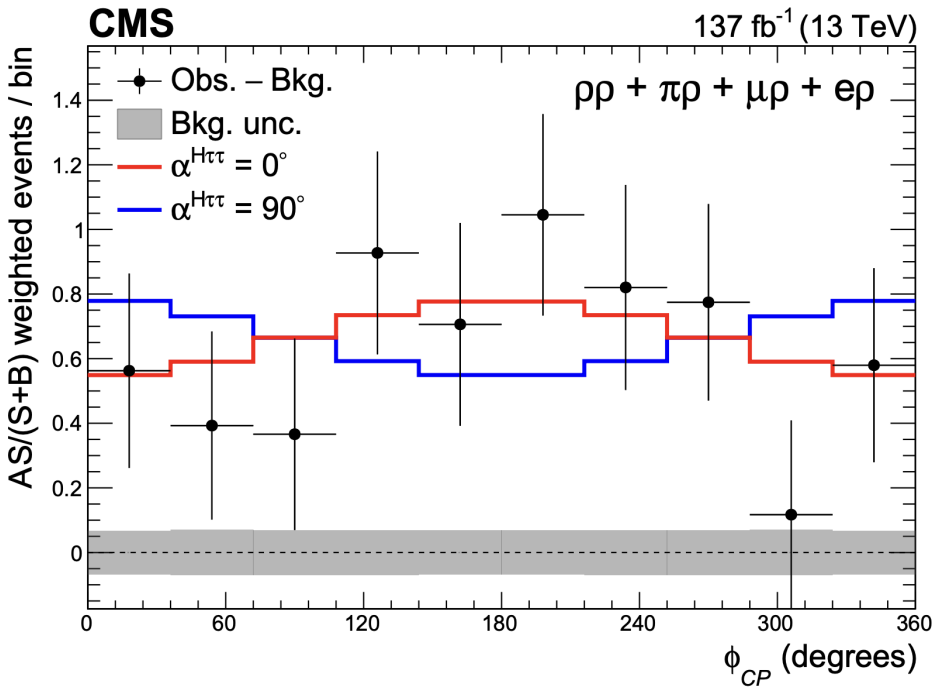


Figure 6.7: The $\phi_{CP}$ distribution for $\rho\rho$, $\pi\rho$, $\mu\rho$, and $e\rho$ final states weighted by $A \cdot S/(S+B)$ as described in Sec. 6.3. The red (blue) histogram corresponds to the distribution of the pure CP-even (CP-odd) hypothesis, black dots correspond to the background-subtracted data. The grey uncertainty band corresponds to the uncertainty of the subtracted background contribution.

# 7 | Summary

This work presents the measurement of the CP properties of the Yukawa coupling between the Higgs boson and tau leptons with the CMS experiment at the LHC in the proton-proton collisions at $\sqrt{s} = 13$ TeV. The data set corresponds to the integrated luminosity of 137 fb$^{-1}$ collected in 2016, 2017, and 2018 years. The analysis was performed in the $\tau_e\tau_h$ channel where one tau lepton decays into a single electron and the other hadronically via the following decay modes: $\pi^\pm$, $\rho \to \pi^\pm\pi^0$, $a_1^\pm \to \pi^\pm\pi^0\pi^0$, $a_1^\pm \to \pi^\pm\pi^\mp\pi^\pm$.

The structure of the Yukawa coupling between the Higgs boson and tau leptons was parametrised in terms of the effective mixing angle $\alpha^{H\tau\tau}$ between the pure CP-even and pure CP-odd contributions. The angle between the tau lepton decay planes in the zero-momentum frame $\phi_{CP}$ was used as an observable to experimentally extract the $\alpha^{H\tau\tau}$ value. For the $H \to \tau\tau$ decays the distribution for $\phi_{CP}$ represents modulations where $\alpha^{H\tau\tau}$ enters as a phase shift with the $\alpha^{H\tau\tau} = 0°$ corresponding to the SM scenario of the pure CP-even coupling. Dedicated methods were developed and optimised to experimentally reconstruct the $\phi_{CP}$ angle for each of the final states.

Considering the $\tau_e\tau_h$ channel only, the expected significance to reject the pure CP-odd hypothesis under the pure CP-even hypothesis was $0.99\sigma$ with $e\rho$ (expected significance $0.57\sigma$) and $e\pi$ (expected significance $0.54\sigma$) being the most sensitive final states. The observed (expected) value of the effective mixing angle was obtained to be:

$$\alpha^{H\tau\tau}_{\tau_e\tau_h} = -48^{+51°}_{-42°}(0 \pm 90°)\ @68.3\%\ \text{CL}. \tag{7.1}$$

This result cannot lead to conclusive statements *per se* since it is large limited by the available statistics and instrumental precision of reconstructing the $\phi_{CP}$ observable. The combination of the measurements in the $\tau_e\tau_h$ channel with those from the $\tau_\mu\tau_h$ and $\tau_h\tau_h$ channels was performed. The observed (expected) significance to reject the pure CP-even hypothesis was measured to be $3.0(2.6)\sigma$. If one considers only $\tau_\mu\tau_h$ and $\tau_h\tau_h$ channels, the corresponding value was measured to be $3.2(2.3)\sigma$.

The observed (expected) value of the effective mixing angle for the combination of the $\tau_e\tau_h$,

$\tau_\mu\tau_\mathrm{h}$, and $\tau_\mathrm{h}\tau_\mathrm{h}$ channels was obtained to be:

$$\alpha^{\mathrm{H}\tau\tau} = -1 \pm 19^\circ(0 \pm 21^\circ) \ @68.3\% \ \mathrm{CL}. \tag{7.2}$$

This value is compatible with the SM prediction as well as with the measurement by the ATLAS experiment: $\alpha^{\mathrm{H}\tau\tau} = 9 \pm 16^\circ(0 \pm 28^\circ)$ @68% CL [212]. The results presented in this work are published in the Journal of High Energy Physics [186].

Studies were performed to improve the tau lepton identification algorithms in CMS. The DeepTau model, extensively used by analysts during the Run 2 data-taking period, was retrained and optimised. This resulted in the consistently reduced misidentification rate against both electrons, muons, and jets by 10-50% at a given efficiency across the phase space regions of interest. The dedicated adversarial fine-tuning provided improved agreement between data and simulation in the distribution of the $D_\mathrm{jet}$ score in the region corresponding to the baseline $\mathrm{H} \rightarrow \tau\tau$ selection without significant decrease in the model performance. The resulting model will be used as a recommended algorithm for the tau lepton identification in CMS during the early Run 3 data-taking period.

Several set-based architectures were developed and adapted to the tau lepton identification task. This includes the ParticleNet and a newly proposed Tau Transformer (TaT) architectures. The training was performed in the unified framework where the tau lepton representation was modelled as the combination of various modalities and the architectures enter as the feature extracting modules. Ablation studies were performed to investigate the contribution of various modalities and the cone size of $\tau_\mathrm{h}$ candidates on the TaT model performance. Overall, compared to the retrained DeepTau baseline, the ParticleNet architecture achieves similarly or better performance on the $\tau_\mathrm{h}$ classification against jets and muons, while performing sizeably worse against electrons. The Tau Transformer architecture consistently outperforms the retrained DeepTau baseline across $p_\mathrm{T}$, $\eta$ and tau decay mode regions of interest providing a reduction in the misidentification rate at a given $\tau_\mathrm{h}$ efficiency by up to 50%.

The measurement of the CP properties of the Higgs boson presented in this work will largely profit from more data. With that respect, the ongoing Run 3 and the upcoming HL-LHC periods of data-taking will allow for a more precise measurement of the structure of Yukawa coupling in the $\mathrm{H}\tau\tau$ sector, which is currently sizeably limited by the statistical uncertainties. Furthermore, other scenarios of the coupling structure, e.g. non-Hermitian, can be investigated, as well as the combination with the measurements of the Higgs coupling with the top quark.

From the experimental side, this is also a great opportunity to design and advance analysis techniques to account for specific features of the CP measurement and the underlying physics. Such ideas as the full ditau system reconstruction and direct $\phi_{CP}$ angle regression are challenging and yet exciting to be explored. On the side of the tau lepton reconstruction, there is room for improvement to recover the inefficiencies of the HPS algorithm when it comes to the missing

neutrino and $\pi^0$ reconstruction. Furthermore, better encoding of inductive biases into the model can be studied, e.g. in the positional encoding or multimodality structure of the input representation. Lastly, the study of self-supervised approaches to the model training will potentially open a new chapter in the object tagging by leveraging enormous amount of data collected by the LHC experiments.

The Standard Model to date provides a very accurate description of various physics phenomena down to the particle scale, thus laying the very foundation of our understanding of nature. Despite this, several fundamental questions are not yet answered within the SM, which indicates that the knowledge puzzle is still missing some pieces. However, it seems reasonable to believe that human curiosity is stubborn enough to make it only a matter of time before the paradigm will broaden and a space for new questions will open up.

# A | Appendix

## A.1 ML glossary

**Activation function:** a non-linear function which is usually applied to the output of the linear transformation within a feed-forward layer to introduce non-linearity.

**Backpropagation/backward pass:** a process of computing the gradient of the loss function with respect to each model weight by the chain rule, computing the gradient one layer at a time, iterating backward from the last model layer to the first one.

**Batch:** a set of input objects from the training data set used for a single step of the model training.

**Categorical variable:** a variable that can take one of a limited and fixed number of possible values.

**Decoder:** a part of the model which maps the representation learned by the encoder into a target prediction.

**Embedding:** a procedure mapping a given representation of the input into another representation.

**Encoder:** a part of the model which maps the input representation (usually high-dimensional) of the event/object into a representation useful for the given optimisation task (usually low-dimensional).

**Epoch:** a period of time corresponding to the number of steps when the model completes the iteration over the entire training dataset.

**Forwardpropagation/forward pass:** a process of propagating a batch of input data through the model layer by layer to the final layer which outputs a prediction and the value of the loss function.

**Hyperparameter:** a parameter whose value is used to control the learning process or the model configuration. Therefore, it usually remains constant during the training rather than being optimised jointly with the model weights.

**Label/Target:** information about the event/object which the model learns to predict given an input representation of the event/object during the optimisation step.

**Loss function:** a function that maps the event or values of one or more variables onto a real number representing a "cost" associated with the event. Usually parametrised per event as a function of the true event information (target) and the corresponding model predictions, the loss function is minimised as a part of an optimisation problem in order to infer the model's parameters. The term also refers to a cost associated with a group of events, usually by summing/averaging the loss function values for each of the events.

**Model:** a mathematical representation of objects and their relationships to each other. Usually represents (a set) of functions or rules parametrised by some parameters, also called weights.

**Optimiser:** the method to update the model weights during the training. Usually it is done using some modified form of the gradient descent. In its original definition, at the given step of the training each of the weights is updated by the negative gradient of the loss function w.r.t. to the corresponding weight as computed for a given batch during the backpropagation step. The key parameter in this update is a so-called learning rate, which is a multiplier to the gradient vector. Many optimisers do not fix it as a constant throughout the training but rather try to dynamically adjust it as the training is ongoing.

**Receiver Operator Characteristic (ROC) curve:** a plot of the true positive rate (positive class efficiency) against the false positive rate (negative class efficiency) at various model threshold settings used to measure the model performance.

**Regularisation:** a reduction of model sensitivity as measured by the change in performance to a certain type of input data variation.

**Representation:** generally, the way a given event/object is described. In the context of ML, the representation is the way the object/event is numerically encoded (manually or by the model itself) in order to be used for the downstream task.

**Training:** the process of inferring the best parameters of the model as obtained by minimising the loss function. In most of the classification problem it is performed in steps where each step corresponds to a forward pass of a single batch to compute the value of the loss function, followed by a backward pass where the weight gradients are computed and the model parameters are updated by the optimiser.

**Up/Downsampling (image preprocessing):** the increase/decrease of the spatial resolution of the image while keeping the same representation.

## A.2 DeepTau loss function

The loss function minimised during the training of the DeepTau v2.1 and v2.5 models takes the form:

$$L(\boldsymbol{y}^{\text{true}}, \boldsymbol{y}^{\text{pred}}; \kappa, \boldsymbol{w}, \gamma) = \frac{1}{N_\tau} \sum_{i=1}^{N_\tau} w_i L_{\text{base}}(\boldsymbol{y}_i^{\text{true}}, \boldsymbol{y}_i^{\text{pred}}; \kappa, \gamma)$$

$$L_{\text{base}}(\boldsymbol{y}^{\text{true}}, \boldsymbol{y}^{\text{pred}}; \kappa, \gamma) = \underbrace{\kappa_\tau H_\tau(y_\tau^{\text{true}}, y_\tau^{\text{pred}})}_{\text{Categorical CE for } \tau \text{ vs. all background classes}} + \underbrace{(\kappa_e + \kappa_\mu + \kappa_{\text{jet}}) \bar{F}_\tau(1 - y_\tau^{\text{true}}, 1 - y_\tau^{\text{pred}}; \gamma_\tau)}_{\text{Focal loss for all background classes vs. } \tau} + \underbrace{\kappa_F \sum_{i \in \{e,\mu,\text{jet}\}} \kappa_i \Theta(y_\tau^{\text{pred}} - 0.1) \bar{F}_{\text{i}}(y_i^{\text{true}}, y_i^{\text{pred}}; \gamma_i)}_{\text{Focal loss for separate background classes with } y_\tau^{\text{pred}} > 0.1 \text{ vs. all the other classes}}$$

$$H_\tau(y_\tau^{\text{true}}, y_\tau^{\text{pred}})) = -y_\tau^{\text{true}} \log y_\tau^{\text{pred}}$$

$$\bar{F}(y^{\text{true}}, y^{\text{pred}}; \gamma)) = N \cdot F(y^{\text{true}}, y^{\text{pred}}; \gamma). \quad F(y^{\text{true}}, y^{\text{pred}}; \gamma)) = -y^{\text{true}}(1 - y^{\text{pred}})^\gamma \log\left(y^{\text{pred}}\right)$$

Here, $\Theta(.)$ is the Heaviside step function, $\bar{F}_i$ are the normalised focal loss terms, $N$ is the factor normalising the focal loss to unity in the range $[0, 1]$, $w_i$ are the individual training weights defined per $\tau_{\text{h}}$ candidate. The following constants are used for the training of the DeepTau v2.5 model (unless it is specified differently in the text):

- $\kappa_e = 0.4$, $\kappa_\mu = 1.$, $\kappa_\tau = 2.$, $\kappa_{\text{jet}} = 0.6$, $\kappa_F = 5.$.
- $N_e = N_\mu = N_{\text{jet}} = 1.63636$, $N_\tau = 1.17153$.
- $\gamma_e = \gamma_\mu = \gamma_{\text{jet}} = 2$, $\gamma_\tau = 0.5$.

## A.3 TaT: input features

Below, the input features are listed in the form of aliases. More detailed description and formal definition of each feature can be found in [213, 214].

- Global features (HPS/$\tau_h$ candidate related):
  - Particle properties: *particle_type (token identifier), tau_pt, tau_eta, tau_mass, tau_E_over_pt, tau_charge.*
  - Isolation variables: *tau_chargedIsoPtSum, tau_chargedIsoPtSumdR03_over_dR05, tau_footprintCorrection, tau_neutralIsoPtSum, tau_neutralIsoPtSumWeight_over_neutralIsoPtSum, tau_neutralIsoPtSumWeightdR03_over_neutralIsoPtSum, tau_neutralIsoPtSumdR03_over_dR05, tau_photonPtSumOutsideSignalCone, tau_puCorrPtSum.*
  - Secondary vertex features: *tau_hasSecondaryVertex, tau_sv_minus_pv_x, tau_sv_minus_pv_y, tau_sv_minus_pv_z, tau_flightLength_x, tau_flightLength_y, tau_flightLength_z, tau_flightLength_sig.*
  - IP features: *tau_dxy_valid, tau_dxy, tau_dxy_sig, tau_ip3d_valid, tau_ip3d, tau_ip3d_sig, tau_dz_sig_valid, tau_dz, tau_dz_sig.*
  - Misc.: *rho, tau_n_charged_prongs, tau_n_neutral_prongs, tau_pt_weighted_deta_strip, tau_pt_weighted_dphi_strip, tau_pt_weighted_dr_signal, tau_pt_weighted_dr_iso, tau_e_ratio_valid, tau_e_ratio, tau_gj_angle_diff_valid, tau_gj_angle_diff, tau_n_photons, tau_emFraction, tau_inside_ecal_crack, tau_leadingTrackNormChi2, tau_leadChargedCand_etaAtEcalEntrance_minus_tau_eta*
- PF candidates:
  - Positional information (in $\eta - \phi$ plane): *r, theta.*
  - Particle properties: *rel_pt, particle_type (token identifier), charge.*
  - PV and impact parameter information: *pvAssociationQuality, fromPV, vertex_dx, vertex_dy, vertex_dz, vertex_dx_tauFL, vertex_dy_tauFL, vertex_dz_tauFL, dxy, dxy_sig, dz, dz_sig.*
  - Tracker information: *lostInnerHits, nPixelHits, hasTrackDetails, nHits, nPixelLayers, nStripLayers, track_ndof, chi2_ndof.*
  - Calorimeter information: *hcalFraction, rawCaloFraction, rawHcalFraction.*
  - Misc.: *tauLeadChargedHadrCand, puppiWeight.*
- RECO electrons:

  - Positional information (in $\eta - \phi$ plane): *r, theta*.
  - Particle properties: *rel_pt, particle_type (token identifier).*
  - MVA variables: *mvaInput_earlyBrem, mvaInput_lateBrem, mvaInput_sigmaEtaEta, mvaInput_hadEnergy, mvaInput_deltaEta.*
  - Track properties: *rel_trackMomentumAtVtx, rel_trackMomentumAtCalo, rel_trackMomentumOut, rel_trackMomentumAtEleClus, rel_trackMomentumAtVtxWithConstraint, gsfTrack_normalizedChi2, gsfTrack_numberOfValidHits, rel_gsfTrack_pt, gsfTrack_pt_sig, has_closestCtfTrack, closestCtfTrack_normalizedChi2, closestCtfTrack_numberOfValidHits.*
  - Cluster variables: *cc_valid, cc_ele_rel_energy, cc_gamma_rel_energy, cc_n_gamma, rel_ecalEnergy, ecalEnergy_sig, eSuperClusterOverP, eSeedClusterOverP, eSeedClusterOverPout, eEleClusterOverPout, deltaEtaSuperClusterTrackAtVtx, deltaEtaSeedClusterTrackAtCalo, deltaEtaEleClusterTrackAtCalo, deltaPhiEleClusterTrackAtCalo, deltaPhiSuperClusterTrackAtVtx, deltaPhiSeedClusterTrackAtCalo.*
  - Shower shape variable: *sigmaEtaEta, sigmaIetaIeta, sigmaIphiIphi, sigmaIetaIphi, e1x5, e2x5Max, e5x5, r9, hcalDepth1OverEcal, hcalDepth2OverEcal, hcalDepth1OverEcalBc, hcalDepth2OverEcalBc, eLeft, eRight, eBottom, eTop, full5x5_sigmaEtaEta, full5x5_sigmaIetaIeta, full5x5_sigmaIphiIphi, full5x5_sigmaIetaIphi, full5x5_e1x5, full5x5_e2x5Max, full5x5_e5x5, full5x5_r9, full5x5_hcalDepth1OverEcal, full5x5_hcalDepth2OverEcal, full5x5_hcalDepth1OverEcalBc, full5x5_hcalDepth2OverEcalBc, full5x5_eLeft, full5x5_eRight, full5x5_eBottom, full5x5_eTop, full5x5_e2x5Left, full5x5_e2x5Right, full5x5_e2x5Bottom, full5x5_e2x5Top.*

- RECO muons:
  - Positional information (in $\eta - \phi$ plane): *r, theta.*
  - Particle properties: *rel_pt, particle_type (token identifier).*
  - Calorimeter information: *segmentCompatibility, caloCompatibility, pfEcalEnergy_valid, rel_pfEcalEnergy.*
  - Track and impact parameter features: *dxy, dxy_sig, normalizedChi2_valid, normalizedChi2, numberOfValidHits.*
  - Muon chambers information: *n_matches_DT_{1,2,3,4}, n_matches_CSC_{1,2,3,4}, n_matches_RPC_{1,2,3,4}, n_hits_DT_{1,2,3,4}, n_hits_CSC_{1,2,3,4}, n_hits_RPC_{1,2,3,4}*

## A.4 TaT ablation study: impact of modalities

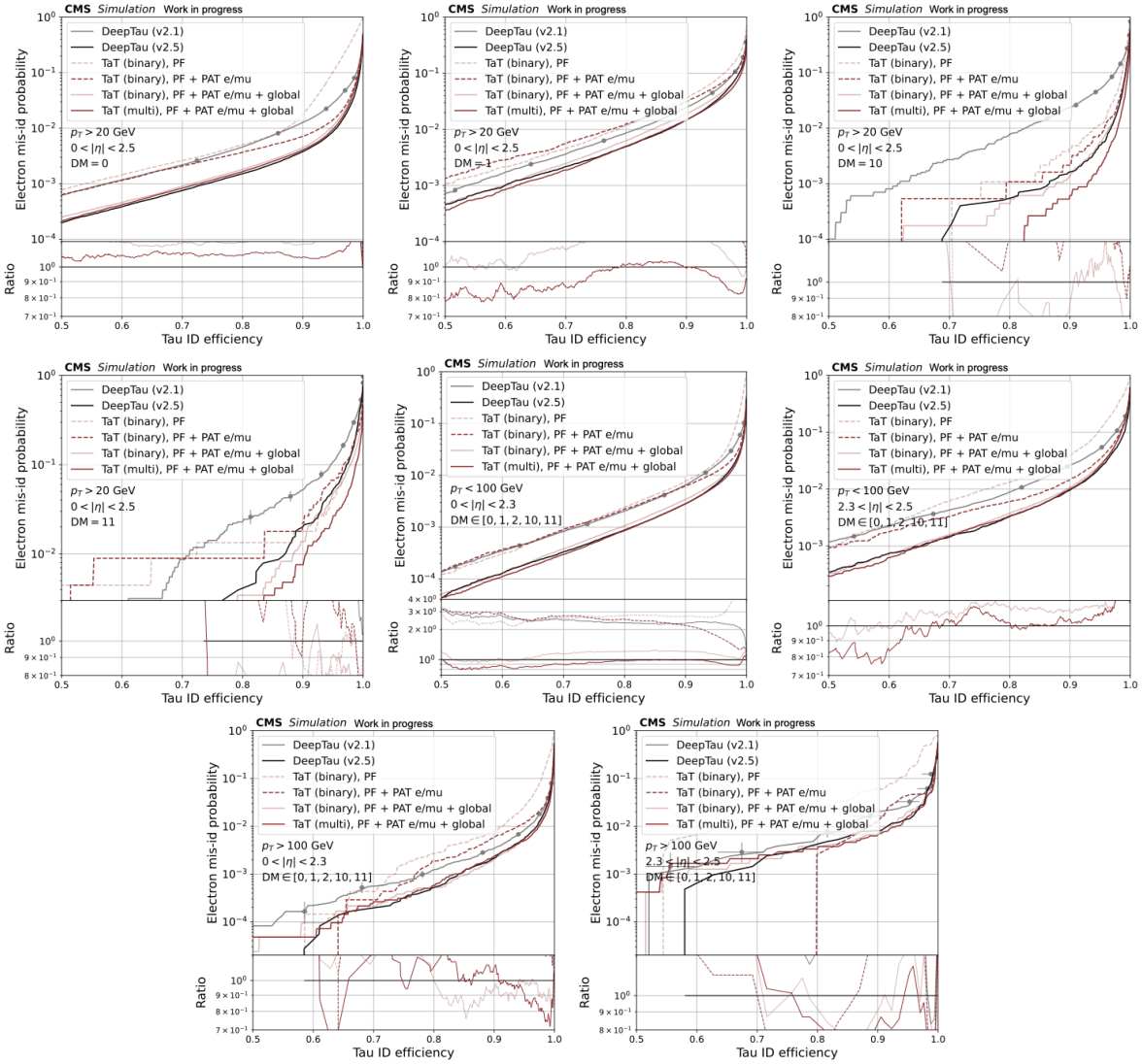


Figure A.1: Efficiency for electrons versus efficiency for genuine $\tau_h$ to pass the corresponding $D_\alpha$ discriminators for DeepTau v2.1, DeepTau v2.5 and TaT models for various $p_T$, $\eta$, and $\tau_h$ decay mode regions. For the TaT model, several configurations are trained with various modalities used as an input: PF candidates only on a binary classification task (dashed red, pale), PF candidates with RECO electrons/muons (PAT e/mu in the legend) on a binary classification task (dashed red, dark), PF candidates with RECO electrons/muons and global features on a binary classification task (solid red, pale), PF candidates with RECO electrons/muons and global features on a multiclass classification task (solid red, dark). Working points (grey dots) for DeepTau v2.1 are also shown, as derived in the original paper. The panel at the bottom of each figure shows the ratio of each of the ROC curves with respect to the one of the DeepTau v2.5 model.

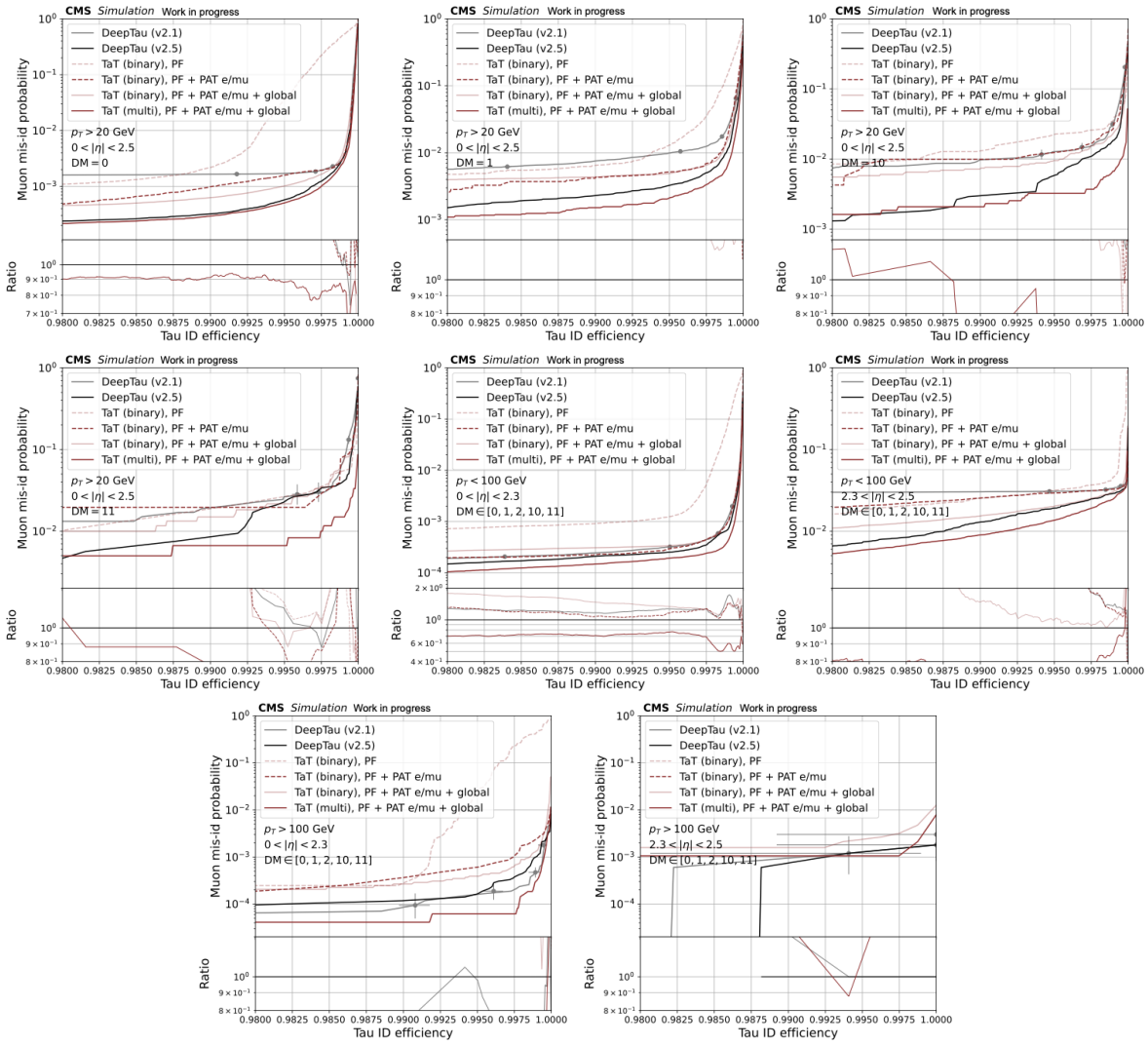


Figure A.2: Efficiency for muons versus efficiency for genuine $\tau_h$ to pass the corresponding $D_\alpha$ discriminators for DeepTau v2.1, DeepTau v2.5 and TaT models for various $p_T$, $\eta$, and $\tau_h$ decay mode regions. For the TaT model, several configurations are trained with various modalities used as an input: PF candidates only on a binary classification task (dashed red, pale), PF candidates with RECO electrons/muons (PAT e/mu in the legend) on a binary classification task (dashed red, dark), PF candidates with RECO electrons/muons and global features on a binary classification task (solid red, pale), PF candidates with RECO electrons/muons and global features on a multiclass classification task (solid red, dark). Working points (grey dots) for DeepTau v2.1 are also shown, as derived in the original paper. The panel at the bottom of each figure shows the ratio of each of the ROC curves with respect to the one of the DeepTau v2.5 model.

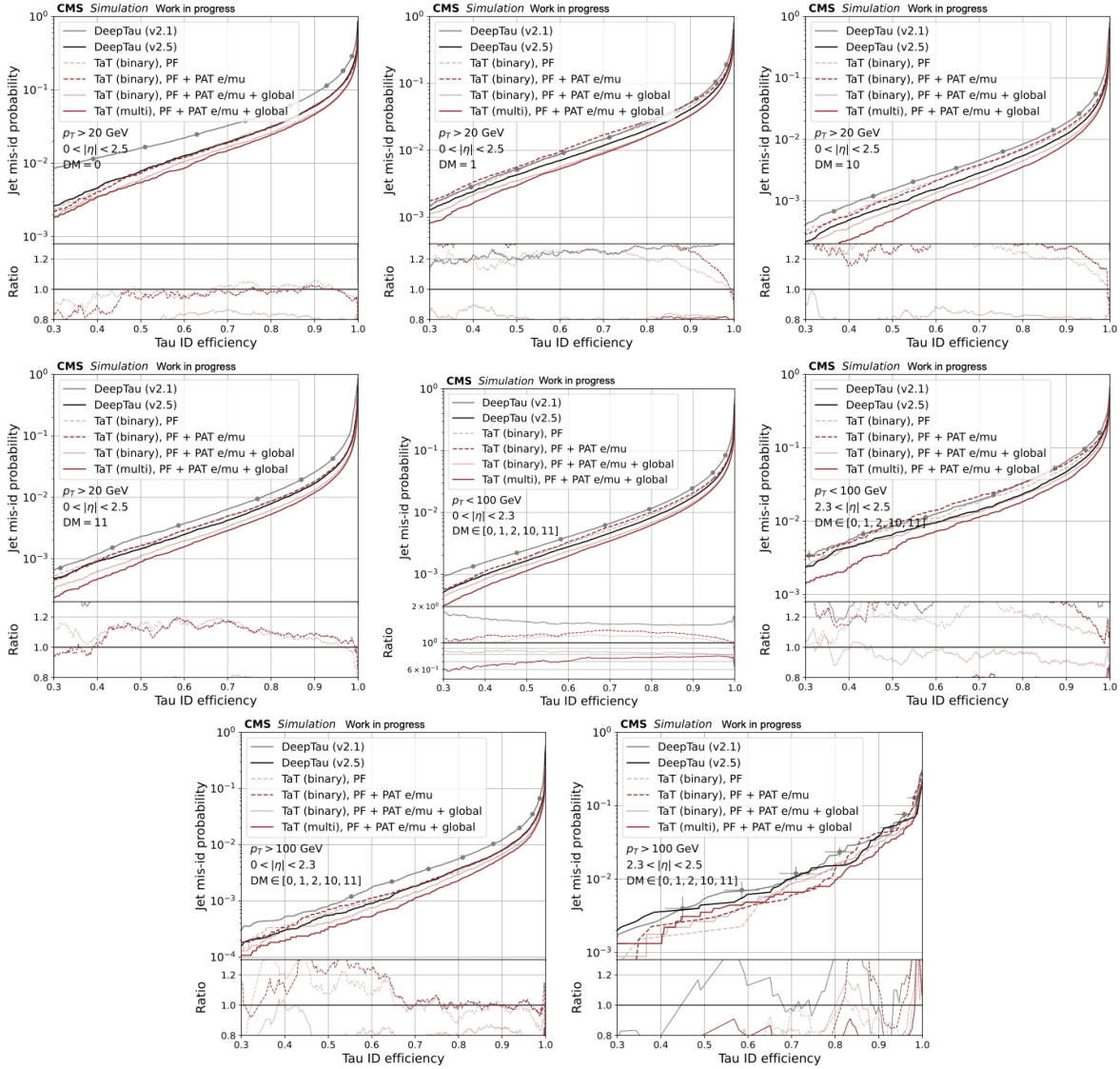


Figure A.3: Efficiency for jets versus efficiency for genuine $\tau_h$ to pass the corresponding $D_\alpha$ discriminators for DeepTau v2.1, DeepTau v2.5 and TaT models for various $p_T$, $\eta$, and $\tau_h$ decay mode regions. For the TaT model, several configurations are trained with various modalities used as an input: PF candidates only on a binary classification task (dashed red, pale), PF candidates with RECO electrons/muons (PAT e/mu in the legend) on a binary classification task (dashed red, dark), PF candidates with RECO electrons/muons and global features on a binary classification task (solid red, pale), PF candidates with RECO electrons/muons and global features on a multiclass classification task (solid red, dark). Working points (grey dots) for DeepTau v2.1 are also shown, as derived in the original paper. The panel at the bottom of each figure shows the ratio of each of the ROC curves with respect to the one of the DeepTau v2.5 model.

## A.5 Performance comparison for TaT and ParticleNet

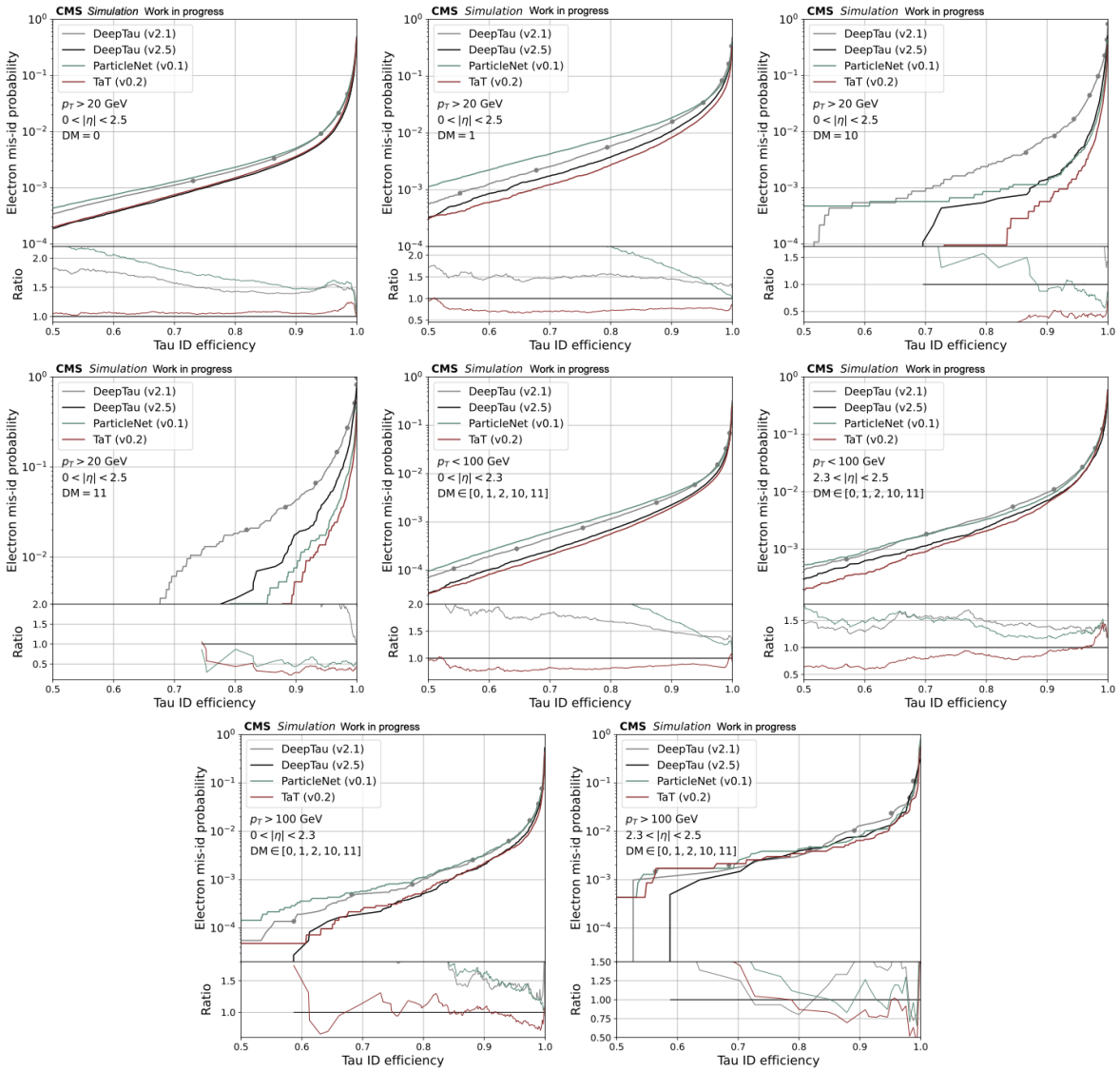


Figure A.4: Efficiency for electrons versus efficiency for genuine $\tau_h$ to pass the corresponding $D_\alpha$ discriminators for DeepTau v2.1 (grey), DeepTau v2.5 (black), ParticleNet v0.1 (dark cyan) and TaT v0.2 (red) models. Working points (grey dots) for DeepTau v2.1 are also shown, as derived in the original paper. The panel at the bottom of each figure shows the ratio of each of the ROC curves with respect to the one of the DeepTau v2.5 model.

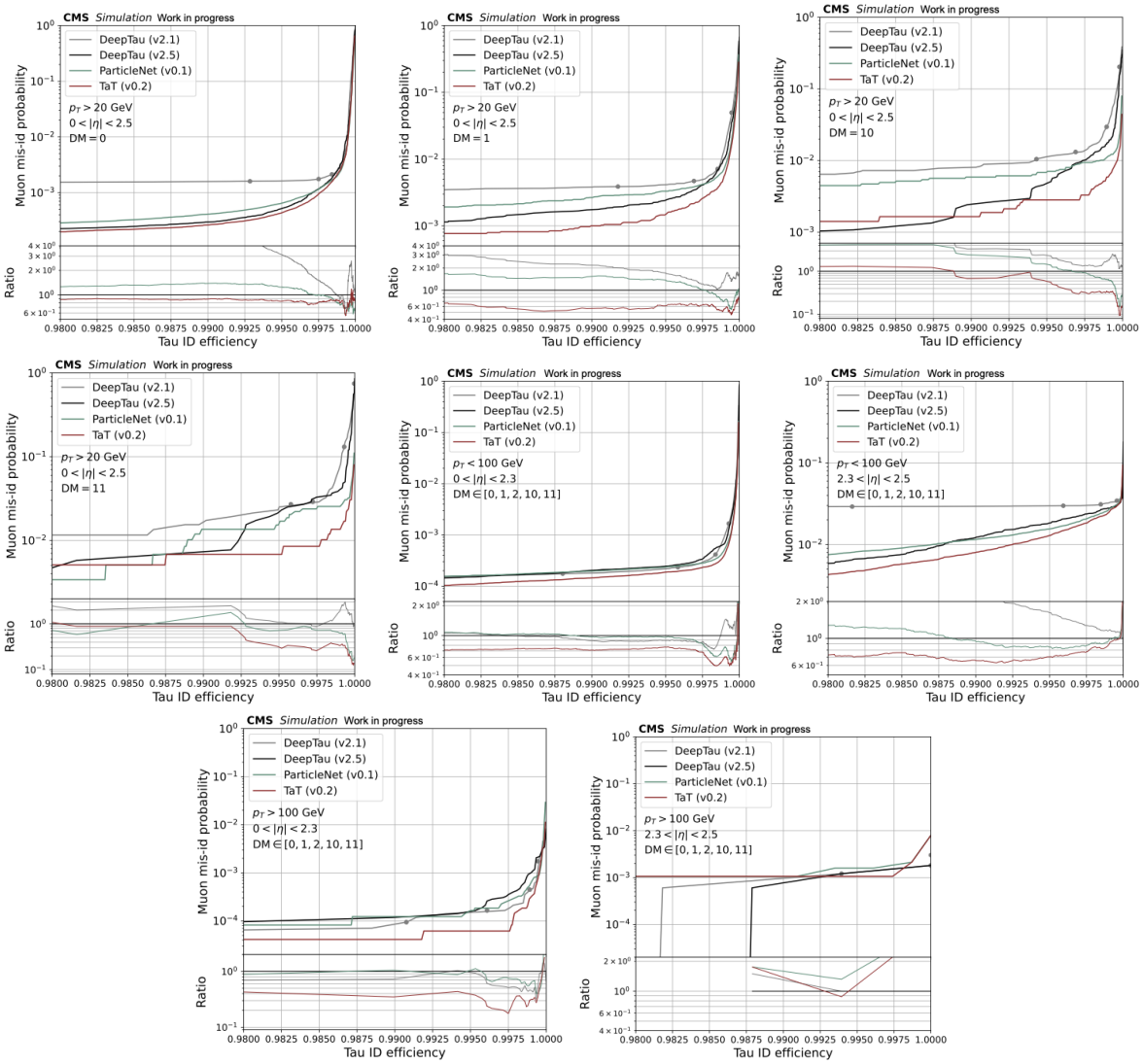


Figure A.5: Efficiency for muons versus efficiency for genuine $\tau_h$ to pass the corresponding $D_\alpha$ discriminators for DeepTau v2.1 (grey), DeepTau v2.5 (black), ParticleNet v0.1 (dark cyan) and TaT v0.2 (red) models. Working points (grey dots) for DeepTau v2.1 are also shown, as derived in the original paper. The panel at the bottom of each figure shows the ratio of each of the ROC curves with respect to the one of the DeepTau v2.5 model.

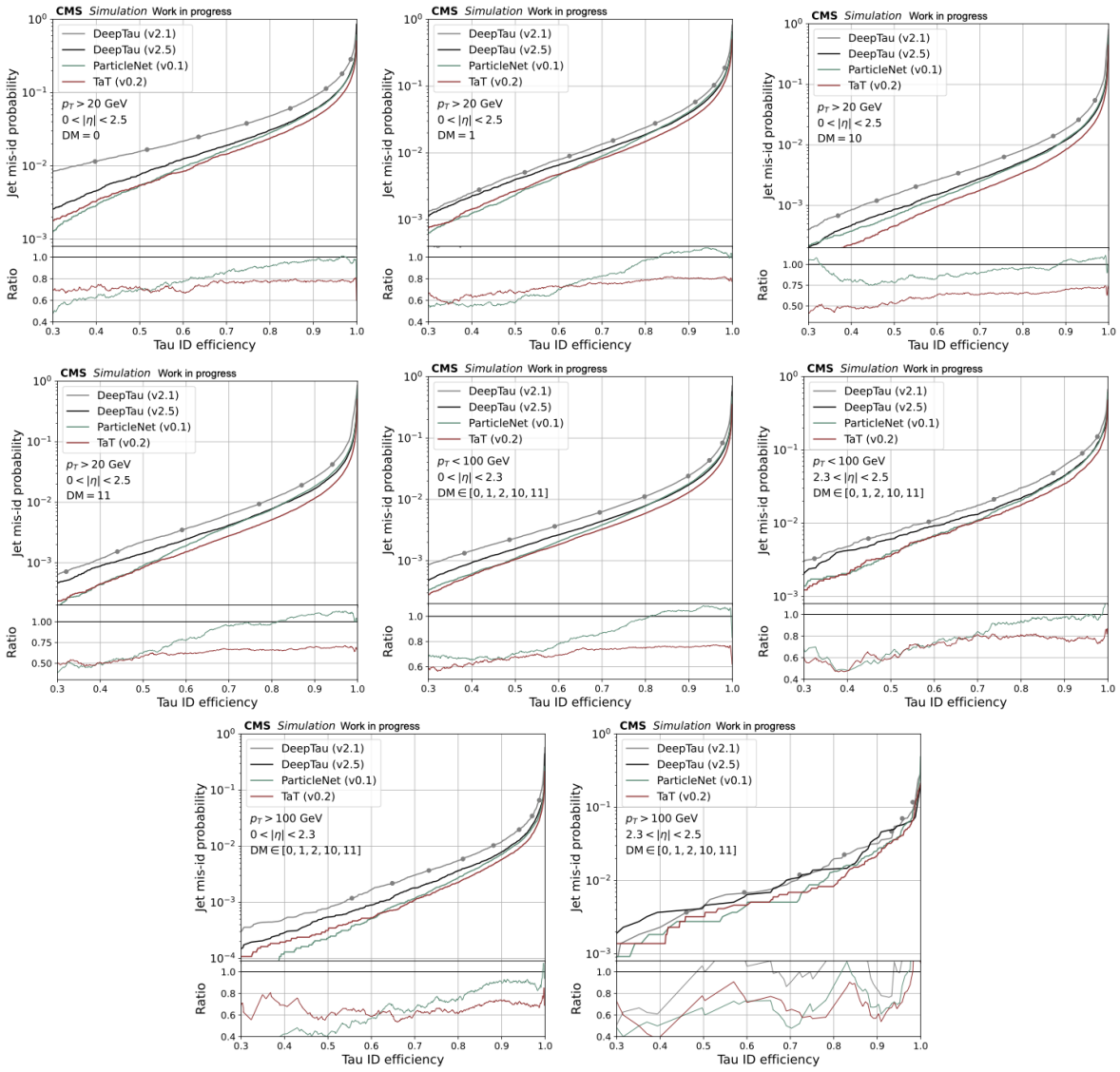


Figure A.6: Efficiency for jets versus efficiency for genuine $\tau_h$ to pass the corresponding $D_\alpha$ discriminators for DeepTau v2.1 (grey), DeepTau v2.5 (black), ParticleNet v0.1 (dark cyan) and TaT v0.2 (red) models. Working points (grey dots) for DeepTau v2.1 are also shown, as derived in the original paper. The panel at the bottom of each figure shows the ratio of each of the ROC curves with respect to the one of the DeepTau v2.5 model.

## A.6 TaT ablation study: variation of the cone size

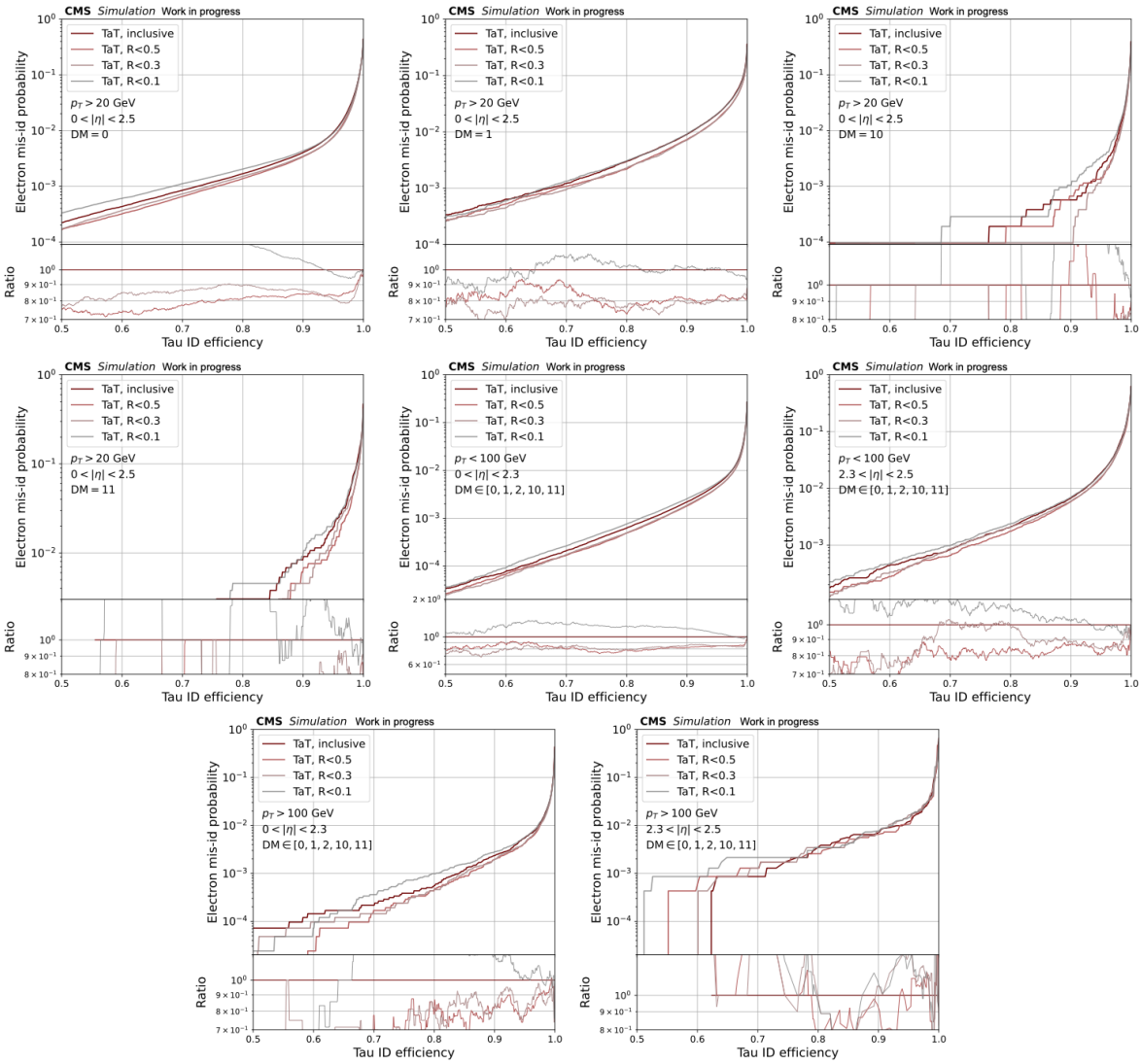


Figure A.7: Efficiency for electrons versus efficiency for genuine $\tau_h$ to pass the corresponding $D_\alpha$ discriminators for a TaT architecture with the various requirements on the cone distance between the directions of flight of constituents and $\tau_h$ candidate (R). The panel at the bottom of each figure shows the ratio of each of the ROC curves with respect to the model without any requirement on the cone distance (inclusive).

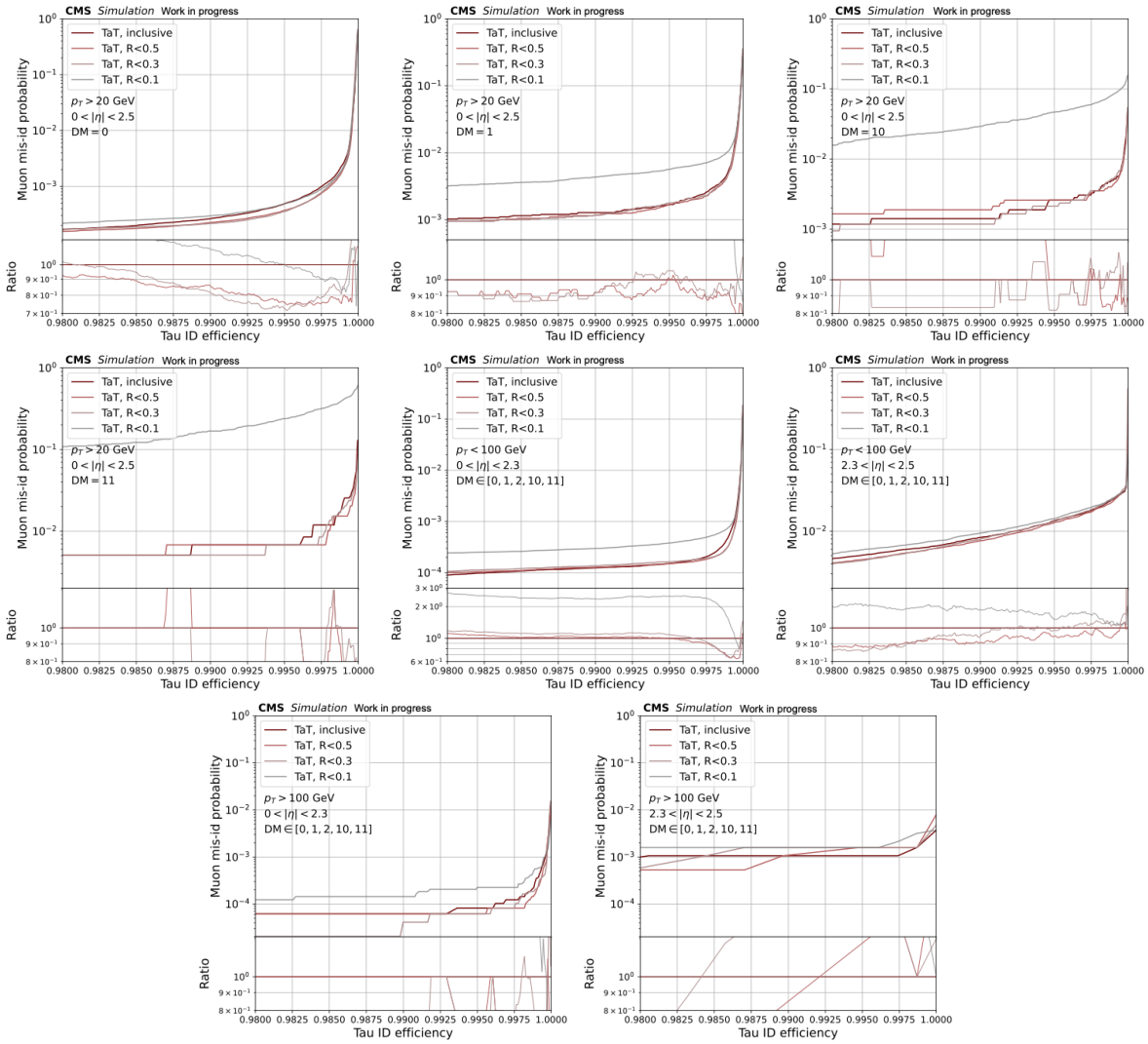


Figure A.8: Efficiency for muons versus efficiency for genuine $\tau_h$ to pass the corresponding $D_\alpha$ discriminators for a TaT architecture with the various requirements on the cone distance between the directions of flight of constituents and $\tau_h$ candidate (R). The panel at the bottom of each figure shows the ratio of each of the ROC curves with respect to the model without any requirement on the cone distance (inclusive).

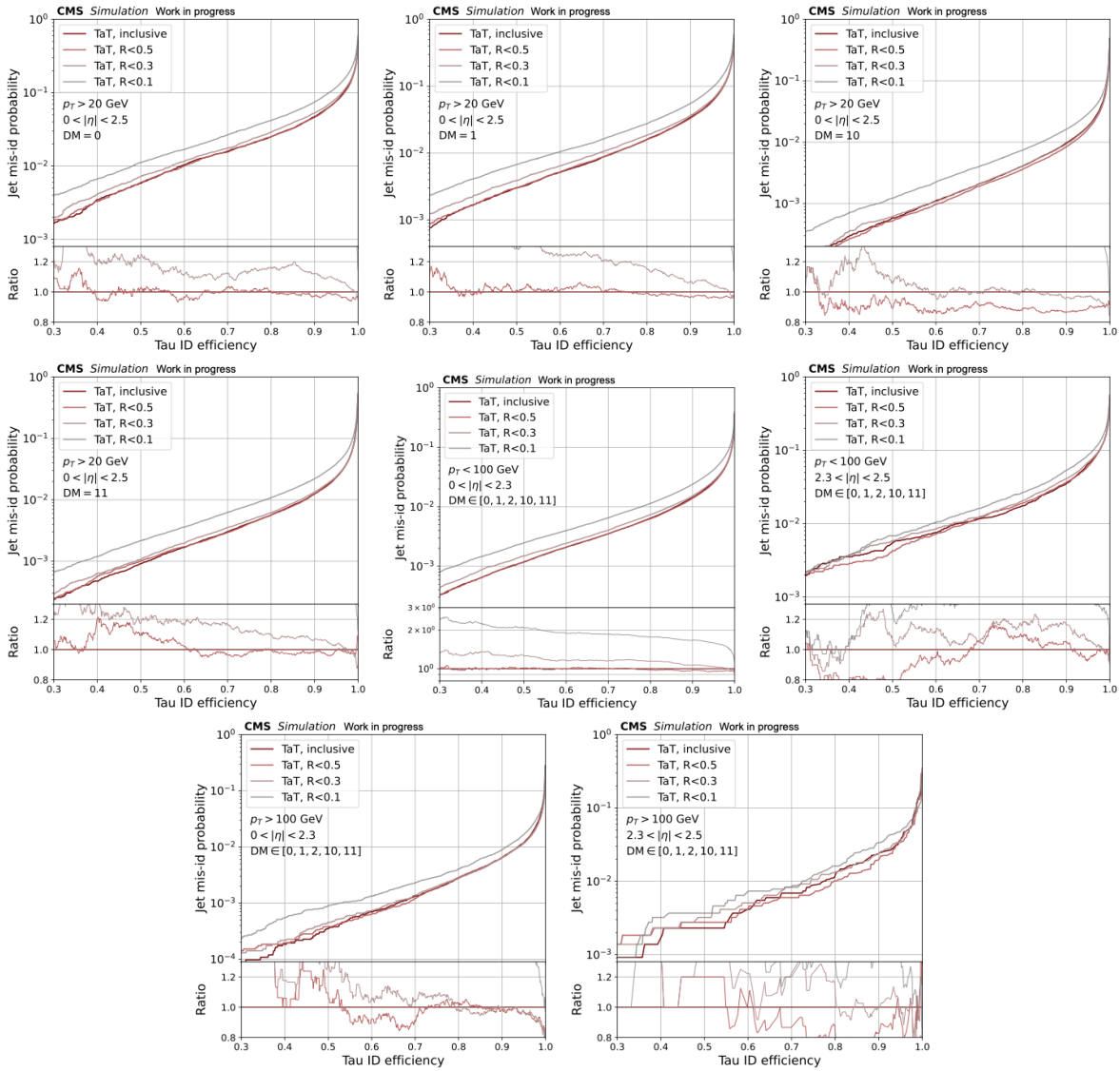


Figure A.9: Efficiency for jets versus efficiency for genuine $\tau_h$ to pass the corresponding $D_\alpha$ discriminators for a TaT architecture with the various requirements on the cone distance between the directions of flight of constituents and $\tau_h$ candidate (R). The panel at the bottom of each figure shows the ratio of each of the ROC curves with respect to the model without any requirement on the cone distance (inclusive).

## A.7 Control plots in the $\tau_e\tau_h$ channel

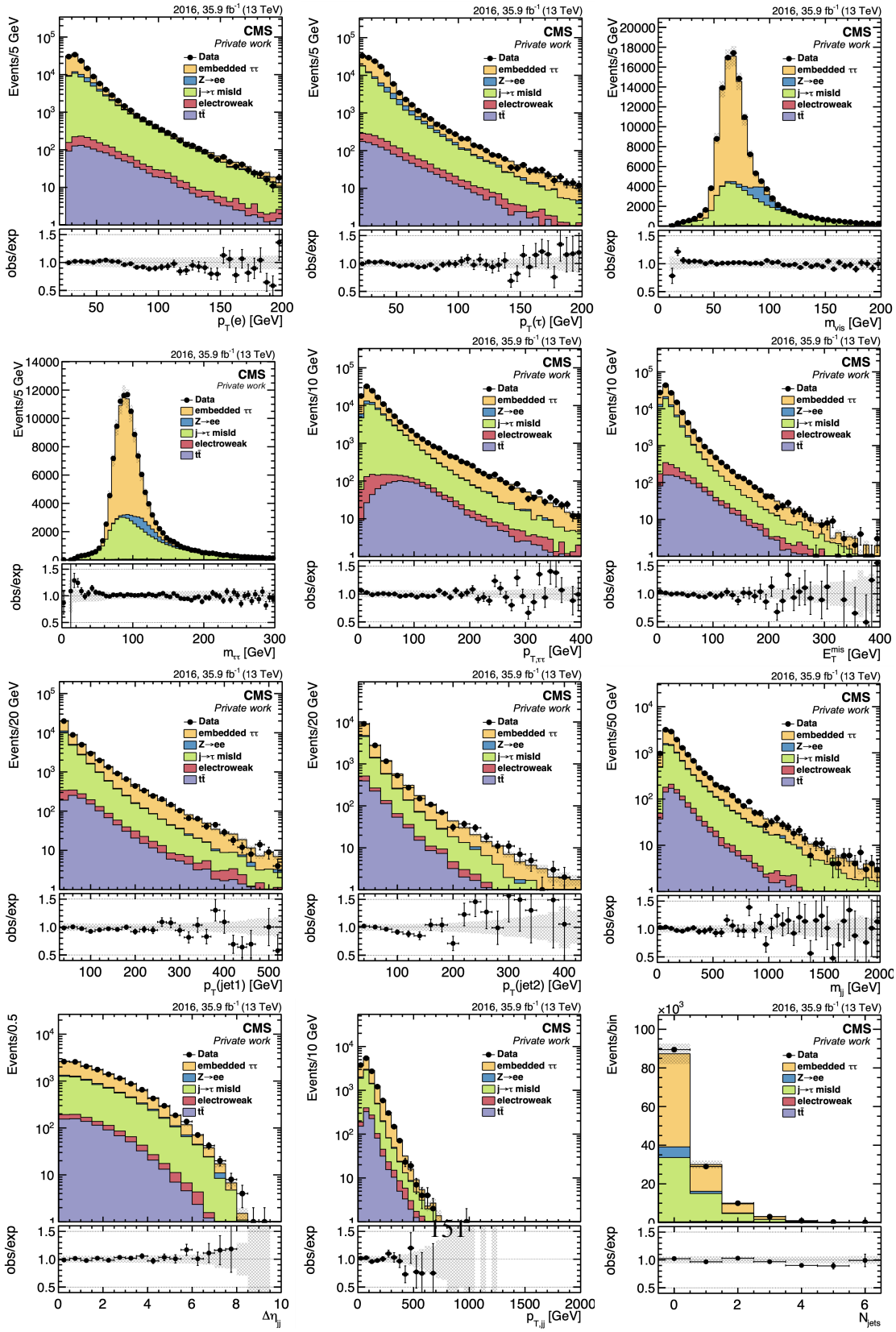


Figure A.10: Comparison of data with simulation for the 2016 data-taking period for the variables used in the neural network training, as described in Sec. 5.6.

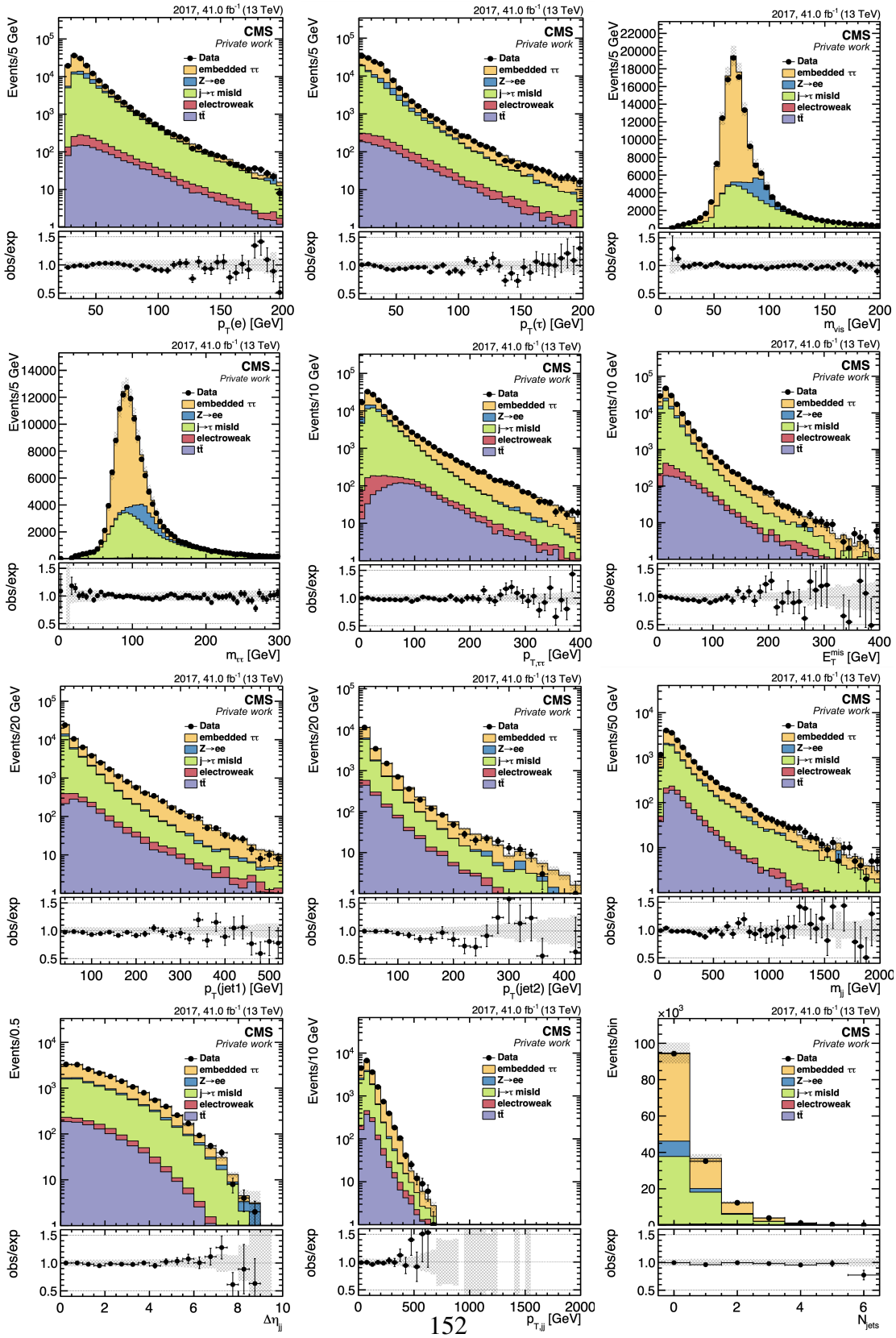


Figure A.11: Comparison of data with simulation for the 2017 data-taking period for the variables used in the neural network training, as described in Sec. 5.6.

## A.8 Pre-fit distributions in the $\tau_e\tau_h$ channel

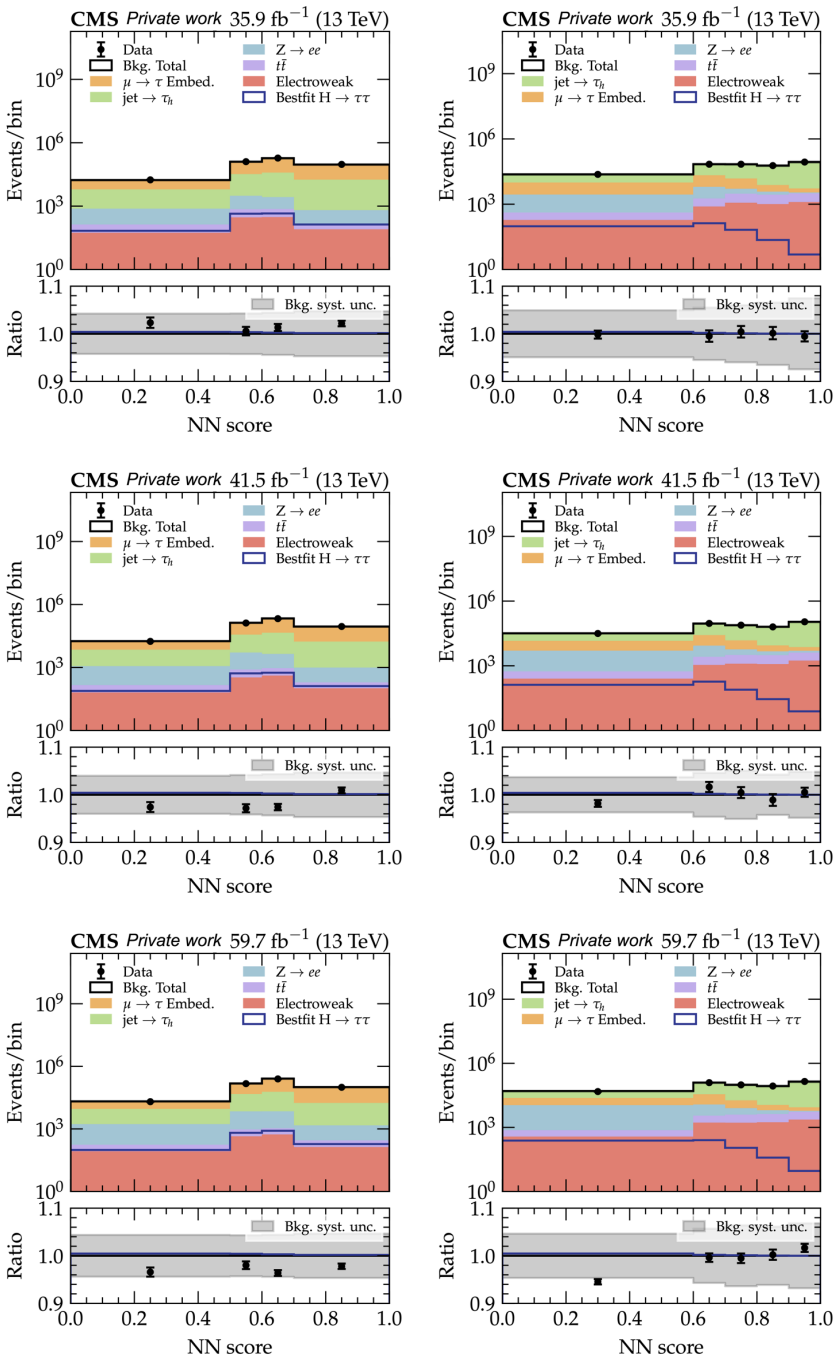


Figure A.12: Pre-fit distribution of the NN score in the genuine $\tau$ (left) and fakes (right) background categories for the 2016 (top), 2017 (middle), and 2018 (bottom) data-taking periods.

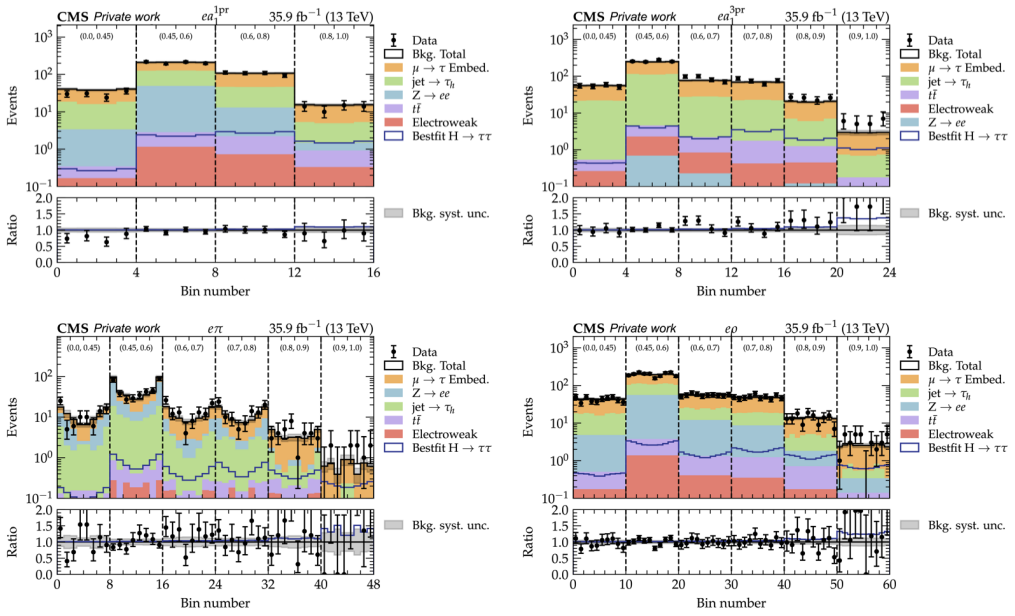


Figure A.13: Pre-fit distributions of the unrolled $\phi_{CP}$ observable in bins of the NN score in the signal categories for the 2016 data-taking period.

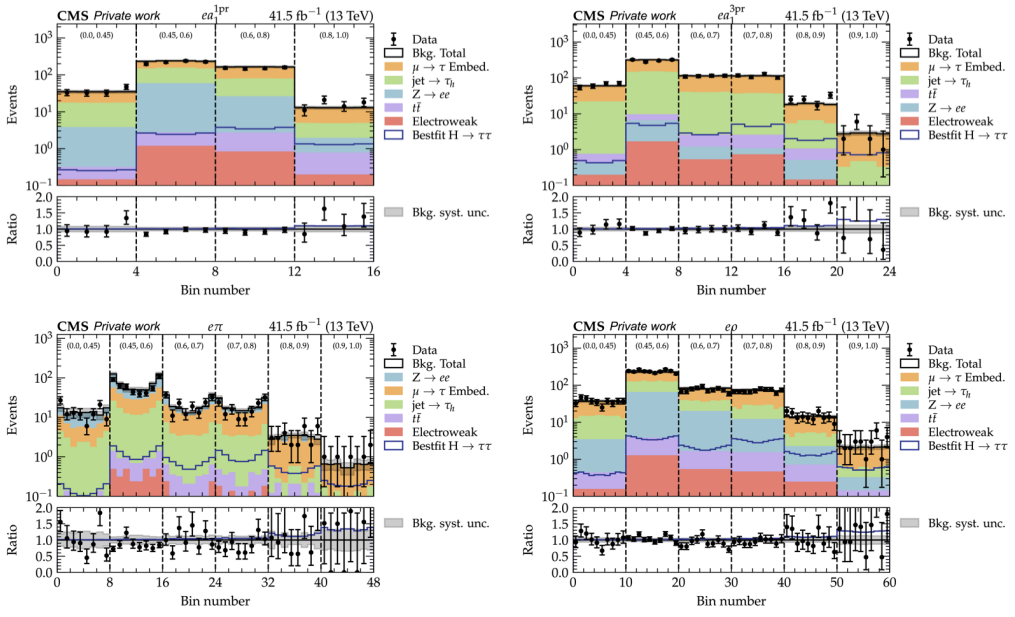


Figure A.14: Pre-fit distributions of the unrolled $\phi_{CP}$ observable in bins of the NN score in the signal categories for the 2017 data-taking period.

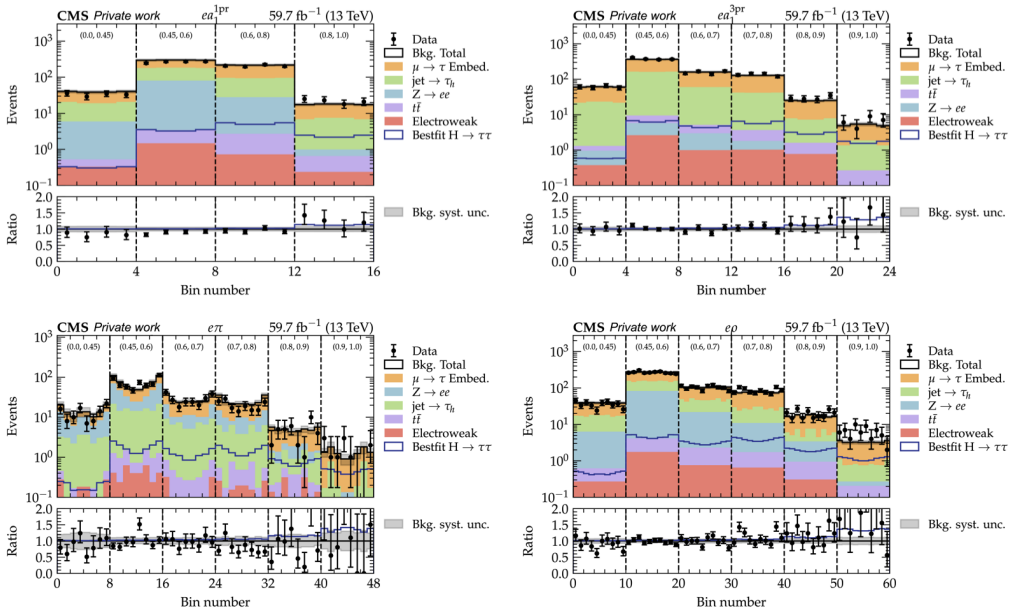


Figure A.15: Pre-fit distributions of the unrolled $\phi_{CP}$ observable in bins of the NN score in the signal categories for the 2018 data-taking period.

## A.9 Post-fit distributions in the $\tau_e\tau_h$ channel

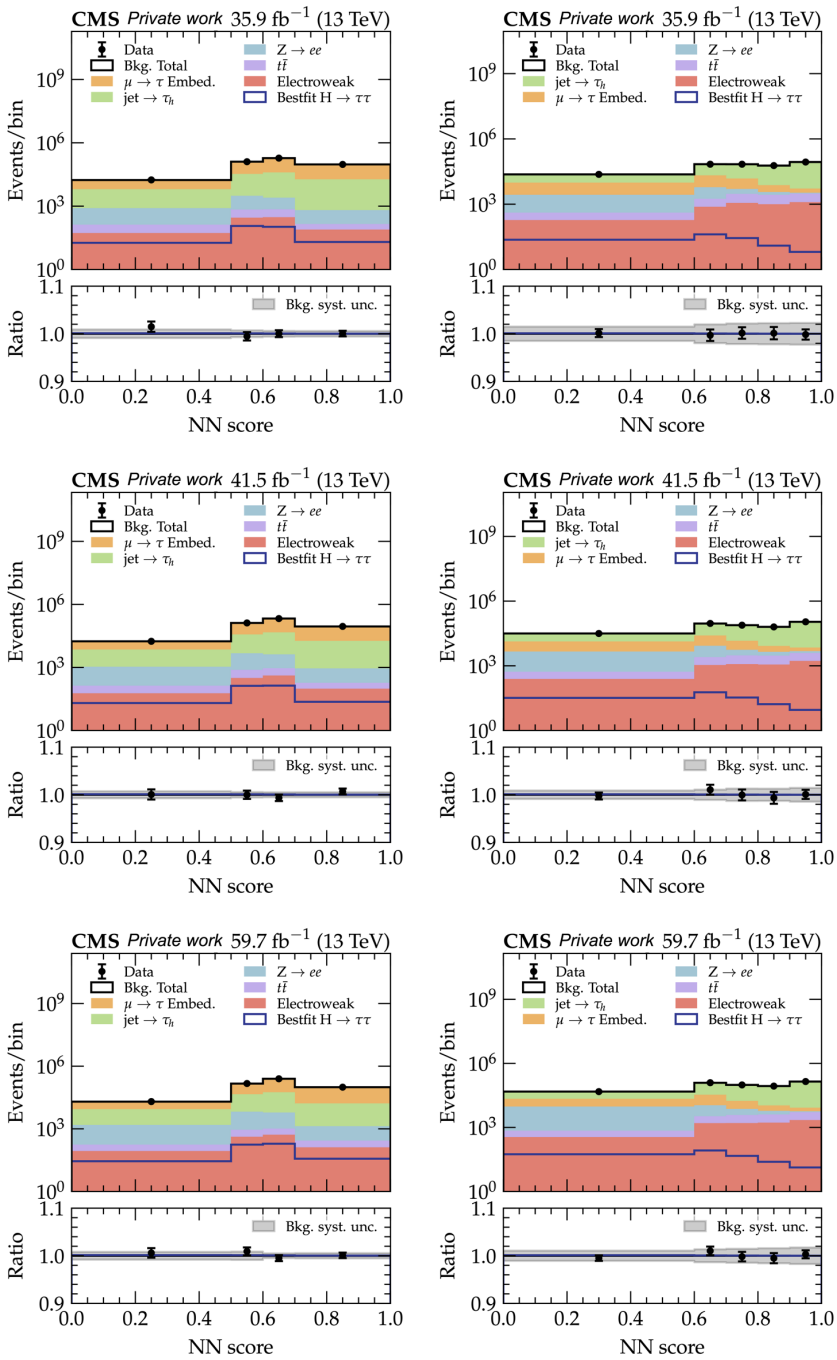


Figure A.16: Post-fit distribution of the NN score in the genuine $\tau$ (left) and fakes (right) background categories for the 2016 (top), 2017 (middle), and 2018 (bottom) data-taking periods.

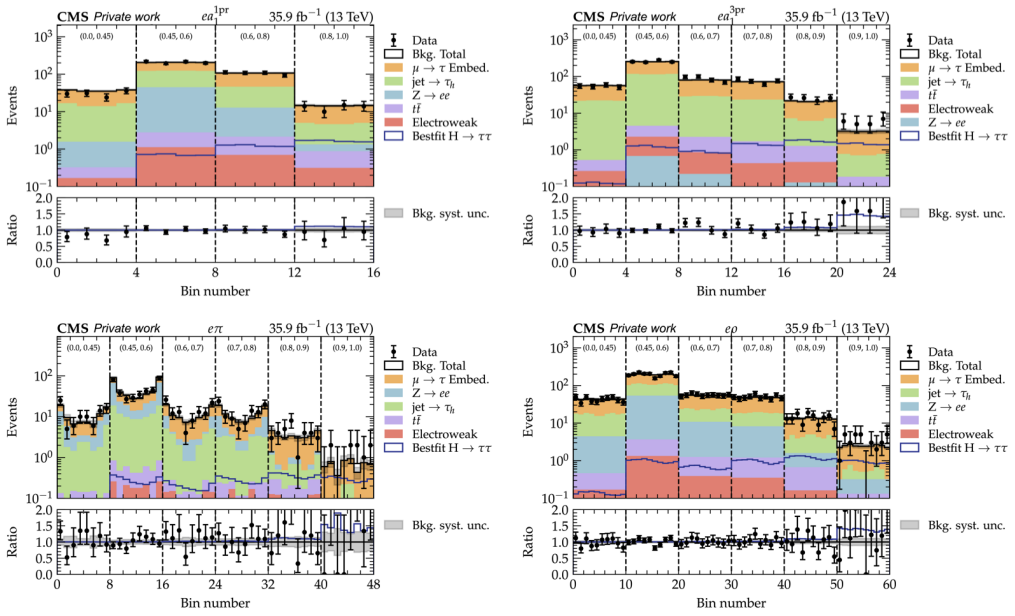


Figure A.17: Post-fit distributions of the unrolled $\phi_{CP}$ observable in bins of the NN score in the signal categories for the 2016 data-taking period.

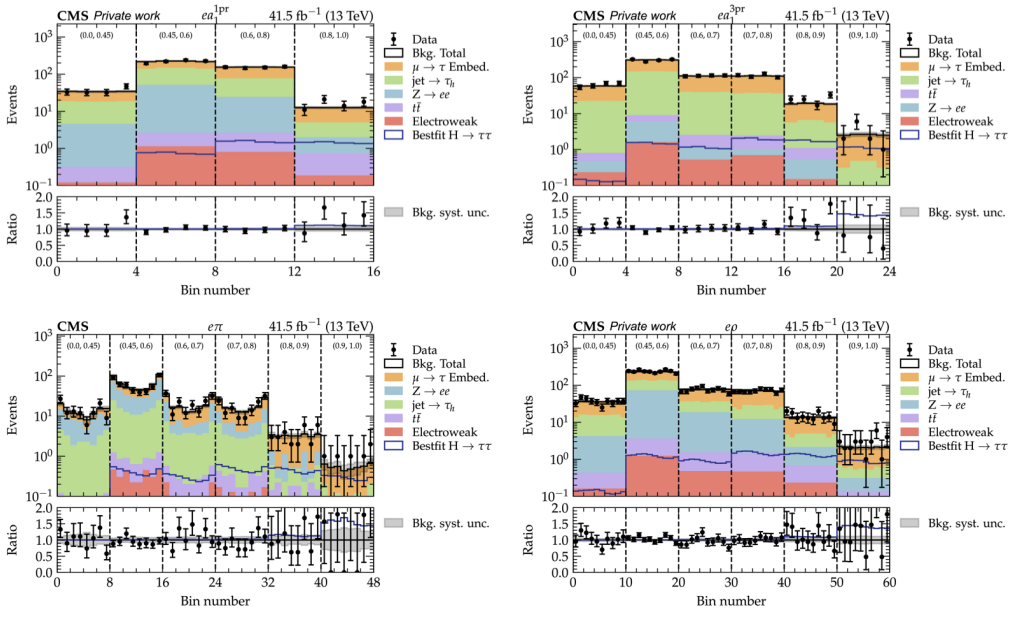


Figure A.18: Post-fit distributions of the unrolled $\phi_{CP}$ observable in bins of the NN score in the signal categories for the 2017 data-taking period.

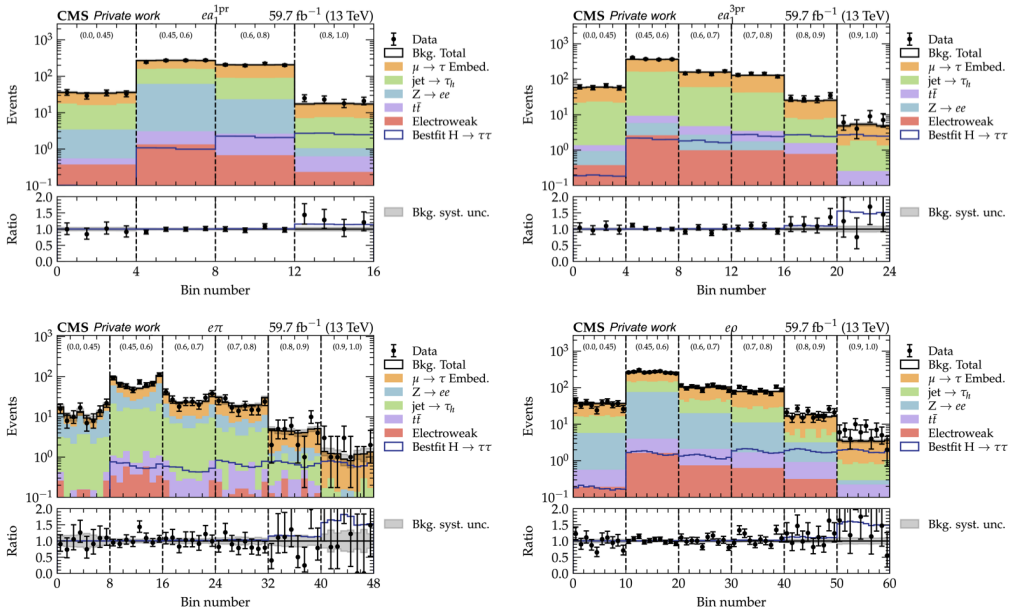


Figure A.19: Post-fit distributions of the unrolled $\phi_{CP}$ observable in bins of the NN score in the signal categories for the 2018 data-taking period.

# Bibliography


[1] A.D. Sakharov, *Violation of CP in variance, C asymmetry, and baryon asymmetry of the universe*, *Phys. Usp.* **34** (1991) 392.

[2] M. Mulders and C. Duhr, eds., *Proceedings, 2018 European School of High-Energy Physics (ESHEP 2018): Maratea, Italy, June 20 – July 3 2018*, vol. 6/2019 of *CERN Yellow Reports: School Proceedings*, (Geneva), CERN, 2019. 10.23730/CYRSP-2019-006.

[3] ATLAS collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, *Phys. Lett. B* **716** (2012) 1 [1207.7214].

[4] CMS collaboration, *Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC*, *Phys. Lett. B* **716** (2012) 30 [1207.7235].

[5] Wikipedia, "Illustration of the Standard Model of Elementary Particles." https://en.wikipedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg.

[6] 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** (1957) 1413.

[7] Particle Data Group, *Review of Particle Physics*, *PTEP* **2020** (2020) 083C01.

[8] D.P. Barber et al., *Discovery of Three Jet Events and a Test of Quantum Chromodynamics at PETRA Energies*, *Phys. Rev. Lett.* **43** (1979) 830.

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

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

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

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

[13] KTeV collaboration, *Observation of direct CP violation in $K_{S,L} \to \pi\pi$ decays*, *Phys. Rev. Lett.* **83** (1999) 22 [hep-ex/9905060].

[14] LHCb collaboration, *First observation of CP violation in the decays of $B_s^0$ mesons*, *Phys. Rev. Lett.* **110** (2013) 221601 [1304.6173].

[15] LHCb collaboration, *Observation of CP Violation in Charm Decays*, *Phys. Rev. Lett.* **122** (2019) 211803 [1903.08726].

[16] CMS collaboration, *A portrait of the Higgs boson by the CMS experiment ten years after the discovery*, *Nature* **607** (2022) 60 [2207.00043].

[17] ATLAS collaboration, *A detailed map of Higgs boson interactions by the ATLAS experiment ten years after the discovery*, *Nature* **607** (2022) 52 [2207.00092].

[18] LHC Higgs Cross Section Working Group collaboration, *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, 1610.07922.

[19] LHC Higgs Cross Section Working Group collaboration, *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties*, 1307.1347.

[20] CMS collaboration, *Measurement of the Higgs boson width and evidence of its off-shell contributions to ZZ production*, *Nature Phys.* **18** (2022) 1329 [2202.06923].

[21] CMS collaboration, *Evidence for Higgs boson decay to a pair of muons*, *JHEP* **01** (2021) 148 [2009.04363].

[22] CMS collaboration, *Search for Higgs boson decay to a charm quark-antiquark pair in proton-proton collisions at $\sqrt{s} = 13$ TeV*, 2205.05550.

[23] CMS collaboration, *Measurement of the Properties of a Higgs Boson in the Four-Lepton Final State*, *Phys. Rev. D* **89** (2014) 092007 [1312.5353].

[24] ATLAS collaboration, *Study of the spin and parity of the Higgs boson in diboson decays with the ATLAS detector*, *Eur. Phys. J. C* **75** (2015) 476 [1506.05669].

[25] CMS collaboration, *Measurements of $t\bar{t}H$ Production and the CP Structure of the Yukawa Interaction between the Higgs Boson and Top Quark in the Diphoton Decay Channel*, *Phys. Rev. Lett.* **125** (2020) 061801 [2003.10866].

[26] ATLAS collaboration, *CP Properties of Higgs Boson Interactions with Top Quarks in the $t\bar{t}H$ and $tH$ Processes Using $H \to \gamma\gamma$ with the ATLAS Detector*, *Phys. Rev. Lett.* **125** (2020) 061802 [2004.04545].

[27] CMS collaboration, *Search for CP violation in ttH and tH production in multilepton channels in proton-proton collisions at $\sqrt{s} = 13$ TeV*, 2208.02686.

[28] ATLAS collaboration, *Probing the CP nature of the top-Higgs Yukawa coupling in $t\bar{t}H$ and $tH$ events with $H \to b\bar{b}$ using the ATLAS detector at the LHC*, *ATLAS-CONF-2022-016*, 2022.

[29] CMS collaboration, *Constraints on anomalous Higgs boson couplings to vector bosons and fermions in its production and decay using the four-lepton final state*, *Phys. Rev. D* **104** (2021) 052004 [2104.12152].

[30] CMS collaboration, *Constraints on anomalous Higgs boson couplings to vector bosons and fermions from the production of Higgs bosons using the $\tau\tau$ final state*, 2205.05120.

[31] ATLAS collaboration, *Study of the CP property of the Higgs boson to electroweak boson coupling in the VBF $H \to \gamma\gamma$ channel with the ATLAS detector*, 2208.02338.

[32] ATLAS collaboration, *Measurements of the Higgs boson inclusive and differential fiducial cross-sections in the diphoton decay channel with pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *JHEP* **08** (2022) 027 [2202.00487].

[33] L. Evans and P. Bryant, eds., *LHC Machine*, *JINST* **3** (2008) S08001.

[34] Mobs, Esma, "The CERN accelerator complex in 2019. Complexe des accÃľlÃľrateurs du CERN en 2019." https://cds.cern.ch/record/2684277, 2019.

[35] CMS collaboration, *The CMS Experiment at the CERN LHC*, *JINST* **3** (2008) S08004.

[36] ATLAS collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, *JINST* **3** (2008) S08003.

[37] LHCb collaboration, *The LHCb Detector at the LHC*, *JINST* **3** (2008) S08005.

[38] ALICE collaboration, *The ALICE experiment at the CERN LHC*, *JINST* **3** (2008) S08002.

[39] TOTEM collaboration, *The TOTEM experiment at the CERN Large Hadron Collider*, *JINST* **3** (2008) S08007.

[40] LHCf collaboration, *The LHCf detector at the CERN Large Hadron Collider*, *JINST* **3** (2008) S08006.

[41] MoEDAL collaboration, *Technical Design Report of the MoEDAL Experiment*, .

[42] J.L. Feng, I. Galon, F. Kling and S. Trojanowski, *ForwArd Search ExpeRiment at the LHC*, *Phys. Rev. D* **97** (2018) 035001 [1708.09389].

[43] SHiP collaboration, *SND@LHC*, 2002.08722.

[44] CMS collaboration, "Public cms luminosity information." https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults, 2022.

[45] I. Zurbano Fernandez et al., *High-Luminosity Large Hadron Collider (HL-LHC): Technical design report*, *CYRM* **10/2020** (2020) .

[46] CMS collaboration, "Illustration of the cms detector." https://cms.cern/detector.

[47] BaBar collaboration, *The BaBar detector*, *Nucl. Instrum. Meth. A* **479** (2002) 1 [hep-ex/0105044].

[48] CMS Tracker Group collaboration, *The CMS Phase-1 Pixel Detector Upgrade*, *JINST* **16** (2021) P02027 [2012.14304].

[49] CMS collaboration, "Track impact parameter resolution for the full pseudo rapidity coverage in the 2017 dataset with the CMS Phase-1 Pixel detector." https://cds.cern.ch/record/2743740, 2020.

[50] S. Banerjee, "Performance of Hadron Calorimeter with and without HO, CMS-NOTE-1999-063."

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

[52] CMS collaboration, T. Hebbeker and A. Korytov, *The Phase-2 Upgrade of the CMS Muon Detectors, CERN-LHCC-2017-012, CMS-TDR-016*, 2017.

[53] CMS collaboration, *The CMS trigger system*, *JINST* **12** (2017) P01020 [1609.02366].

[54] I. Bird et al., *LHC computing Grid. Technical design report, CERN-LHCC-2005-024*, 2005.

[55] P. Fayet, *Supergauge Invariant Extension of the Higgs Mechanism and a Model for the electron and Its Neutrino*, *Nucl. Phys. B* **90** (1975) 104.

[56] P. Fayet, *Spontaneously Broken Supersymmetric Theories of Weak, Electromagnetic and Strong Interactions*, *Phys. Lett. B* **69** (1977) 489.

[57] CMS collaboration, *Searches for additional Higgs bosons and for vector leptoquarks in $\tau\tau$ final states in proton-proton collisions at $\sqrt{s} = 13$ TeV*, 2208.02717.

[58] HFLAV collaboration, *Averages of b-hadron, c-hadron, and $\tau$-lepton properties as of 2021*, 2206.07501.

[59] Belle II collaboration, *Overview of R(D) and R(D*)*, in *20th Conference on Flavor Physics and CP Violation* , 7, 2022 [2207.14146].

[60] LHCb collaboration, *Measurement of the ratio of branching fractions $\mathcal{B}(B_c^+ \to J/\psi\tau^+\nu_\tau)/\mathcal{B}(B_c^+ \to J/\psi\mu^+\nu_\mu)$*, *Phys. Rev. Lett.* **120** (2018) 121801 [1711.05623].

[61] LHCb collaboration, *Test of lepton universality in beauty-quark decays*, *Nature Phys.* **18** (2022) 277 [2103.11769].

[62] M.L. Perl et al., *Evidence for Anomalous Lepton Production in e+ - e- Annihilation*, *Phys. Rev. Lett.* **35** (1975) 1489.

[63] M.L. Perl, *The Discovery of the tau lepton*, in *3rd International Symposium on the History of Particle Physics: The Rise of the Standard Model*, pp. 79–100, 9, 1992.

[64] M.L. Perl, *Reflections on the discovery of the tau lepton, SLAC-PUB-7049*, 3, 1996.

[65] M.L. Perl, "Searches for heavy leptons and anomalous leptonic behavior – the past and the future."
https://www.slac.stanford.edu/pubs/slacpubs/1000/slac-pub-1062.pdf, 1972.

[66] G.J. Feldman et al., *Inclusive Anomalous Muon Production in e+ e- Annihilation*, *Phys. Rev. Lett.* **38** (1977) 117.

[67] PLUTO collaboration, *Anomalous Muon Production in e+ e- Annihilation as Evidence for Heavy Leptons*, *Phys. Lett. B* **68** (1977) 297.

[68] A. Barbaro-Galtieri et al., *Electron-Muon and electron-Hadron Production in e+ e- Collisions*, *Phys. Rev. Lett.* **39** (1977) 1058.

[69] W. Bartel et al., *A Precise Determination of the $\tau$ - Mass*, *Phys. Lett. B* **77** (1978) 331.

[70] W. Bacino et al., *Measurement of the Threshold Behavior of tau+ tau- Production in e+ e- Annihilation*, *Phys. Rev. Lett.* **41** (1978) 13.

[71] W. Bacino et al., *Measurement of the Branching Ratios for $\tau \to \pi\nu_\tau$ and $\tau \to \mu\nu\_\mu\nu_\tau$*, *Phys. Rev. Lett.* **42** (1979) 6.

[72] DONUT collaboration, *Observation of tau neutrino interactions*, *Phys. Lett. B* **504** (2001) 218 [hep-ex/0012035].

[73] CMS collaboration, *Identification of hadronic tau lepton decays using a deep neural network*, *JINST* **17** (2022) P07023 [2201.08458].

[74] CMS collaboration, *Particle-flow reconstruction and global event description with the CMS detector*, *JINST* **12** (2017) P10003 [1706.04965].

[75] CMS collaboration, *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software, CERN-LHCC-2006-001, CMS-TDR-8-1, CERN-LHCC-2006-001, CMS-TDR-8-1*, 2006.

[76] CMS collaboration, *Description and performance of track and primary-vertex reconstruction with the CMS tracker*, *JINST* **9** (2014) P10009 [1405.6569].

[77] W. Adam, B. Mangano, T. Speer and T. Todorov, *Track reconstruction in the CMS tracker, CERN-CMS-NOTE-2006-041, CMS-NOTE-2006-041*, 12, 2005.

[78] W. Adam, R. Frühwirth, A. Strandlie and T. Todorov, *Reconstruction of electrons with the gaussian-sum filter in the cms tracker at the lhc*, *Journal of Physics G: Nuclear and Particle Physics* **31** (2005) N9.

[79] CMS collaboration, *Energy Calibration and Resolution of the CMS Electromagnetic Calorimeter in pp Collisions at* $\sqrt{s}$ *= 7 TeV*, *JINST* **8** (2013) P09009 [1306.2016].

[80] CMS collaboration, *Calibration of the CMS hadron calorimeters using proton-proton collision data at* $\sqrt{s}$ *= 13 TeV*, *JINST* **15** (2020) P05002 [1910.00079].

[81] J.L. Bentley, *Multidimensional binary search trees used for associative searching*, *Commun. ACM* **18** (1975) 509âĂŞ517.

[82] CMS collaboration, *Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC*, *JINST* **16** (2021) P05014 [2012.06888].

[83] CMS collaboration, *Performance of* $\tau$*-lepton reconstruction and identification in CMS*, *Journal of Instrumentation* **7** (2012) P01001.

[84] CMS collaboration, *Reconstruction and identification of* $\tau$ *lepton decays to hadrons and* $\nu$ *at CMS*, *JINST* **11** (2016) P01019 [1510.07488].

[85] CMS collaboration, *Performance of reconstruction and identification of* $\tau$ *leptons decaying to hadrons and* $\nu_\tau$ *in pp collisions at* $\sqrt{s}$ *= 13 TeV*, *JINST* **13** (2018) P10005 [1809.02816].

[86] M. Cacciari, G.P. Salam and G. Soyez, *The anti-$k_t$ jet clustering algorithm*, *JHEP* **04** (2008) 063 [0802.1189].

[87] CMS collaboration, *Reconstruction of decays to merged photons using end-to-end deep learning with domain continuation in the CMS detector*, 2204.12313.

[88] J. Pata, J. Duarte, J.-R. Vlimant, M. Pierini and M. Spiropulu, *MLPF: Efficient machine-learned particle-flow reconstruction using graph neural networks*, *Eur. Phys. J. C* **81** (2021) 381 [2101.08578].

[89] C. Olah, A. Mordvintsev and L. Schubert, *Feature visualization*, *Distill* (2017) .

[90] Y. Lecun, L. Bottou, Y. Bengio and P. Haffner, *Gradient-based learning applied to document recognition*, *Proceedings of the IEEE* **86** (1998) 2278.

[91] A. Krizhevsky, I. Sutskever and G.E. Hinton, *Imagenet classification with deep convolutional neural networks*, in *Advances in Neural Information Processing Systems*, F. Pereira, C. Burges, L. Bottou and K. Weinberger, eds., vol. 25, Curran Associates, Inc., 2012, https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf.

[92] K. Simonyan and A. Zisserman, *Very deep convolutional networks for large-scale image recognition*, *arXiv preprint arXiv:1409.1556* (2014) .

[93] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov et al., *Going deeper with convolutions*, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9, 2015.

[94] K. He, X. Zhang, S. Ren and J. Sun, *Deep residual learning for image recognition*, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[95] G. Huang, Z. Liu, L. Van Der Maaten and K.Q. Weinberger, *Densely connected convolutional networks*, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700–4708, 2017.

[96] M. Tan and Q. Le, *Efficientnet: Rethinking model scaling for convolutional neural networks*, in *International conference on machine learning*, pp. 6105–6114, PMLR, 2019.

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

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

[99] M. Kagan, *Image-Based Jet Analysis*, 2012.09719.

[100] A. Butter et al., *The Machine Learning landscape of top taggers*, *SciPost Phys.* **7** (2019) 014 [1902.09914].

[101] ATLAS collaboration, *Quark versus Gluon Jet Tagging Using Jet Images with the ATLAS Detector, ATL-PHYS-PUB-2017-017*, 7, 2017.

[102] CMS collaboration, *Identification of heavy, energetic, hadronically decaying particles using machine-learning techniques*, *JINST* **15** (2020) P06005 [2004.08262].

[103] KM3NeT collaboration, *Event reconstruction for KM3NeT/ORCA using convolutional neural networks*, *JINST* **15** (2020) P10005 [2004.08254].

[104] J. Collado, K. Bauer, E. Witkowski, T. Faucett, D. Whiteson and P. Baldi, *Learning to isolate muons*, *JHEP* **21** (2020) 200 [2102.02278].

[105] J. Collado, J.N. Howard, T. Faucett, T. Tong, P. Baldi and D. Whiteson, *Learning to identify electrons*, *Phys. Rev. D* **103** (2021) 116028 [2011.01984].

[106] R. Abbasi et al., *A Convolutional Neural Network based Cascade Reconstruction for the IceCube Neutrino Observatory*, *JINST* **16** (2021) P07041 [2101.11589].

[107] M. Paganini, L. de Oliveira and B. Nachman, *CaloGAN : Simulating 3D high energy particle showers in multilayer electromagnetic calorimeters with generative adversarial networks*, *Phys. Rev. D* **97** (2018) 014021 [1712.10321].

[108] G.R. Khattak, S. Vallecorsa, F. Carminati and G.M. Khan, *Fast simulation of a high granularity calorimeter by generative adversarial networks*, *Eur. Phys. J. C* **82** (2022) 386 [2109.07388].

[109] E. Buhmann, S. Diefenbacher, D. Hundhausen, G. Kasieczka, W. Korcari, E. Eren et al., *Hadrons, better, faster, stronger*, *Mach. Learn. Sci. Tech.* **3** (2022) 025014 [2112.09709].

[110] T. Dozat, *Incorporating Nesterov Momentum into Adam*, in *Proceedings of the 4th International Conference on Learning Representations*, pp. 1–4.

[111] T.-Y. Lin, P. Goyal, R. Girshick, K. He and P. Dollár, *Focal loss for dense object detection*, in *Proceedings of the IEEE international conference on computer vision*, pp. 2980–2988, 2017.

[112] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro et al., *TensorFlow: Large-scale machine learning on heterogeneous systems*, 2015.

[113] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan et al., *Pytorch: An imperative style, high-performance deep learning library*, in *Advances in Neural Information Processing Systems 32*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox and R. Garnett, eds., pp. 8024–8035, Curran Associates, Inc. (2019), http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.

[114] D.P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, *arXiv preprint arXiv:1412.6980* (2014) .

[115] M. Wang and W. Deng, *Deep visual domain adaptation: A survey*, *Neurocomputing* **312** (2018) 135.

[116] Y. Ganin and V. Lempitsky, *Unsupervised domain adaptation by backpropagation*, in *International conference on machine learning*, pp. 1180–1189, PMLR, 2015.

[117] G. Louppe, M. Kagan and K. Cranmer, *Learning to Pivot with Adversarial Networks*, 1611.01046.

[118] CMS collaboration, *A deep neural network to search for new long-lived particles decaying to jets*, *Mach. Learn. Sci. Tech.* **1** (2020) 035012 [1912.12238].

[119] B. Denby, *Neural networks and cellular automata in experimental high energy physics*, *Computer Physics Communications* **49** (1988) 429.

[120] P.T. Komiske, E.M. Metodiev and J. Thaler, *Energy Flow Networks: Deep Sets for Particle Jets*, *JHEP* **01** (2019) 121 [1810.05165].

[121] R.T. de Lima, *Sequence-based Machine Learning Models in Jet Physics*, 2102.06128.

[122] S. Thais, P. Calafiura, G. Chachamis, G. DeZoort, J. Duarte, S. Ganguly et al., *Graph Neural Networks in Particle Physics: Implementations, Innovations, and Challenges*, in *2022 Snowmass Summer Study*, 3, 2022 [2203.12852].

[123] J.M. Munoz, I. Batatia and C. Ortner, *BIP: Boost Invariant Polynomials for Efficient Jet Tagging*, 2207.08272.

[124] F.A. Dreyer and H. Qu, *Jet tagging in the Lund plane with graph networks*, *JHEP* **03** (2021) 052 [2012.08526].

[125] P. Baldi, P. Sadowski and D. Whiteson, *Deep Learning From Four Vectors*, 2203.03067.

[126] A. Bogatskiy et al., *Symmetry Group Equivariant Architectures for Physics*, in *2022 Snowmass Summer Study*, 3, 2022 [2203.06153].

[127] C. Shimmin, *Particle Convolution for High Energy Physics*, 7, 2021 [2107.02908].

[128] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez et al., *Attention is all you need*, *Advances in neural information processing systems* **30** (2017) .

[129] M. Phuong and M. Hutter, *Formal algorithms for transformers*, *arXiv preprint arXiv:2207.09238* (2022) .

[130] V. Mikuni and F. Canelli, *ABCNet: An attention-based method for particle tagging*, *Eur. Phys. J. Plus* **135** (2020) 463 [2001.05311].

[131] V. Mikuni and F. Canelli, *Point cloud transformers applied to collider physics*, *Mach. Learn. Sci. Tech.* **2** (2021) 035027 [2102.05073].

[132] H. Qu, C. Li and S. Qian, *Particle Transformer for Jet Tagging*, 2202.03772.

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

[134] P. Xu, X. Zhu and D.A. Clifton, *Multimodal learning with transformers: A survey*, *arXiv preprint arXiv:2206.06488* (2022) .

[135] H. Qu and L. Gouskos, *ParticleNet: Jet Tagging via Particle Clouds*, *Phys. Rev. D* **101** (2020) 056019 [1902.08570].

[136] CMS collaboration, *Search for nonresonant pair production of highly energetic Higgs bosons decaying to bottom quarks*, 2205.06667.

[137] P.-C. Chen, H. Tsai, S. Bhojanapalli, H.W. Chung, Y.-W. Chang and C.-S. Ferng, *Demystifying the better performance of position encoding variants for transformer*, *arXiv preprint arXiv:2104.08698* (2021) .

[138] J. Hoffmann, S. Borgeaud, A. Mensch, E. Buchatskaya, T. Cai, E. Rutherford et al., *Training compute-optimal large language models*, *arXiv preprint arXiv:2203.15556* (2022) .

[139] L. Liu, H. Jiang, P. He, W. Chen, X. Liu, J. Gao et al., *On the variance of the adaptive learning rate and beyond*, *arXiv preprint arXiv:1908.03265* (2019) .

[140] D. Hendrycks and K. Gimpel, *Gaussian error linear units (gelus)*, *arXiv preprint arXiv:1606.08415* (2016) .

[141] A.V. Gritsan, R. Röntsch, M. Schulze and M. Xiao, *Constraining anomalous Higgs boson couplings to the heavy flavor fermions using matrix element techniques*, *Phys. Rev. D* **94** (2016) 055023 [1606.03107].

[142] E. Accomando et al., *Workshop on CP Studies and Non-Standard Higgs Physics*, hep-ph/0608079.

[143] P.D. Mannheim, *Extension of the Goldstone and the Englert-Brout-Higgs mechanisms to non-Hermitian theories*, 9, 2021 [2109.08714].

[144] A.Y. Korchin and V.A. Kovalchuk, *Decay of the Higgs boson to $\tau^-\tau^+$ and non-Hermiticy of the Yukawa interaction*, *Phys. Rev. D* **94** (2016) 076003 [1607.02827].

[145] A.Y. Korchin and V.A. Kovalchuk, *Decay of the Higgs Boson $h \to \tau^-\tau^+ \to \pi^-\nu_\tau\pi^+\bar{\nu}_\tau$ for a Non-Hermitian Yukawa Interaction*, *Acta Phys. Polon. B* **53** (2022) 2 [2103.13665].

[146] S. Berge, W. Bernreuther and S. Kirchner, *Determination of the Higgs CP-mixing angle in the tau decay channels*, *Nucl. Part. Phys. Proc.* **273-275** (2016) 841 [1410.6362].

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

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

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

[150] E. Bagnaschi, G. Degrassi, P. Slavich and A. Vicini, *Higgs production via gluon fusion in the POWHEG approach in the SM and in the MSSM*, *JHEP* **02** (2012) 088 [1111.2854].

[151] P. Nason and C. Oleari, *NLO Higgs boson production via vector-boson fusion matched with shower in POWHEG*, *JHEP* **02** (2010) 037 [0911.5299].

[152] T. Ježo and P. Nason, *On the Treatment of Resonances in Next-to-Leading Order Calculations Matched to a Parton Shower*, *JHEP* **12** (2015) 065 [1509.09071].

[153] F. Granata, J.M. Lindert, C. Oleari and S. Pozzorini, *NLO QCD+EW predictions for HV and HV +jet production including parton-shower effects*, *JHEP* **09** (2017) 012 [1706.03522].

[154] K. Hamilton, P. Nason, E. Re and G. Zanderighi, *NNLOPS simulation of Higgs boson production*, *JHEP* **10** (2013) 222 [1309.0017].

[155] K. Hamilton, P. Nason and G. Zanderighi, *Finite quark-mass effects in the NNLOPS POWHEG+MiNLO Higgs generator*, *JHEP* **05** (2015) 140 [1501.04637].

[156] T. Sjöstrand, S. Ask, J.R. Christiansen, R. Corke, N. Desai, P. Ilten et al., *An introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* **191** (2015) 159 [1410.3012].

[157] T. Przedzinski, E. Richter-Was and Z. Was, *Documentation of TauSpinner algorithms: program for simulating spin effects in $\tau$ -lepton production at LHC*, *Eur. Phys. J. C* **79** (2019) 91 [1802.05459].

[158] NNPDF collaboration, *Parton distributions for the LHC Run II*, *JHEP* **04** (2015) 040 [1410.8849].

[159] NNPDF collaboration, *Parton distributions from high-precision collider data*, *Eur. Phys. J. C* **77** (2017) 663 [1706.00428].

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

[161] J. Alwall et al., *Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions*, *Eur. Phys. J. C* **53** (2008) 473 [0706.2569].

[162] E. Re, *Single-top Wt-channel production matched with parton showers using the POWHEG method*, *Eur. Phys. J. C* **71** (2011) 1547 [1009.2450].

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

[164] S. Alioli, S.-O. Moch and P. Uwer, *Hadronic top-quark pair-production with one jet and parton showering*, *JHEP* **01** (2012) 137 [1110.5251].

[165] CMS collaboration, *Event generator tunes obtained from underlying event and multiparton scattering measurements*, *Eur. Phys. J. C* **76** (2016) 155 [1512.00815].

[166] CMS collaboration, *Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements*, *Eur. Phys. J. C* **80** (2020) 4 [1903.12179].

[167] GEANT4 collaboration, *GEANT4–a simulation toolkit*, *Nucl. Instrum. Meth. A* **506** (2003) 250.

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

[169] CMS collaboration, *Deep learning techniques for energy clustering in the CMS ECAL*, in *20th International Workshop on Advanced Computing and Analysis Techniques in Physics Research: AI Decoded - Towards Sustainable, Diverse, Performant and Effective Scientific Computing*, 4, 2022 [2204.10277].

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

[171] L. Bianchini, J. Conway, E.K. Friis and C. Veelken, *Reconstruction of the Higgs mass in* $H \to \tau\tau$ *Events by Dynamical Likelihood techniques*, *J. Phys. Conf. Ser.* **513** (2014) 022035.

[172] CMS collaboration, "Identification of hadronic tau decay channels using multivariate analysis (MVA decay mode)." https://cds.cern.ch/record/2727092, 2020.

[173] M. Cacciari, G.P. Salam and G. Soyez, *FastJet User Manual*, *Eur. Phys. J. C* **72** (2012) 1896 [1111.6097].

[174] S.D. Ellis and D.E. Soper, *Successive combination jet algorithm for hadron collisions*, *Phys. Rev. D* **48** (1993) 3160 [hep-ph/9305266].

[175] M. Wobisch and T. Wengler, *Hadronization corrections to jet cross-sections in deep inelastic scattering*, in *Workshop on Monte Carlo Generators for HERA Physics (Plenary Starting Meeting)*, pp. 270–279, 4, 1998 [hep-ph/9907280].

[176] CMS collaboration, "Pileup Removal Algorithms, CMS-PAS-JME-14-001." https://cds.cern.ch/record/1751454, 2014.

[177] CMS collaboration, *Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV*, *JINST* **12** (2017) P02014 [1607.03663].

[178] CMS collaboration, *Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV*, *JINST* **13** (2018) P05011 [1712.07158].

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

[180] D. Bertolini, P. Harris, M. Low and N. Tran, *Pileup Per Particle Identification*, *JHEP* **10** (2014) 059 [1407.6013].

[181] K. Rose, *Deterministic annealing for clustering, compression, classification, regression, and related optimization problems*, *Proceedings of the IEEE* **86** (1998) 2210.

[182] R. Fruhwirth, W. Waltenberger and P. Vanlaer, *Adaptive vertex fitting*, *J. Phys. G* **34** (2007) N343.

[183] S. Berge and W. Bernreuther, *Determining the CP parity of Higgs bosons at the LHC in the tau to 1-prong decay channels*, *Phys. Lett. B* **671** (2009) 470 [0812.1910].

[184] S. Berge, W. Bernreuther, B. Niepelt and H. Spiesberger, *How to pin down the CP quantum numbers of a Higgs boson in its tau decays at the LHC*, *Phys. Rev. D* **84** (2011) 116003 [1108.0670].

[185] S. Berge, W. Bernreuther and S. Kirchner, *Determination of the Higgs CP-mixing angle in the tau decay channels at the LHC including the Drell–Yan background*, *Eur. Phys. J. C* **74** (2014) 3164 [1408.0798].

[186] CMS collaboration, *Analysis of the CP structure of the Yukawa coupling between the Higgs boson and $\tau$ leptons in proton-proton collisions at $\sqrt{s}$ = 13 TeV*, *JHEP* **06** (2022) 012 [2110.04836].

[187] CMS collaboration, *An embedding technique to determine $\tau\tau$ backgrounds in proton-proton collision data*, *JINST* **14** (2019) P06032 [1903.01216].

[188] CMS collaboration, *Measurement of the $Z\gamma^* \to \tau\tau$ cross section in pp collisions at $\sqrt{s}$ = 13 TeV and validation of $\tau$ lepton analysis techniques*, *Eur. Phys. J. C* **78** (2018) 708 [1801.03535].

[189] CMS collaboration, *Search for additional neutral MSSM Higgs bosons in the $\tau\tau$ final state in proton-proton collisions at $\sqrt{s}$ = 13 TeV*, *JHEP* **09** (2018) 007 [1803.06553].

[190] CMS collaboration, *Measurements of Inclusive W and Z Cross Sections in pp Collisions at $\sqrt{s}$ = 7 TeV*, *JHEP* **01** (2011) 080 [1012.2466].

[191] CMS collaboration, *Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s}$ = 13 TeV*, *JINST* **15** (2020) P10017 [2006.10165].

[192] J.S. Conway, *Incorporating Nuisance Parameters in Likelihoods for Multisource Spectra*, in *PHYSTAT 2011*, pp. 115–120, 2011, DOI [1103.0354].

[193] ATLAS, CMS, LHC Higgs Combination Group collaboration, *Procedure for the LHC Higgs boson search combination in Summer 2011*, .

[194] CMS collaboration, *Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV*, *Eur. Phys. J. C* **75** (2015) 212 [1412.8662].

[195] "Combine statistical toolkit." https://cms-analysis.github.io/HiggsAnalysis-CombinedLimit/.

[196] G. Cowan, K. Cranmer, E. Gross and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, *Eur. Phys. J. C* **71** (2011) 1554 [1007.1727].

[197] A. Cardini, *Measurement of the CP properties of the Higgs boson in its decays to $\tau$ leptons with the CMS experiment*, Ph.D. thesis, Hamburg U., Universität Hamburg, Hamburg U., 2021. 10.3204/PUBDB-2021-03550.

[198] CMS collaboration, *CMS Luminosity Measurements for the 2016 Data Taking Period, CMS-PAS-LUM-17-001*, 2017.

[199] CMS collaboration, *CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s}$ = 13 TeV, CMS-PAS-LUM-17-004*, 2018.

[200] CMS collaboration, *CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s}$ = 13 TeV, CMS-PAS-LUM-18-002*, 2019.

[201] CMS collaboration, *Measurement of the WZ production cross section in pp collisions at $\sqrt{s}$ = 13 TeV*, *Phys. Lett. B* **766** (2017) 268 [1607.06943].

[202] CMS collaboration, *Cross section measurement of t-channel single top quark production in pp collisions at $\sqrt{s}$ = 13 TeV*, *Phys. Lett. B* **772** (2017) 752 [1610.00678].

[203] R.D. Cousins, *Generalization of chisquare goodness-offit test for binned data using saturated models , with application to histograms*, 2013.

[204] CMS collaboration, *Analysis of the CP structure of the Yukawa coupling between the Higgs boson and $\tau$ leptons in proton-proton collisions at $\sqrt{s}$ = 13 TeV, CMS-PAS-HIG-20-006*, 2020.

[205] V. Cherepanov and A. Zotz, *Kinematic reconstruction of $Z/H \to \tau\tau$ decay in proton-proton collisions*, 1805.06988.

[206] M. Davier, L. Duflot, F. Le Diberder and A. Rouge, *The Optimal method for the measurement of tau polarization*, *Phys. Lett. B* **306** (1993) 411.

[207] J.H. Kuhn and E. Mirkes, *Structure functions in tau decays*, *Z. Phys. C* **56** (1992) 661.

[208] S. Jadach, J.H. Kuhn and Z. Was, *TAUOLA: A Library of Monte Carlo programs to simulate decays of polarized tau leptons*, *Comput. Phys. Commun.* **64** (1990) 275.

[209] M. Jezabek, Z. Was, S. Jadach and J.H. Kuhn, *The tau decay library TAUOLA, update with exact O(alpha) QED corrections in tau —> mu. (e) neutrino anti-neutrino decay modes*, *Comput. Phys. Commun.* **70** (1992) 69.

[210] S. Jadach, Z. Was, R. Decker and J.H. Kuhn, *The tau decay library TAUOLA: Version 2.4*, *Comput. Phys. Commun.* **76** (1993) 361.

[211] CMS collaboration, *Measurements of Higgs boson production in the decay channel with a pair of $\tau$ leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV*, 2204.12957.

[212] ATLAS collaboration, *Measuring CP properties of Higgs boson interactions with $\tau$ leptons with the ATLAS detector*, ATLAS-CONF-2022-032.

[213] "Variables declaration in the TauMLTools framework." https://github.com/cms-tau-pog/TauMLTools/blob/b063593eb2753b75f7945ea27a04ed4e7587f7e0/Analysis/interface/TauTuple.h.

[214] "Feature preprocessing module of the tatoo framework." https://github.com/ofivite/tatoo/blob/4a97d728a11e8ec3eb36c3e0c1dbe4b669c0f92d/utils/data_preprocessing.py.

# Acknowledgements

I should be honest: the time of my PhD has not been an easy ride. Now I am thinking that maybe I should have thought several times before committing to such an experience if I had known how this would unfold. But luckily I did not, so for me, this was a step into the unknown with an open heart and no expectations whatsoever.

The only metaphor of this time which persistently swirls in my mind is the one of light shining through the darkness. I guess it is the reflection of my personality: as my mother told me once, "you are a careful pessimist". If I were to make a metaphoric image of this world, it would be a calm grey water surface at twilight with light ripples fading away; when it is not clear if it is a sunset or a sunrise, if it calls for a storm or contemplates its end.

It was not really the PhD studies which added up to the dark component of this (very naive and misleading!) dual depiction of reality. Of course, pushing the analysis further and trying to understand its nuances was not a piece of cake – rather a marathon towards deeper satisfaction of the Eureka moment. It was more about the world turning upside down and downside up back again. It is probably its natural state, but in any case, it felt hard to recover from the loss of ground and push myself away from dissolving in the dark. The harmony felt so distant as it was stormy all along.

Despite this, I would like to acknowledge the whole experience of my time in Hamburg: as it was, with all the ups and downs, struggles, suffering, moments of enlightenment, excitement, and enthusiasm. I view the world as a vibration, and retrospectively I think the rhythm of these times has been unique and special to be experienced.

This experience would be empty and lacking depth without all the wonderful people with whom I have had a chance to share it. I would certainly say that they have been one of the sources of light in this metaphorical darkness. Thank you so much:

- Alexei Raspereza & Elisabetta Gallo for babysitting me, giving space for exploration and supporting my growth as a researcher. I cannot think of the supervisor-student relationship differently than those of parents and kids. Of course, I exaggerate a bit here – but still, the PhD step for me is reminiscent of late adolescence: there is little babysitting involved, which is why it is easy to get lost as one is establishing their research values. That is why I am very thankful for your guidance and wisdom as I was navigating my way through the scientific landscape. I definitely could not ask for better supervisors.
- My parents Viktor & Tatiana and brother Andrey for unconditional support and love. I realised this during the past years, as I was growing older, what it is actually like to be a parent (in particular of myself) and what it actually means to love someone. There is enormous work behind it, but at its core, it goes easily and naturally from the depth of the

heart. You helped me a lot when I felt broken and scared, and I am deeply grateful for all your kind words and advice.

- DeepTau team, in particular Konstantin and Mykyta. I remember the (absolutely random) moment when I joined the team, and I love such turning points. There is no decisive moment *per se*, only following the flow as it leads you where one should be. It has been a pleasure to work with you as we were developing DeepTau v2.5 and beyond. It was the time when I learned so much about how things should be written and designed in code. Moreover, the leadership style and the expertise standards of Konstantin are something which I personally admire, and I am grateful for the chance to learn it from you.

- All the nice DESY people with whom I worked together, in particular Aliya, Andrea, Daina, Mareike, Maryam, Sam, Sandra, Teresa, Valeria, and Yiwen. It is the people who create the environment, and I am very happy that I was a part of the welcoming and supportive DESY research space. I think it is crucial for every researcher to have peers to share and exchange ideas with. And so it is good that at DESY we have got ideas flowing and new interactions being explored.

- My close friends: Alyona, Bettina, Denis, Dima, Eldar, Gosha, Leo, Michał, Nastya, Valentina, Vlad (1), Vlad (2), Vova, Stepan, Yana. I am happy that with you I could not care less about being serious: it was in the natural order of things to make ridiculous (meta)ironic jokes, form a self-proclaimed beer party, ride a bike with a bunch of balloons, and in general behave like kids. And yet we could sit on the floor of our WG kitchen and talk about the depth of the world, or discuss the implications of ancient Greek philosophy, or argue what it is to be a scientist. We were exploring together the landscapes of music in the most intimate ways (i.e. jamming with each other) and the landscapes of the world through the lenses of film cameras (with occasional stops in craft beer shops). We were drinking wine on Elbstrand and talking about art or were slurping lots of coffee in Surf to dial in that sweat spot. You are all deeply kind and loving in your heart, and this is the precious thing I cherish.

It is funny that after writing this, I truly realised how lucky I am to have such people close by. There is a saying that it is easy to overlook and not notice the value of things one possesses until they are lost. In that respect, I would like to finally acknowledge these acknowledgements for an opportunity to contemplate what I have experienced during the years of my PhD and what I value as important to stick to before it is lost.

In the end, I would like to conclude this work with a reminder for my future self and a curious reader: there is a bright side, and it shines within oneself.

~~~
~~~