
Machine learning approach to $H \rightarrow \tau\tau$ analysis in the CMS experiment

MICHAŁ OLSZEWSKI



Doctoral dissertation prepared under the supervision of
dr hab. Marcin Konecki and **dr Michał Bluj**
at the Institute of Experimental Physics, Faculty of Physics
University of Warsaw

March 15, 2020

Abstract

The ATLAS and CMS experiments at the LHC discovered a 125 GeV Higgs boson in 2012. The measurements of its properties give us an insight to many important physical parameters. In this thesis the CMS $H \rightarrow \tau\tau \rightarrow \mu\nu_\mu\nu_\tau\tau_h\nu_\tau$ analysis is discussed, including the Higgs physics theory, construction of the CMS apparatus and event reconstruction algorithms. Moreover, a set of machine learning methods is presented together with their utilization in event identification. The construction of these methods allows them to be easily accommodated to official Monte Carlo CMS data samples in the supervised learning mode. We showed that the result of 0.875 for average area under receiver operating curves for all considered event final states is feasible. The best performance is obtained for neural network model.

The thesis contains the author analysis of significance implemented on top of both cut based and machine learning based distributions of discriminating variables in the aforementioned Higgs boson decay channel. The results indicate that the incorporation of the output of machine learning estimator can boost the performance of the analysis by 20 to 35%.

Streszczenie

W 2012 roku eksperymenty ATLAS i CMS ogłosiły odkrycie nowej cząstki skalarniej o masie około 125 GeV - bozonu Higgsa. Pomiary własności tej cząstki pozwalają na oszacowanie wielu ważnych parametrów fizycznych Modelu Standardowego. W szczególności sprzężenia bosonu Higgsa do leptonów mogą być mierzone w rozpadach na dwa taony. W niniejszej pracy omówiona jest analiza $H \rightarrow \tau\tau \rightarrow \mu\nu_\mu\nu_\tau\tau\nu_\tau$: przedstawiono podstawy teoretyczne Modelu Standardowego i spontanicznego łamania symetrii wraz z fizyką cząstki Higgsa, budowę detektora CMS przy akceleratorze LHC oraz algorytmy rekonstrukcji zderzeń. Zaprezentowano ponadto przegląd metod uczenia maszynowego razem z ich zastosowaniem do identyfikacji przypadków w omawianej analizie. Wygoda użycia tych metod wynika z faktu łatwego ich trenowania przy użyciu standardowych próbek Monte Carlo dostępnych dla analiz fizycznych kolaboracji CMS. Uzyskane wyniki wskazują na możliwość uzyskania wartości 0.875 dla uśrednionej powierzchni pod krzywymi ROC (*ang.* „Receiver Operating Curve”, charakterystyka operacyjna odbionika) dla rozpoznawania czternastu zdefiniowanych stanów końcowych istotnych w analizie $H \rightarrow \tau\tau$.

Praca kończy się autorską analizą znaczości zaimplementowaną dla wyżej wspomnianego kanału rozpadu bozonu Higgsa. Porównano tutaj rozkłady zmiennej dyskryminującej dla standardowej analizy opartej o cięcia kinematyczne oraz analizy opartej o wynik z modelu sieci neuronowej. Wykazano, że uczenie maszynowe pozwala na poprawę wyniku końcowego analizy o 20% do 35% .

Acknowledgements

I would like to thank All, which in any way helped me in scientific work and have contributed to creation of this thesis.

I am very grateful to my supervisor dr hab. Marcin Konecki and auxiliary supervisor dr Michał Bluj for their care, numerous discussions, sacrificed time, precious comments and observations which helped me to obtain many results needed to write this thesis. I also want to thank other members of the Warsaw CMS Group, especially Małgorzata Kazana, Krzysztof Doroba, Piotr Zalewski, Karol Buńkowski and Artur Kalinowski for their guidance, fruitful ideas, collaboration, support and always relevant comments. I would like to thank my brothers in arms at PhD studies: Tomasz Früboes for his help in solving numerous computing infrastructure troubles; Maciej Misiura, Łukasz Obara, Marek Walczak and Andrzej Pyskir for their ideas and honest conversations.

Special thanks to Valeria Botta and Teresa Lenz (DESY), Laura Margaret Dodd and Tyler Henry Ruggles (University of Wisconsin Madison) and Andrew Gilbert (Northwestern University). It's only their vast knowledge and selfless help allowed me to finish Chapter 4 of this thesis.

I want to thank the most keen and hard-working researcher that I ever came across Michail Bachtis (University of California Los Angeles) for a lot of positive inspiration and for showing me why and how to use Python language in HEP.

I could not forget to thank Krystian Zawistowski for his enormous help, long and fruitful conversations and collaboration in developing the machine learning part of this thesis.

A warm thanks go to Ms Anna Kaczor for patience, understanding and dedication.

At last but not at least, I would like to dearly thank my family - Zosia and Nel - for your endless love, understanding and assistance through the duration of my studies.

Contents

List of Abbreviations	ix
Introduction	1
1 Introduction to Higgs boson and tau lepton physics	3
1.1 The Standard Model	3
1.2 Symmetry breaking and the Brout-Englert-Higgs mechanism	5
1.3 The Higgs boson and its properties in the Standard Model	6
1.3.1 Higgs couplings	7
1.3.2 Higgs boson production at the LHC	8
1.3.3 Higgs decay modes	11
1.4 Tau lepton	14
2 The LHC and CMS detector	17
2.1 Large Hadron Collider	17
2.2 Compact Muon Solenoid detector	20
2.2.1 Superconducting magnet	20
2.2.2 Tracker	21
2.2.3 Calorimeters	22
2.2.4 Muon system	25
2.2.5 Trigger system	26
3 Offline object reconstruction	31
3.1 Particle Flow algorithm	32
3.2 Jet reconstruction	37
3.3 Tau reconstruction	39
3.4 Higgs mass reconstruction in decays into a pair of taus	41
4 The Higgs boson analysis in the ditau decay channel	43
4.1 Samples	45
4.2 Data pre-processing	46
4.3 Analysis categories	53
4.4 Simulated samples corrections	54
4.5 Simulated samples normalization	57
4.6 Background estimation	58
4.7 Signal variables	59
4.8 Analysis cross check	60
4.9 Official CMS analysis results	64
5 Event identification using Machine Learning techniques	67
5.1 Initial analysis setup and data preprocessing	69
5.2 Decision Tree and ML performance measurements	71
5.2.1 Confusion matrix	75

5.2.2	ROC and ROC AUC	77
5.2.3	DT model results	79
5.3	Random Forest	81
5.4	Gradient boosted decision trees	84
5.5	Deep learning	87
6	Machine Learning application into the $H \rightarrow \tau\tau$ analysis	93
6.1	Setup	93
6.2	An insight into machine learning approach	94
6.3	Systematic uncertainties	97
6.4	Results	99
	Conclusions	101
	Bibliography	103
	A Datasets	115
	B Technical Appendix for Chapter 4	121
	C Control plots	125
	D Features list	127
	E Statistics	129

List of Abbreviations

AI	Artificial Intelligence
API	Application Programming Interface
ASIC	Application-Specific Integrated Circuit
ATLAS	A Toroidal LHC ApparatuS
AUC	Area Under Curve
BDT	Boosted Decision Tree
BEH	Brout-Englert-Higgs mechanism
BMTF	Barrel Muon Track Finder
BR	Branching Ratio
CART	Classification And Regression Trees
CDF	Cumulative Distribution Function
CERN	European Organization for Nuclear Research
CMS	Compact Muon Solenoid
CMSSW	CMS SoftWare (application framework)
CRAB	CMS Remote Analysis Builder
CSC	Cathode Strip Chamber
CTF	Combinatorial Track Finder
DAS	Data Aggregation Service
DC	Data Certification
DIS	Deep Inelastic Scattering
DL	Deep Learning
DNN	Deep Neural Network
DQM	Data Quality Monitoring
DT	Drift Tube (Chapter 2) or Decision Tree (Chapter 5)
DY	Drell–Yan
ECAL	Electromagnetic CALorimeter
ECDF	Empirical Cumulative Distribution Function
EDM	Event Data Model
EMTF	Endcap Muon Track Finder
FN(R)	False Negative (Rate)
FP(R)	False Positive (Rate)
FPGA	Field-Programmable Gate Array
GGF/ggH	Gluon-Gluon Fusion
GMT	Global Muon Trigger
GPU	Graphics Processing Unit
GSF	Gaussian Sum Filter
GT	Global Trigger
HB	Hadronic Barrel calorimeter
HCAL	Hadronic CALorimeter
HE	Hadronic Endcap calorimeter
HEP	High Energy Physics
HF	Hadronic Forward calorimeter
HLT	High Level Trigger

HO	Hadronic Outer calorimeter
HPS	Hadron-Plus-Strip algorithm
IP	Interaction Point
JEC	Jet Energy Correction
L1	Level 1 (trigger)
LEP	Large Electron-Positron collider
LHC	Large Hadron Collider
LO	Leading Order
LS1	Long Shutdown 1 (2013-2015)
LS2	Long Shutdown 2 (2018-2021)
MC	Monte Carlo
MET	Missing Transverse Energy
MLE	Maximum Likelihood Estimator/Estimation
ML	Machine Learning
MP	Misio-Pączek
MSSM	Minimal Supersymmetric Standard Model
MVA	MultiVariate Analysis
N(NLO)	Next-to-(Next-to-)Leading Order
NPU	Number of Primary Vertexes
OMTF	Overlap Muon Track Finder
OS/SS	Opposite/Same Side (about lepton pair charge)
PDF	Parton Distribution Function
PF	Particle Flow (algorithm)
POG	Particle Object Group
PV	Primary Vertex
QCD	Quantum ChromoDynamics
RF	Radio Frequency (Chapter 2) or Random Forest (Chapter 5)
RMS	Root Mean Square
ROC	Receiver Operating Characteristic
RPC	Resistive Plate Chamber
S/B	Signal to Background ratio
SF	Scale Factor
SLAC	Stanford Linear Accelerator Center
SM	Standard Model
SU(n)	Special Unitary group of order n
SV	Secondary Vertex
SVfit	Secondary Vertex fit (algorithm)
TEC	Tracker End Caps
TES	Tau Energy Scale (correction)
TIB	Tracker Inner Barrel
TID	Tracker Inner Disks
TN(R)	True Negative (Rate)
TOB	Tracker Outer Barrel
TP(R)	True Positive (Rate)
U(n)	Unitary group of order n
UN	United Nations
UNESCO	United Nations Educational, Scientific and Cultural Organization
VBF	Vector Boson Fusion
WLCG	Worldwide LHC Computing Grid
WLS	WaveLength-Shifting
WP	Working Point

Introduction

At the time of writing this thesis the Higgs boson is already a well known and confirmed part of the Standard Model. Since its discovery in 2012, the ATLAS (A Toroidal LHC ApparatuS) and CMS (Compact Muon Solenoid) collaborations re-defined their analyses' main objective to the Higgs particle property studies and Beyond Standard Model searches. As part of the former task, for the Run II (2015-2018) of the Large Hadron Collider (LHC), direct coupling measurements into fermions were conducted in the ditau Higgs decay channel.

This channel is characterized by a relatively large Higgs boson branching ratio (BR) of 6%. However, tau lepton decays very rapidly and it's only possible to search for its decay products. On top of that, there are many possible tau decay modes, not excluding the semi-leptonic ones where hadrons (visible as jets) are present in the final state (about 60% of the BR). As a consequence the single $H \rightarrow \tau\tau$ analysis is split into multiple sub-analyses. Moreover, the large hadronic background present at the LHC collision point plays a crucial role, and the invisible neutrinos worsen the mass reconstruction resolution. The signal-to-background ratio is boosted by considering the topology of Higgs boson production processes and applied into analysis with a set of carefully selected constraints on kinematical variables in the event. The above process is called *categorization* and has been the place of major improvements for the Run II (2015-2018) versus the initial Run I (2010–2013) design. In particular, the one dimensional distributions of discriminating variable (Higgs mass) were replaced by two dimensional ones, with observables chosen per category and presenting the best discriminative power between signal and backgrounds. This allows to limit the number of categories as 2D approach does not require designing complex selection criteria based on a variety of variables. Additionally, the remaining categories have similar definitions across final states.

Work on the construction of the analysis (e.g. categorization) may be considered very technical, but it is crucial for enhancing the sensitivity of the measurement and influence the final precision of the result. This is especially important when only a finite amount of physical data is available. Therefore, I have decided to devote this thesis to present yet another approach to the $H \rightarrow \tau\tau$ analysis, where Run II categorization is backed up – or even replaced – by a single classifier, derived using the set of popular machine learning techniques. The aim of this thesis is to find the machine learning classifier which effectively reconstructs the origin of the event and find if and how this classifier can be used for Higgs boson signal extraction. Within that task we do not constrain the target of the classification only to Higgs-like and not-Higgs-like discrimination as a general insight into the event can bring valuable information e.g. for the signal (background) estimation in the control regions defined in the analysis.

The thesis is organized as follows. *Chapter 1* describes the foundations of the Higgs particle physics. The Standard Model is introduced as a field theory with appropriate gauge group symmetry and with spontaneous symmetry breaking. Furthermore the properties of the Higgs boson, i.e. mass, couplings, main production

processes and decay modes are presented. The chapter ends with a brief description of the tau lepton properties.

In *Chapter 2* the Compact Muon Solenoid apparatus is described. The chapter begins with a short presentation of historical background and the reasoning behind the LHC and CMS creation. The rest of the text is devoted to the description of constituents of the CMS: the magnet, inner tracking detector, calorimeters, muon trackers and trigger systems.

The vast majority of analyses in the CMS are based on high level objects (electrons, muons, photons, jets and so on) reconstructed from signals left by particles traversing the matter of the detector. The sophisticated algorithms employed for this task are enumerated and detailed in *Chapter 3*. In the Compact Muon Solenoid experiment, the core algorithm used is called Particle Flow and described first. It uses data from all CMS subsystems and allows to distinguish between muons, electrons, photons and charged/neutral hadrons. The jets are reconstructed on top of the Particle Flow particles and taus are identified from jets. At the end, the Higgs boson mass reconstruction in the ditau events is presented.

The results of this thesis are based upon the cut-based $H \rightarrow \tau\tau$ analysis, and as such it was necessary to have working re-implementation of the standard (and official) CMS analysis. *Chapter 4* follows the path of the 2016 CMS $H \rightarrow \tau\tau$ analysis, all the way from official CMS data samples to the nominal distributions of discriminating variables in this channel. Within that, the analysis-specific variables are defined, the baseline and the category selections are presented and the Monte Carlo to Data samples weights and scale factors are discussed. Moreover, the relevant systematic uncertainties are given.

Chapter 5 explains exactly what *Machine Learning* is and what it has to do with the term *classifier*. In the main part of this chapter a variety of classifiers are presented, trained and compared. The analysis goes from the simplest white box model – decision tree, through newer and more advanced random forest and boosted decision trees, and ends with neural networks as an example of lately worldwide favored approach to machine learning.

The last *Chapter 6* uses the best classifier found earlier to propose its utilization for the Higgs boson events classification. The results are compared to the performance of the standard CMS analysis classification (presented in *Chapter 4*) on the ground of the significance calculation.

The results of the thesis are summarized in *Conclusions*. Five appendices contain a list of data samples used (Appendix A), additional technical information for Chapter 4 (Appendix B), control plots for the $H \rightarrow \tau_\mu \tau_h$ analysis (Appendix C), the input variables for the machine learning models (Appendix D), and derivation of the equation for the significance used in Chapter 6 (Appendix E).

Chapter 1

Introduction to Higgs boson and tau lepton physics

The currently dominant model in high energy physics is called the *Standard Model (SM)* or *Glashow–Weinberg–Salam Model* [1]. SM is a quantum field theory that fully describes three of the four fundamental forces (the only exception is gravity) in the Universe and all elementary particles discovered so far. The Lagrangian of the SM is composed of three parts related to the quark sector (*Quantum chromodynamics (QCD)*), fermion sector (*Electroweak interaction theory*) and Higgs sector. In this chapter we will shortly introduce the basics of the SM, emphasizing the spontaneous electroweak symmetry breaking, Higgs boson properties and tau lepton physics. We will describe the classical approach to the Standard Model, where neutrinos are massless and the lepton flavor mixing effects related to Pontecorvo–Maki–Nakagawa–Sakata matrix are neglected.

1.1 The Standard Model

The Standard Model is described by the Lagrangian density

$$\mathcal{L}_{\text{SM}} = \mathcal{L}_{\text{kinetic+gauge}} + \mathcal{L}_{\text{Yukawa}} + \mathcal{L}_{\text{Higgs}}. \quad (1.1)$$

The above formula describes all particles and their interactions in the Standard Model. Lagrangian in Eq. 1.1 contains all allowed terms dimension 4 or less respecting Lorentz invariance and gauge group

$$\text{SU}(3) \times \text{SU}(2) \times \text{U}(1). \quad (1.2)$$

In accordance to Noether theorem [2] every gauge symmetry corresponds to a conservation of a single physical quantity. In the Standard Model they are called *color*, (*weak*) *isospin* and *hypercharge*, respectively for gauge groups given in Eq. 1.2.

The field content of the Standard Model contains the Higgs field (spin 0), fermionic fields (spin 1/2) and vector gauge fields (spin 1).

The Higgs field ϕ is the only complex scalar field in the Standard Model. It has four degrees of freedom and is described by the following quantum numbers

$$\phi = (1, 2, +1/2) \quad (1.3)$$

In this notation the quantum numbers describe the transformation properties with respect to the gauge group (Eq. 1.2). This means that ϕ transforms as a singlet under SU(3) (it is color neutral and thus does not take part in the strong interactions) and as a doublet under SU(2). The last entry of the triplet is the value of hypercharge.

The three fermion generations (labeled by the subscript i) are described by five representations of the fields and their transformation properties with respect to the gauge group (Eq. 1.2). They are as follows:

- the doublet of left-handed quarks: $Q_{Li} = (3, 2, +1/6)$,
- the singlet of right-handed up quarks: $u_{Ri} = (3, 1, +2/3)$,
- the singlet of right-handed down quarks: $d_{Ri} = (3, 1, -1/3)$,
- the doublet of left-handed leptons: $L_{Li} = (1, 2, -1/2)$,
- the singlet of right-handed leptons: $e_{Ri} = (1, 1, -1)$.

The gauge interactions bosonic degrees of freedom are given by eight SU(3) fields $G_a^\mu = (8, 1, 0)$, three SU(2) fields $W_a^\mu = (1, 3, 0)$ and one U(1) field $B^\mu = (1, 1, 0)$. The first type are called the gluonic fields and are responsible for transmitting color interactions. The two latter ones are combined together into a unified Glashow–Weinberg–Salam theory [3, 4, 5] of the electromagnetic and weak interactions, shortly called the electroweak theory. The field strength tensors are given by

$$\begin{aligned} G^{\mu\nu} &= \partial^\mu G_a^\nu - \partial^\nu G_a^\mu - g_S f_{abc} G_b^\mu G_c^\nu, \\ W^{\mu\nu} &= \partial^\mu W_a^\nu - \partial^\nu W_a^\mu - g_W \varepsilon_{abc} W_b^\mu W_c^\nu, \\ B^{\mu\nu} &= \partial^\mu B^\nu - \partial^\nu B^\mu \end{aligned} \quad (1.4)$$

Here g_S and g_W are the strong SU(3) and the weak SU(2) coupling constants. The U(1) coupling constant will be called g_Y .

To finish the construction of the Standard Model we need to write down the definition of the covariant derivative in terms of the fields G_a^μ , W_a^μ and B^μ :

$$D^\mu = \partial^\mu + ig_S G_a^\mu L_a + ig_W W_b^\mu T_b + ig_Y B^\mu \mathcal{Y}. \quad (1.5)$$

The generators of each group are as follows. The L_a 's are the SU(3) generators, which are described by the 3×3 Gell-Mann matrices $\frac{1}{2}\lambda_a$ for triplets and 0 for singlets. The T_b 's are the SU(2) generators, given by 2×2 Pauli matrices $\frac{1}{2}\sigma_b$ for doublets and 0 for singlets and \mathcal{Y} is the hypercharge.

Let us now expand the Eq. 1.1 and elaborate on the physics behind it.

The first term gives

$$\begin{aligned} \mathcal{L}_{\text{kinetic+gauge}} &= -\frac{1}{4} G_a^{\mu\nu} G_{a\mu\nu} - \frac{1}{4} W_b^{\mu\nu} W_{b\mu\nu} - \frac{1}{4} B^{\mu\nu} B_{\mu\nu} \\ &\quad - i\bar{Q}_{Li} \not{D} Q_{Li} - i\bar{u}_{Ri} \not{D} u_{Ri} - i\bar{d}_{Ri} \not{D} d_{Ri} - i\bar{L}_{Li} \not{D} L_{Li} - i\bar{e}_{Ri} \not{D} e_{Ri} \\ &\quad - (D^\mu \phi)^\dagger (D_\mu \phi) \end{aligned} \quad (1.6)$$

and describes the free propagation of the fields in spacetime, and gauge interactions.

The remaining terms in the Lagrangian contain interactions between appropriate fields. The Yukawa term $\mathcal{L}_{\text{Yukawa}}$ describes the interaction between the Higgs field and fermion fields. The coupling is proportional to the fermion masses [6]

$$\mathcal{L}_{\text{Yukawa}} = Y_{ij}^d \bar{Q}_{Li} \phi d_{Rj} + Y_{ij}^u \bar{Q}_{Li} \tilde{\phi} u_{Rj} + Y_{ij}^e \bar{L}_{Li} \phi e_{Rj} + \text{h.c.} \quad (1.7)$$

In this equation $\tilde{\phi} = i\sigma_2 \phi^\dagger$ is a doublet with the hypercharge $\mathcal{Y}(\tilde{\phi}) = -1$ and the Y 's represent general 3×3 matrices of dimensionless couplings between scalars and

fermions. The couplings depend on the flavor of the particles and this part of the Lagrangian is the only source of CP symmetry violation in the SM¹.

The last term in the Lagrangian is the so-called scalar potential

$$\mathcal{L}_{\text{Higgs}} = -V(\phi) = -\mu^2\phi^\dagger\phi - \lambda(\phi^\dagger\phi)^2, \quad (1.8)$$

where μ and λ are constant scalar coefficients. This part is responsible for symmetry breaking and is the essence of the Brout-Englert-Higgs mechanism [9, 10, 11, 12], which will be described in the following section.

1.2 Symmetry breaking and the Brout-Englert-Higgs mechanism

There are two ways of breaking the exact symmetry of the Lagrangian: explicit and spontaneous. In the former case the Lagrangian does not exhibit symmetry at all. In the latter case the Lagrangian is symmetrical under some transformation, but the ground state of the theory is not. Therefore, it is the nonvanishing vacuum expectation value (vev) of a field present in the theory that breaks the symmetry. The usefulness of the spontaneous symmetry breaking mechanism stems from the fact that thanks to it in the theory particles acquire masses. In the unbroken $SU(2) \times U(1)$ theory of weak and electromagnetic interactions all gauge bosons are massless which is inconsistent with short-range nature of weak forces and clearly massive force carriers. The spontaneous breaking of this continuous local² gauge symmetry is exactly what is needed to give the masses to the weak interaction bosons (via Higgs mechanism), leaving the photon massless (in accordance with Goldstone theorem [13]). Since we want three out of four (the number of generators of symmetry group is $3 + 1$) massless bosons to acquire mass, the symmetry of the ground state should be symmetry with single generator. The broken Lagrangian $U(1)$ symmetry group is associated with electromagnetism with electric charge as its generator.

In the Standard Model the Higgs field has a non-zero vev v

$$\langle 0|\phi(x)|0\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} v \\ 0 \end{pmatrix}.$$

The placement of the real-valued vev in the higher component of the doublet does not cause any loss of generality. In order to get the mass part of the Lagrangian we need to evaluate its kinetic term at the vev. We obtain:

$$\mathcal{L}_{\text{mass}} = -(D^\mu\phi)^\dagger (D_\mu\phi) = -\frac{1}{8}v^2(1,0) \begin{pmatrix} g_W W_\mu^3 - g_Y B_\mu & g_W(W_\mu^1 - iW_\mu^2) \\ g_W(W_\mu^1 + W_\mu^2) & -g_W W_\mu^3 - g_Y B_\mu \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

The diagonalization of this matrix leads to the following formula

$$\mathcal{L}_{\text{mass}} = -M_W^2 W^{\mu+} W_\mu^- - \frac{1}{2}M_Z^2 Z^\mu Z_\mu, \quad (1.9)$$

¹In the extended SM theory with right-handed neutrinos the lepton CP symmetry also can be violated. The T2K experiment reported a 2σ confidence interval for the CP violating phase (δ_{CP}) of PMNS matrix [7] which does not include CP-conservation cases ($\delta_{CP} = 0, \pi$) [8].

²Every local symmetry is also a global one, but not vice versa.

where the fields

$$W_\mu^\pm \equiv \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2); \quad Z_\mu \equiv \cos \theta_W W_\mu^3 - \sin \theta_W B_\mu; \quad \theta_W \equiv \tan^{-1}(g_Y/g_W)$$

are associated with W^\pm and Z^0 particles and their masses are equal

$$M_W = g_W \cdot v/2 \approx 80.4 \text{ GeV}; \quad M_Z = M_W / \cos \theta_W \approx 91.2 \text{ GeV}.$$

and θ_W is the Weinberg mixing angle (electroweak mixing angle). The fourth massless particle, which is ruled out of Eq. 1.9 by the zero element of the Higgs vev spinor, is a photon represented by the field:

$$A_\mu \equiv \sin \theta_W W_\mu^3 + \cos \theta_W B_\mu.$$

In the Standard Model the Higgs field is also present in terms responsible for giving mass of the fermions via the Yukawa interactions (see Eq. 1.7). These interactions explain how the masses of fermions are generated in the Standard Model. Fermions, which are initially massless (before spontaneous symmetry breaking in the so-called gauge interaction eigenstate) acquire a mass, determined by the Yukawa coupling after symmetry breaking (mass eigenstate). The mass term is omitted for the neutrinos, since right-handed neutrinos are not present in the Standard Model³. To obtain the fermion particle masses after the Higgs field gains a vev, we perform a rotation to the fermion mass eigenstate basis. After performing the diagonalization the masses are given by

$$m_f = \frac{Y^f v}{\sqrt{2}} \quad (1.10)$$

Y^f being appropriate parameters of the Lagrangian in Eq. 1.7.

1.3 The Higgs boson and its properties in the Standard Model

From the four initial degrees of freedom of the Higgs field, the W^\pm and Z bosons absorb one degree of freedom each in the Higgs mechanism. Therefore, after spontaneous symmetry breaking only one degree of freedom is left. It corresponds to the physical Higgs boson.

The mass of the Higgs boson (in terms of the parameters of the Eq. 1.8) is

$$m_H = \sqrt{2\lambda}v, \quad (1.11)$$

where λ is the so-called Higgs self-coupling parameter.

The expectation value of the Higgs field must be fixed experimentally by relation to the Fermi coupling constant G_F ⁴, giving the value of v ,

$$v = \left(\sqrt{2}G_F\right)^{-1/2} \sim 246 \text{ GeV}. \quad (1.12)$$

³The SM does not explain how neutrinos obtain masses. One of the explanations which goes beyond the SM is that they are Majorana particles. Another one assumes the presence of additional right-handed (sterile) neutrinos and Yukawa-like mass terms.

⁴...which in turn comes from measurements of the muon lifetime, gives strength of the weak force and is equal to $1.1663787(6) \times 10^{-5} \text{ GeV}^{-2}$

The lack of the prediction of the Higgs measured mass is due to the fact that the λ coefficient is a Standard Model free parameter. Given the Higgs mass approximately equal to 125 GeV, $\lambda \simeq 0.13$. Furthermore, the minimum of the Higgs potential (Eq. 1.8) is for

$$|\phi_0|^2 \equiv v^2 = \frac{-\mu^2}{\lambda},$$

so we have $|\mu| \simeq 88.8$ GeV. An illustration of the Higgs potential and spontaneous symmetry breaking mechanism is shown in Fig. 1.1.

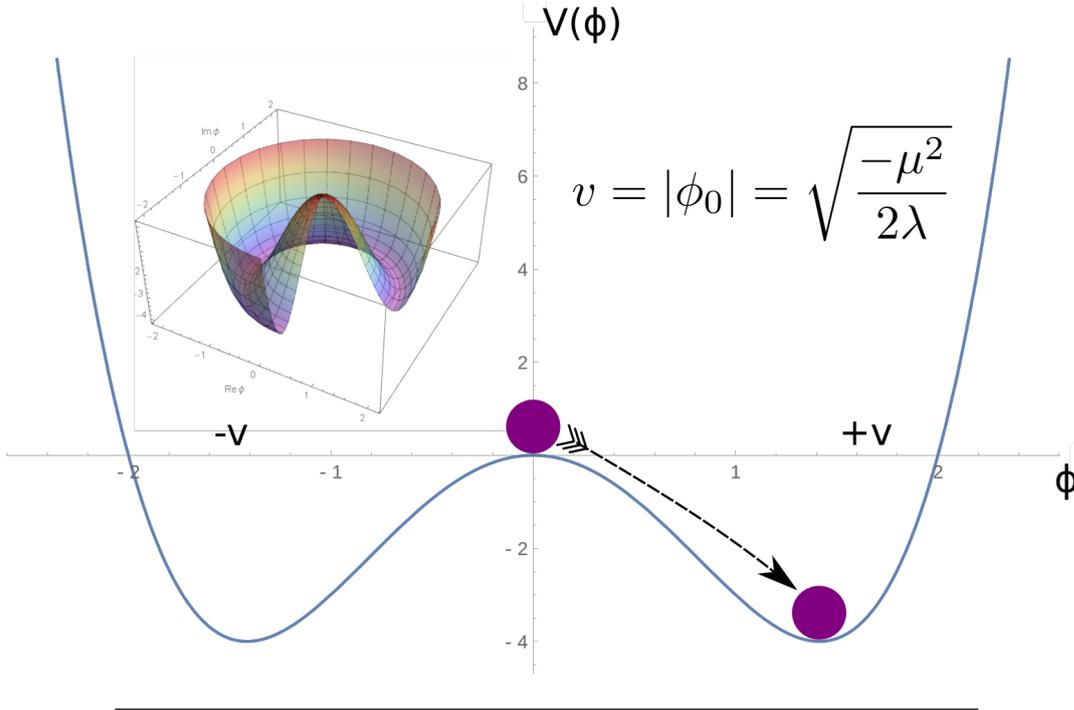


Figure 1.1: Illustration of the Higgs potential (Eq. 1.8) and symmetry breaking mechanism. The main plot shows the projection of the complex scalar potential $V(\phi)$ (top left 3D plot) onto a plane where two (out of an infinite number) of vevs are chosen. The shape of the potential requires the μ^2 constant to be negative, i.e. μ is a purely imaginary number. After the Higgs field ϕ acquires a vev v (the purple ball falls down the slope), the initial symmetry of the potential is broken (the purple ball can no longer move freely and must stay in the groove).

1.3.1 Higgs couplings

We can write down explicitly the full expression describing Higgs boson (self) couplings as

$$\mathcal{L} = -g_{Hff}\bar{f}fH + \frac{g_{HHH}}{6}H^3 + \frac{g_{HHHH}}{24}H^4 + \delta_V V_\mu V^\mu \left(g_{HVV}H + \frac{g_{HHVV}}{2}H^2 \right) \quad (1.13)$$

with $V = W^\pm$ or $V = Z$ and $\delta_W = 1$ and $\delta_Z = 1/2$. In the formula above we have

- $g_{Hff} = \frac{m_f}{v}$ as the Higgs coupling constant to the fermions - it is linearly proportional to the mass of the lepton/quark m_f .
- $g_{HVV} = \frac{2m_V^2}{v}$ as a single Higgs coupling constant to vector bosons. It is proportional to the square of the vector boson mass m_V , so it is much larger than g_{Hff} (for $m_V \approx m_f$).

- $g_{HHVV} = \frac{2m_V^2}{v^2}$ as a double Higgs coupling constant to the vector bosons - also proportional to the square of the vector boson mass, but with squared vev in the denominator.
- $g_{HHH} = \frac{3m_H^2}{v}$ and $g_{HHHH} = \frac{3m_H^2}{v^2}$ - are the Higgs self couplings in a triple and quadrupole vertex.

In the LHC one-loop processes also play an important role, in which a Higgs boson is produced from gluons and decays into photons. In the first case, the loop contains mainly a virtual top-antitop pair. In the second case, the photons are created dominantly via a virtual vector boson loop (top loop contributes at a $\sim 20\%$ level compared to W^\pm at leading order).

1.3.2 Higgs boson production at the LHC

In the LHC the Higgs particle is produced mainly via gluon-gluon fusion (ggH), vector boson fusion (VBF), associated production with a weak-boson and finally associated production with a $t\bar{t}$ quarks pair (see Fig. 1.2). The total production cross section corresponding to these processes depends on (for the fixed Higgs mass of 125 GeV) the center of mass energy, and is equal to 55.1 pb for 13 TeV.

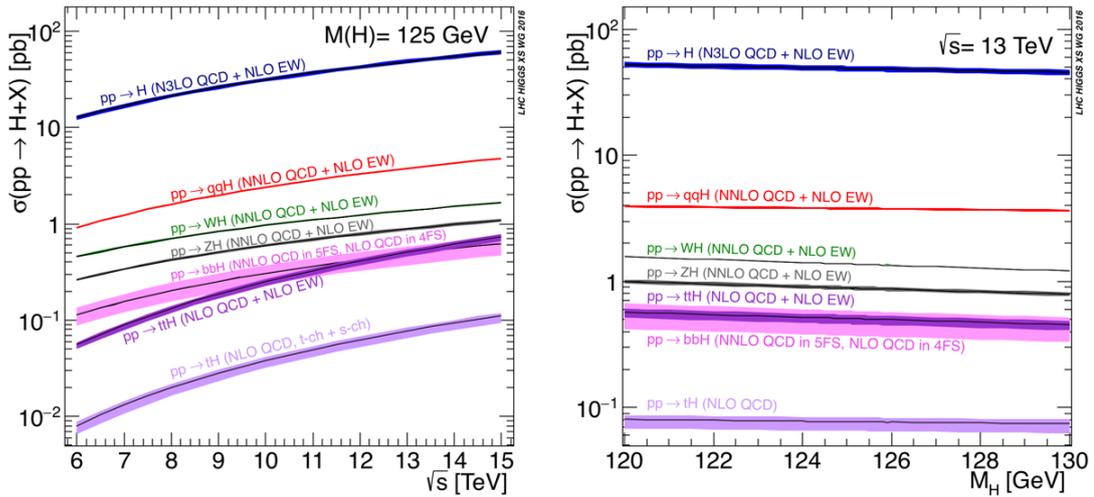


Figure 1.2: The Standard Model Higgs boson prediction for the cross section. On the *left* plot the cross section as a function of center-of-mass energy for 125 GeV Higgs boson, and on the *right* plot as a function of Higgs particle mass for $\sqrt{s} = 13$ TeV are shown [14].

Gluon fusion ($gg \rightarrow H$)

Gluon fusion (Fig. 1.3) is the largest cross section Higgs boson production mechanism at the LHC. In the SM $gg \rightarrow H$ process at Leading Order (LO) is mediated only by an exchange of virtual quarks, where the particular quark contribution is proportional to its mass [15]. The following terms constitute to the current best gluon fusion prediction for the cross section σ at the 13 TeV center-of-mass energy [15]:

- LO cross section of 16.00 pb. Although this contribution can be calculated precisely, its value is computed using effective theory to keep consistency with higher order corrections. The effective theory cross section is derived assuming infinite top quark mass and then rescaled (at each order) by the ratio $\sigma_{m_t}^{LO} / \sigma_{m_t=\infty}^{LO}$.

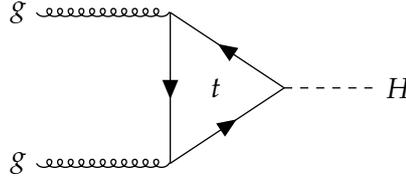


Figure 1.3: Main process Feynman diagram for gluon fusion Higgs production.

Here $\sigma_{m_t}^{LO}$ denotes the SM cross section with five massless quarks and a single massive top quark. In this approach the contributions from bottom and charm quarks are neglected. They are added separately using the precise SM result.

- NLO (Next-to-Leading Order) cross section of 18.79 pb. This contribution is also known exactly and computed analogously to the previous case. The additional correction is applied to match the exact theory (i.e. which takes into account the finite top quark mass) with the effective one.
- NNLO (Next-to-Next-to-Leading Order) cross section of 9.90 pb. The NNLO results are obtained only in the frame of effective theory and return 9.56 pb. The additional 0.34 pb comes from the inclusion of subleading corrections given through an expansion of the inverse top quark mass.
- Electroweak (EW) and N³LO ([Next-to]³ = Next-to-Next-to-Next-to-Leading Order) cross section of 3.89 pb. These consist of electroweak radiative corrections computed at the NLO (mixed QCD-electroweak corrections are known only in an effective theory [16]) and rescaled effective theory N³LO cross section.

All above sum up to the 125 GeV Higgs production cross section of:

$$\sigma^{ggH} = 48.58 \text{ pb}_{-6.72\%}^{+4.56\%} (\text{theory}) \pm 3.20\% (\text{PDF}+\alpha_s).$$

Vector boson fusion ($qq \rightarrow q'q'H$)

The second-largest Higgs production cross section at LHC is Vector Boson Fusion (VBF). In this process (Fig. 1.4) the scattering of two (anti-)quarks is mediated by a vector boson with radiation of the Higgs boson into the final state.

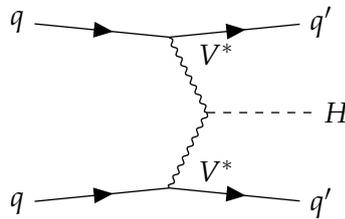


Figure 1.4: Feynman diagram for VBF Higgs production.

In the event the two scattered quarks are visible as two hard jets going in opposite direction along the beam. Furthermore, the final state gluon radiation is strongly suppressed in VBF due to color-singlet nature of the W/Z boson exchange. Because of these features, the VBF process is fairly easily distinguished from both large QCD

background and Higgs associated production. This makes VBF particularly useful for Higgs searches in channels with hadrons (jets) in the final state ($H \rightarrow b\bar{b}$, $H \rightarrow \tau\bar{\tau}$).

The VBF cross section is currently calculated at the NNLO QCD, where VBF process is considered as double deep-inelastic scattering (DIS) with standard DIS structure functions $F_i(x, Q^2)$; $i = 1, 2, 3$. Additionally, the relative NLO EW corrections and contribution from photon-induced channels are added. The final predicted cross section value at 13 TeV is the following:

$$\sigma^{qqH} = 3.7817 \text{ pb}_{-0.33\%}^{+0.43\%} (\text{theory}) \pm 2.1\% (\text{PDF}+\alpha_s)$$

Higgs Strahlung ($qq' \rightarrow Z/WH$)

The associated production of Higgs boson with a vector gauge boson (W or Z) is called *Higgs Strahlung* (Fig. 1.6) and is another important production process next to gluon fusion and VBF.

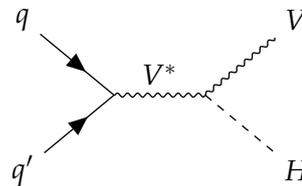


Figure 1.5: Feynman diagram for Higgs Strahlung.

Although the large QCD background limits the usefulness of this channel in dedicated Higgs search analyses, it can still be relevant e.g. for high- p_T $VH(H \rightarrow b\bar{b})$ events, where information about substructure of jets can be used to extract the signal [17] and where Higgs particle decays into a pair of W bosons or into invisible (undetectable) Beyond Standard Model particles. Moreover, the couplings of the Higgs boson to the W and Z bosons can be measured separately and uniquely throughout this channel.

The total inclusive theoretical cross section predictions for Higgs Strahlung are known in the frame of Standard Model up to NNLO QCD and NLO EW corrections. In the case of NLO QCD and majority of NNLO QCD corrections they can be reduced to the Drell-Yan production of a virtual gauge boson [18]. The irreducible part of NNLO QCD corrections contain Higgs boson radiation from top-quark loops (for $qq' \rightarrow VH$) and where Z and H couple to gluons via top-quark loop (for $gg \rightarrow ZH$) [19]. The NLO EW cannot be factorized into Drell-Yan-like process due to irreducible box diagrams already at one loop [20]. For the 125 GeV Higgs boson at 13 TeV we have:

$$\sigma^{VH} = 1.373 \text{ pb}_{-0.7\%}^{+0.5\%} (\text{theory}) \pm 1.9\% (\text{PDF}+\alpha_s).$$

Associated production ($gg \rightarrow t\bar{t}H$)

At LHC at 13 TeV, the Higgs boson is created in about 1% of events via associated production with a top-antitop quark pair. This channel can be used to probe the top quark Yukawa coupling y_t directly and model-independently⁵. Also, this channel is

⁵Indirect measurements of y_t can be performed with $gg \rightarrow H$ events assuming no contribution from unknown particles in the loop.

relatively experimentally clear but with a complex final state and the Higgs signal can be extracted using multivariate analysis (MVA) techniques [21].

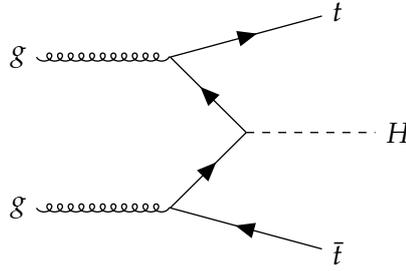


Figure 1.6: Feynman diagram for Higgs associated production with top-antitop quark pair.

The $t\bar{t}H$ cross section is known at the LO [22], NLO QCD [23] and NLO EW [24] orders. The predicted value for 125 GeV Higgs boson at 13 TeV is:

$$\sigma^{t\bar{t}H} = 0.5071 \text{ pb}_{-9.2\%}^{+5.8\%} (\text{theory}) \pm 3.6\% (\text{PDF}+\alpha_s).$$

In this thesis we neglect the $t\bar{t}H$ production because of its very small contribution to the analyzed data sample.

Other production processes

From the other 125 GeV Higgs boson production processes, it is worth to note 1) the associated production with single (anti-)top quark (74.25 fb at 13 TeV); 2) the associated production with charm-quark pair (85 fb at 13 TeV) and 3) the associated production with bottom-quark pair (479.20 fb at 13 TeV) [25]. The first two have negligible cross sections. The third one although having a cross section comparable to $gg \rightarrow t\bar{t}H$ process, has also a very large background at LHC.

1.3.3 Higgs decay modes

Theoretical total width of a 125 GeV SM Higgs boson is $4.07 \times 10^{-3} \text{ GeV}_{-3.9\%}^{+4.0\%}$ [26]. The dominant branching ratios with theoretical uncertainties are shown in Fig. 1.7. In this section we will give a brief introduction into the most important Higgs particle decay modes with extended consideration about the most interesting for us $H \rightarrow \tau^+\tau^-$ decay channel.

Diphoton decay mode ($H \rightarrow \gamma\gamma$)

The $H \rightarrow \gamma\gamma$ search channel (Fig. 1.8) is characterized by its high mass resolution of $\approx 2\%$ (no neutrinos in the final state and efficient reconstruction in the electromagnetic calorimeter) and very good background rejection (Higgs particle creates a narrow peak over otherwise smoothly falling diphoton mass distribution). The main backgrounds are prompt two photon production, photon with jet and pair of jet processes (where the jet fakes a photon). The final result is calculated by simultaneous fitting the signal yield from multiple categories (created with respect to the production process, expected mass resolution and signal-to-background ratio). The measured signal strength relative to the Standard Model prediction is equal to $1.18_{-0.14}^{+0.17}$ for CMS [27] and $0.99_{-0.14}^{+0.14}$ for ATLAS [28]. The CMS Collaboration additionally gave

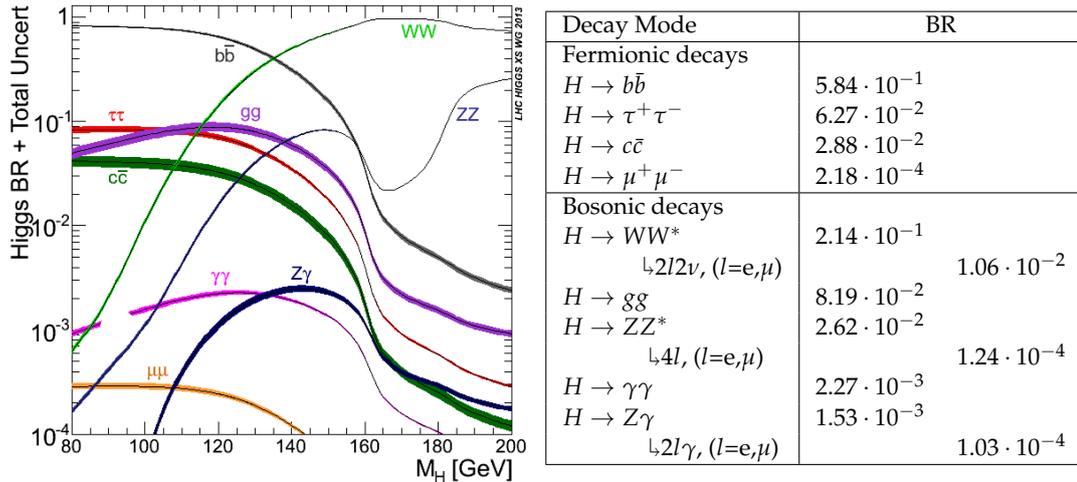


Figure 1.7: The Standard Model Higgs boson decay branching ratios and total uncertainties. On the left plot the 80-200 GeV particle mass range and on the right a table with numerical values for 125 GeV particle are shown [14].

the estimated significances for different categories: 6.7 (ggH), 1.7 (VBF), 3.2 ($t\bar{t}H$) and 2.4 (VH) standard deviations.

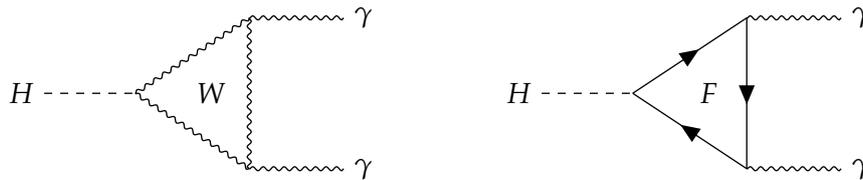


Figure 1.8: Leading-order Feynman diagrams of the Higgs boson decays to photons. $H \rightarrow \gamma\gamma$ is a one-loop process at LO. W stands for W boson and F for fermions in the loop.

Higgs to four leptons decay mode ($H \rightarrow 4l$)

Another channel with high mass resolution and good background separation is $H \rightarrow ZZ^* \rightarrow l^+l^-l'^+l'^-$. The main irreducible backgrounds are $q\bar{q} \rightarrow ZZ^*$ and $gg \rightarrow ZZ^*$. The reducible background is mainly from Z boson production together with a pair of bottom or top quarks or jets. Both CMS and ATLAS experiments performed measurements of Higgs boson mass (m_H) using this channel. For 13 TeV LHC data, the obtained values are 125.26 ± 0.21 GeV (CMS, 35.9 fb^{-1}) [29] and 124.79 ± 0.37 GeV (ATLAS, 36.1 fb^{-1}) [30].

Higgs into b-quark ($H \rightarrow b\bar{b}$) and W-boson ($H \rightarrow WW^*$) pair decay modes

The two above Higgs boson decay channels are often and interchangeably called *golden channels*. Other decay modes, although sometimes with a higher branching ratio, suffer from large background and/or poor Higgs boson mass determination. Both problems are valid for $H \rightarrow b\bar{b}$ and $H \rightarrow W^+W^- \rightarrow l^+l^-\nu\bar{\nu}$ channels. In the former case the vast majority of events do not have a lepton in the final state, which makes it very hard to separate them from an enormous QCD background at the LHC. The current state-of-the-art result in $H \rightarrow b\bar{b}$ channel for observed (expected) significance is 5.6 (5.5) standard deviations for CMS [31] and 5.3 (4.8) standard deviations for ATLAS [32]. The results were obtained by a combination of analyses

from Run 1 and Run 2 and from all considered topologies (associated production, VBF, etc.). The $H \rightarrow W^+W^- \rightarrow l^+l^-\nu\bar{\nu}$ analysis is usually subdivided into three, depending on the flavor of the leptons in the final state, i.e. e^+e^- , $\mu^+\mu^-$ and $e^\pm\mu^\mp$. These three channels are distinctive with regard to a set of dominant backgrounds, where the main events contamination may stem from non-resonant vector boson production, Drell-Yan process, vector boson plus jet or multijet production. The Higgs boson decay into a pair of W bosons was first observed by the CMS experiment for 35.9 fb^{-1} of $\sqrt{s} = 13 \text{ TeV}$ data and published in a 2019 paper [33]. ATLAS reported $H \rightarrow WW^*$ observation using only 25 fb^{-1} of data from $\sqrt{s} = 7$ and 8 TeV proton-proton collisions with an excess of 6.1 standard deviations over the estimated background [34].

Ditau decay mode ($H \rightarrow \tau^+\tau^-$)

The $H \rightarrow \tau^+\tau^-$ channel (branching fraction about 6%) is currently one of two (next to $H \rightarrow b\bar{b}$ with branching fraction of about 57%) at the LHC (or any hadronic collider with a comparable luminosity) best suited to perform measurements of Higgs boson couplings to fermions. In order to do this the overwhelming QCD background needs to be significantly reduced and the interesting events need to be extracted. This can be done comparatively easier with decays into pair of taus than bottom quarks thanks to tagging leptons from leptonic tau decays and powerful identification of taus-to-hadrons decays⁶.

The final result of the analysis is obtained by merging the calculations from multiple sub-analyses. The main analysis splitting is twofold:

- by tau decay mode - to electrons (τ_e), to muons (τ_μ) and to hadrons (τ_h) with corresponding neutrinos. This gives four detached sub-analyses of Higgs boson into: $\tau_h\tau_h$, $\tau_\mu\tau_h$, $\tau_e\tau_h$ and $\tau_e\tau_\mu$. Currently the $H \rightarrow \tau_\mu\tau_\mu$ and $H \rightarrow \tau_e\tau_e$ searches are excluded from $H \rightarrow \tau\tau$ analysis due to low branching fraction and significant background contributions,
- by number of jets: zero, two or all other cases. This is meant to optimize (with regard to signal-to-background ratio) the final cuts with respect to the Higgs boson production mode.

For all above cases the Higgs boson signal is searched for as a broad⁷ excess over the background in the Higgs boson mass (or better *mass of ditau system*) $m_{\tau\tau}$ distribution. The main background sources are: Z boson (Drell-Yan production) decays to taus and electrons⁸ (where the electron fakes the hadronic tau), W boson production associated with jets in the event (where W decays leptonically and jet fakes the hadronic tau), top-anti-top production and multijet production (again jet fakes the hadronic tau).

The Higgs boson decays into a pair of tau leptons was first observed in a single experiment by the CMS. In the discovery publication from 2017 the combined data from LHC 7, 8 and 13 TeV runs were used [35]. The $H \rightarrow \tau\tau$ was established with significance of 5.9 standard deviations (4.9 for 35.9 fb^{-1} with 13 TeV data only) compared to expected significance of 5.9 (4.7). The corresponding result from 2018

⁶In $H \rightarrow b\bar{b}$ channel the signal is mainly extracted for vector boson associated production (VH), where leptonic W/Z decays are used for event tagging.

⁷This is due to $m_{\tau\tau}$ poor resolution (one order of magnitude worse than for $H \rightarrow ZZ$ and $H \rightarrow \gamma\gamma$ channels) caused by undetectable neutrinos in the final state.

⁸The $Z \rightarrow ee$ background is however irrelevant for $H \rightarrow \tau_h\tau_\mu$ process, which is of main interest in this thesis.

ATLAS publication gave 6.4 standard deviations (4.4 for 36.1 fb^{-1} with 13 TeV data) compared to expected significance of 5.4 (4.1) [36].

1.4 Tau lepton

The *tau lepton* or *tauon* is a third-generation charged lepton [37, 38] discovered at SLAC-LBL detector in 1975 [39]. The tau particle has a mass of $1776.86 \pm 0.12 \text{ MeV}$ and its lifetime is $(290.3 \pm 0.5) \times 10^{-15} \text{ s}$ [26]. Tauon can decay either leptonically or semi-leptonically (tauon is the only lepton heavy enough to be able to decay into hadrons). In the former case, there is an electron or a muon with two neutrinos in the final state. The semi-leptonic decays are often referred to as *hadronic decays*, and in such case there is only one tauonic neutrino in the final state.

The basic possible decays and approximate branching fractions follows a simple counting rule. First, please note that taus are weakly interacting particles and can decay only via intermediate virtual weak boson (W^\pm). Since W^\pm can only decay into a fermion-anti-fermion pair, tau decays into baryons are forbidden. Moreover, the charge conservation enforces the final state meson to be build out of different flavored quarks. Out of the six quarks only three can occur in tau lepton decays due to energy conservation law: top ($m_t \approx 173 \text{ GeV}$) and bottom ($m_b \approx 4.18 \text{ GeV}$) are heavier than tau by itself and the lightest meson with a charm quark (D-meson with mass about 1.87 GeV) is also heavier. All above leaves us with following symbolic formula for the inclusive rate (the equation for anti-tau τ^+ is analogical) :

$$\begin{aligned} \Gamma(\tau^- \rightarrow \text{everything}) &= \Gamma(\tau^- \rightarrow \nu_\tau + e^- + \bar{\nu}_e) + \Gamma(\tau^- \rightarrow \nu_\tau + \mu^- + \bar{\nu}_\mu) \\ &+ \Gamma(\tau^- \rightarrow \nu_\tau + d_j + \bar{u}_j) + \Gamma(\tau^- \rightarrow \nu_\tau + s_j + \bar{u}_j). \end{aligned} \quad (1.14)$$

In the formula above all terms are a function of mass of the particles and hadronic terms are additionally associated with $|V_{ud}|^2$ and $|V_{us}|^2$ matrix elements of Cabibbo — Kobayashi — Maskawa matrix which contains information about the strength of the flavour-changing weak interaction. However, if we notice that $|V_{ud}|^2 + |V_{us}|^2 \approx 1$ and neglect the particles masses, we obtain the branching ratio of 20% separately for two fully leptonic channels and 60% for hadronic decays. The last figure ($3 \cdot 20\%$) stems from counting every quark color separately. More precise calculations require taking into account mass and QCD corrections effects. For reference, the current best experimental values for main branching fractions of tauon are presented in Table 1.1.

Mode	Decay channel	BR (Γ_i/Γ) [%]
Leptonic	$\tau^\pm \rightarrow \mu^\pm \bar{\nu}_\mu \nu_\tau$	17.392 ± 0.04
	$\tau^\pm \rightarrow e^\pm \bar{\nu}_e \nu_\tau$	17.816 ± 0.04
1-prong	$\tau^\pm \rightarrow h^\pm \nu_\tau$	11.506 ± 0.05
	$\tau^\pm \rightarrow h^\pm \pi^0 \nu_\tau$	25.935 ± 0.09
	$\tau^\pm \rightarrow h^\pm 2\pi^0 \nu_\tau$	9.458 ± 0.10
3-prong	$\tau^\pm \rightarrow h^\pm h^\pm h^\mp \nu_\tau$	9.780 ± 0.05
	$\tau^\pm \rightarrow h^\pm h^\pm h^\mp \pi^0 \nu_\tau$	4.790 ± 0.05
-	others	≈ 3.323

Table 1.1: Main tauon decay channels branching fractions [40]. h^\pm stands for π^\pm or K^\pm .

The lifetime of tauon translates into approx. $87 \mu\text{m} c\tau$ decay length. From the perspective of large general-purpose detector (like CMS or ATLAS) taking into account Lorentz boost, the decay length is too small to be solely and reliably used for tau lepton identification. Tauon fully leptonic decays are reconstructed as leptons

since electron and muon are stable (the latter in the range of the detector radius). The hadronic decays are reconstructed from hadrons. Charged pions and kaons

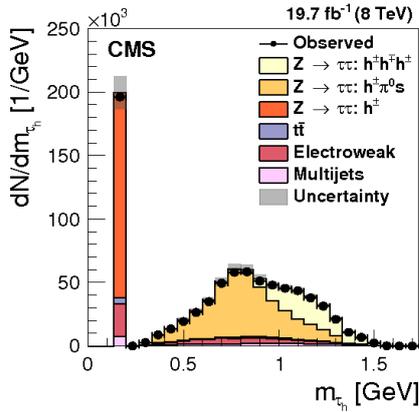


Figure 1.9: Hadronic tau mass distribution in $Z/\gamma^* \rightarrow \tau\tau$ events. [41]

are also semi-stable with decay length of several to several dozens meters. The remaining neutral pion decays almost instantaneously (with decay length of 67 nm) into almost exclusively a pair of photons (branching ratio ≈ 0.99)⁹. It may be interesting to consider the visual mass distribution of the hadronic tau lepton (Fig. 1.9). On the right hand side, the continuous distribution have two dominant contributions from the two resonances (vector mesons): ρ ($\tau^\pm \rightarrow \rho^\pm \nu \rightarrow \pi^\pm \pi^0 \nu$, $m_\rho = 770$ MeV, $c\tau \simeq 0$) and a_1 ($\tau^\pm \rightarrow a_1 \nu \rightarrow \pi^\pm \pi^\pm \pi^\mp \nu$ and $\tau^\pm \rightarrow a_1 \nu \rightarrow \pi^\pm \pi^0 \pi^0 \nu$, $m_{a_1} = 1260$ MeV, $c\tau \simeq 0$). The single peak on the left is the mass of the charged pion (about 135 MeV).

⁹Technically speaking CMS covers 100% of neutral pions decays, as its remaining decay modes involve an electron-positron pair in place of photon(s) which is taken into account in the reconstruction as $\gamma \rightarrow e^+e^-$ conversion (both processes are indistinguishable within CMS resolution).

Chapter 2

The Large Hadron Collider and the Compact Muon Solenoid detector

2.1 Large Hadron Collider

On 24 October 1945 the United Nations (UN) was created. The system of specialized agencies of this intergovernmental organization, tasked to promote international cooperation and to maintain international order, include UNESCO (United Nations Educational, Scientific and Cultural Organization), which has been authorized to "assist and encourage the formation of regional research laboratories in order to increase international scientific collaboration..." [42]. At an intergovernmental meeting of UNESCO in Paris in December 1951, the first resolution concerning the establishment of a European Council for Nuclear Research was adopted. Two months later, 11 countries signed an agreement establishing the provisional council – the acronym CERN (*Conseil Européen pour la Recherche Nucléaire*) was born [43].

CERN's mandate of establishing a world-class fundamental physics research organization in Europe led to multiple fundamental discoveries in the field of nuclear and particle physics, e.g.: discovery of the weak neutral currents (1954), first observations of antinuclei (1965), W and Z particles discovery (1983) and first antiatoms production (1995). The progress in theoretical physics and incorporation of Higgs Mechanism (or Brout–Englert–Higgs Mechanism, 1964) into SM (Weinberg and Salam, late 1960s) created the need for its experimental confirmations. The idea of large proton-proton synchrotron was originally conceived in the 1980s, and in 1994 CERN Council approved the construction of Large Hadron Collider (LHC) [44].

Neither time nor place for the LHC was coincidental. The first serious attempt for the Higgs boson discovery was conducted with Large Electron–Positron Collider (LEP) which was constructed between 1984 and 1989 at CERN and operated from 1989 until 2000. This largest electron-positron accelerator ever built worked at energies up to 209 GeV. The search was not conclusive with lower bound limit for the Higgs boson mass set at 114.4 GeV at the 95% confidence level, and a decision was taken to stop LEP and switch to LHC as soon as possible. Instead of using only weakly and electromagnetically interacting leptons colliding on relatively small energies and luminosity, the LHC collides protons which allows for large center-of-mass energy (low synchrotron radiation losses) and luminosity (no antiparticles) on the price of uncertainty in interacting partonic state and some design difficulties. Around 2001 LEP was dismantled to make way for the LHC, which re-used its 26.7 km long tunnel.

The decision to place LHC in LEP tunnel let to reduce the project costs but also influenced many technical solutions. For example the LEP tunnel has eight short arcs and eight long straight sections which allowed to compensate the high synchrotron radiation losses with long RF cavities. The arcs internal diameter of only 3.7 m was

adequate for the particle–antiparticle collider in which two counter-rotating beams share the same ring, but not in case of particle–particle collider where two rings are needed (beams require opposite magnetic dipole fields). Eventually the twin-bore magnet design was adapted [45].

The performance of LHC is determined by multiple factors:

- The maximal collision energy is constructed by the value of magnetic field created by an array of 1232 dipole magnets around the pipe. 8 T superconducting magnets cooled with superfluid helium used in the LHC allow for accelerating particles to designed energy of 8 TeV. At this energy, the magnet superconducting cables have to handle electric currents up to 12 kA and the cooling system has to maintain constant temperature of 1.9 K, which is difficult taking into account very low heat capacity of the cable in such environment. In order to avoid very long training of magnets and to minimize the number of quenches the collision energy for Run II of LHC was limited to 13 TeV [46].
- The number of events N_{event} occurring at the interaction point (IP) in every second is given by:

$$N_{event} = L\sigma_{event}, \quad (2.1)$$

where σ_{event} is the cross section for the event and L is the instantaneous luminosity. The machine luminosity depends only on the beam parameters and can be written for a beam with Gaussian profile of proton distribution as [45]:

$$L = \frac{N_b^2 n_b f \gamma_r}{4\pi\epsilon_n \beta^*} \left[1 + \left(\frac{\theta_c \sigma_z}{2\sigma^*} \right)^2 \right]^{-1/2}, \quad (2.2)$$

where N_b is the number of particles per bunch, n_b - the number of bunches per beam, f - the revolution frequency, γ_r - the relativistic gamma factor, ϵ_n - the normalized transverse beam emittance, β^* - the beta function at the collision point, θ_c - the full crossing angle at the IP¹, σ_z - the RMS (root mean square) bunch length, and σ^* - the transverse RMS beam size at the IP. The both integrated and maximum peak luminosity is limited by a number of reasons. The properties of storage ring constrains in a quite sophisticated manner the maximum β -function to 180 meters in arcs. The beam screen aperture of 44.04×34.28 restrict the maximum transverse beam size to 0.119 cm when combined with prescribed minimum aperture of 10σ RMS beam size. The obtained values of β -function and beam size translate into $3.75 \mu\text{m}$ limit on the transverse beam emittance [48]. Then the beam-beam interaction at all IPs measured by the linear tune shift:

$$\xi = \frac{N_b r_p}{4\pi\epsilon_n}, \quad (2.3)$$

should not exceed 0.015. In the above formula r_p is the classical proton radius. From Eq. 2.3 one can find that linear beam-beam shift and the aperture of the LHC impose the limit on the maximum bunch intensity of $N_b^{max} = 1.15 \times 10^{11}$ (this number was actually reached in 2017 [49]). Further limitations stem from e.g.: aperture of the triplet magnets, total stored beam energy and synchrotron radiation; and set the maximum designed peak luminosity for LHC at $L_{peak}^{max} = 10^{34} \text{ cm}^{-2}\text{s}^{-1}$ (the real maximum peak luminosity obtained is more than twice

¹The LHC collision crossing angle is of the order of couple of hundred microradians (typically $\sim 150 \mu\text{rad}$) and slowly decreases as the LHC fill goes [47].

the designed and equal to $2.14 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$ [50]). The luminosity of the beam traveling through the tunnel decreases with time due to degradation and emittance. For LHC the luminosity lifetime and average turnaround time can be calculated for 15 and 1.2 hours, respectively. Knowing these values a maximum total integrated luminosity can be estimated to be 120 fb^{-1} assuming 200 days of operation per year.

Collected luminosity

The LHC was successfully commissioned in 2010, colliding protons with 7 TeV center-of-mass energy till April 2012 and 8 TeV later at the end of the year. During this time the collider reached a peak luminosity of $7.7 \times 10^{33} \text{ cm}^{-2}\text{s}^{-1}$ and delivered about 25 fb^{-1} to both CMS and ATLAS experiments. After the Long Shutdown 1 (LS1, 2013 – 2015) the LHC unleashed its full potential at 13 TeV center-of-mass energy and delivered 4.2 fb^{-1} (2015) + 41 fb^{-1} (2016) + 49.8 fb^{-1} (2017) + 67.9 fb^{-1} (2018) of integrated luminosity. Fig. 2.1 shows the peak luminosity and the increase in integrated luminosity over time at the CMS. At the time of writing, we are during the LHC Long Shutdown 2 (LS2, 12.2018 – 02.2021).

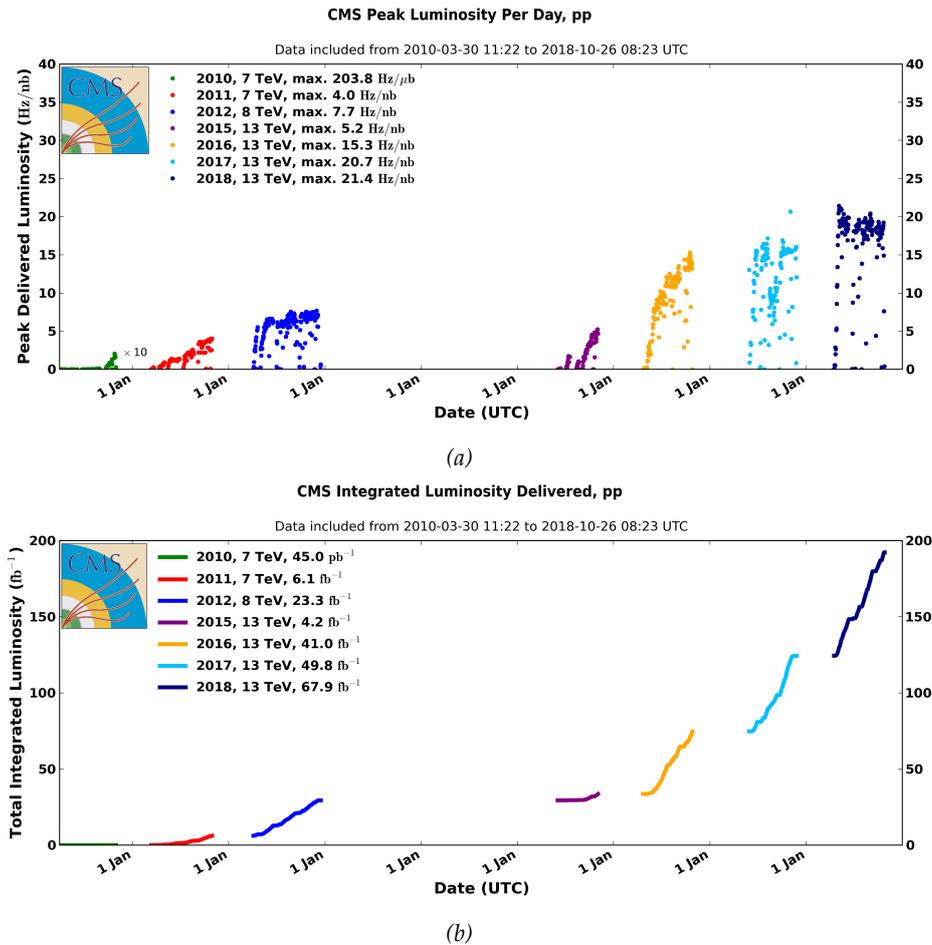


Figure 2.1: (a) Peak luminosity versus time for 2010-2012 and 2015-2018 (pp data only). (b) Delivered luminosity versus time for 2010-2012 and 2015-2018 in the CMS experiment (pp data only) [50].

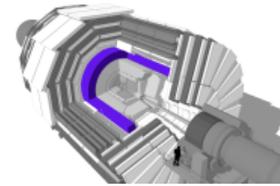
2.2 Compact Muon Solenoid detector

Compact Muon Solenoid (CMS) is a general purpose detector build upon scheme of a cylinder 29 m long and 15 m in diameter². This 14 kilotons heavy apparatus is divided into 5 segments: the central barrel, two endcaps and two outwardly placed forward calorimeters [52, 53].

The cartesian coordinate system of the CMS originates at the center of the detector and has x, y, z -axes pointing to the center of the LHC, up and along the anticlockwise beam direction, respectively. The θ angle between the particle three-momentum and the positive direction of the z -axis (polar angle) is used to define pseudorapidity η as $-\ln[\tan(\theta/2)]$. The azimuthal angle is defined on the $x - y$ plane and relative to the x -axis.

2.2.1 Superconducting magnet

The central feature of the CMS is a huge solenoid, which produces 3.8 T of magnetic field and stores 2.66 GJ of total energy. The design of the magnet was determined by the following considerations:



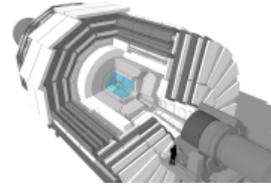
- The generated field value. The charged particle path curvature decreases with its energy and increases with external magnetic field. The precision measurement of a particle momentum in a finite detector range requires high values of magnetic field. In other words, the resolution of the tracker and muon system is a function of particle energy and field induction.
- The generated field geometry. There are three factors here. First, the magnet has to be long enough, so it can cover the highest possible pseudorapidity range and the (ultra-)peripheral events can be detected. Second, the shape of the field influences the detector resolution. Additionally, the more homogeneous magnetic field is, the simpler algorithm for reconstruction can be used.
- The presence of other CMS subsystems. The magnet coil cannot interfere with the tracker, calorimeters, nor the muon system.
- Costs. A strong magnet uses considerable amount of power and enforces utilization of an efficient cooling system. On the other hand, a weak magnet requires the whole detector to be enlarged in order to obtain comparable usability. In particular, the big part of the construction budget was devoted to creation of an underground detector cavern. All institutes involved in the CMS Collaboration participate in the cost of the magnet.

The CMS magnet is a 12.9 m long cylinder with inner bore of 5.9 m and is constructed out of five coil modules. Each module is made of four 2.65 km long Rutherford-type niobium-titanium cables co-extruded with high purity aluminum for thermal stabilization. The four-layer winding and large cross-section of conductor of $6.4 \times 2.2 \text{ cm}^2$ help withstand 64 atmospheres of hoop stress inside the detector coming from its own magnetic field and voltage of almost 19 kV. The magnet ensures the momentum resolutions of $\Delta p/p \sim 10\%$ at the TeV scale.

²CMS scheme icons in this chapter are taken from *Sketchup images highlighting the sub-detectors* [51].

2.2.2 Tracker

The innermost part of the CMS detector – the tracker – is responsible for the reconstruction of tracks of charged particles. The high spatial and temporal resolution is achieved by using fast and tiny semiconductor detectors, which are stacked together to create the world’s largest silicon detector with 5.8 m in length and 2.5 m in diameter.



The principle of operation of a semiconducting detector is comparable to the operation of the diode. A charged particle traveling through the volume of the detector creates an electron-hole pair. The free electrons from the conduction band are first captured by the electrode and then multiplied in an amplifier. The scheme of silicon detector and CMS pixel module is shown in Fig. 2.2.

The particle track reconstruction is based on the signal localization coming from electrodes. Due to a small band gap detectors based on silicon are able to operate stably at room temperatures although the CMS tracker is cooled to sub-zero temperatures in order to mitigate material degradation in the harsh environment in the vicinity of the beam.

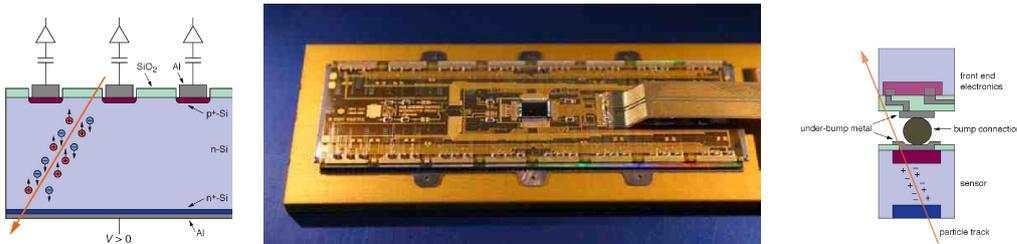


Figure 2.2: Working scheme of typical silicon detector (left), CMS pixel detector module (middle) and pixel scheme (right) [54].

The tracker is comprised of two sub-detectors: the pixel tracker and the strip tracker. They both create two concentric cylinders along the CMS z -axis, contain barrel ($|\eta| < 0.9$) and endcap ($|\eta| < 2.5$) parts and have an independent power supply, read-out and cooling systems [55].

The pixel detector³ consists of 66 million pixels made of 285 μm thick silicon n^+ -in- n sensors and creating cells of size $100 \times 150 \mu\text{m}$ in $r\phi \times z$. The pixel barrel system consists of three layers of radii of 4.4, 7.3 and 10.2 cm, and pixel endcap system consists of two pairs of disks at $z = \pm 34.5$ and $z = \pm 46.5$ cm. The whole pixel detector is build of 1440 modules covering about 1 m^2 sensing area. The overall resolution is approximately 20–40 μm in the longitudinal (z) coordinate and 10 μm in the transverse ($r\phi$) coordinate. The third coordinate is determined by the sensor plane position [57].

The strip detector has 9.3 mln strips/channels covering about 198 m^2 and consists of:

- Tracker Inner Barrel (TIB) - made of 4 layers of strip detectors, which are placed between $r = 20$ cm and $r = 55$ cm from the beam axis. It uses 320 μm thick silicon detectors with 80 to 120 μm distance between neighboring strips. TIB

³We will give the description of the pixel tracker that was present at the time of 2016 data taking (i.e. for data samples used in this thesis). In 2017, CMS has upgraded pixel detector to mitigate expected inefficiencies of the readout chip at the LHC luminosities after 2017. The new pixel detector has new readout chips with increased readout bandwidth, an additional layer in the barrel part and also an additional disk per side in the endcaps (which all adds up to 124 million pixels/channels) [56].

provides hit position measurements resolution of up to $23 \mu\text{m}$ in both $r\phi$ and z directions.

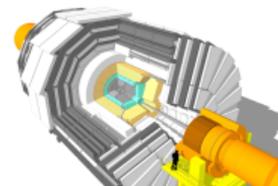
- Tracker Outer Barrel (TOB) - made of 6 layers of strip detectors, which are placed between $r = 55 \text{ cm}$ and $r = 116 \text{ cm}$ from the beam axis. It uses $500 \mu\text{m}$ thick silicon detectors with 122 to $183 \mu\text{m}$ distance between neighboring strips. TOB provides hit position measurements resolution of up to $35 \mu\text{m}$ in $r\phi$ and $52 \mu\text{m}$ in z directions.
- Tracker Inner Disks (TID) - made of 3 disks, which are placed between $z = 58 \text{ cm}$ and $z = 124 \text{ cm}$. It uses $320 \mu\text{m}$ thick silicon detectors with 100 to $141 \mu\text{m}$ distance between neighboring strips.
- Tracker End Caps (TEC) - made of 9 disks, which are placed between $z = 124 \text{ cm}$ and $z = 282 \text{ cm}$. It uses $320 \mu\text{m}$ thick silicon detectors in the three innermost rings and $500 \mu\text{m}$ thick detectors in the remaining rings. The distance between neighboring strips varies between 97 and $184 \mu\text{m}$.

The tracker allows to reconstruct tracks for pseudorapidity below 2.5 , particle transverse momentum above 0.1 GeV and produced up to 60 cm from the beam line. The efficiency of the reconstruction and the quality of the track depend on the pile-up and particle type, momentum and pseudorapidity. For reference, the prompt isolated muons with $|\eta| < 1.4$ and $p_T = 100 \text{ GeV}$ are reconstructed with essentially 100% efficiency and with p_T , transverse and longitudinal impact parameter resolution of approximately 2.8% , $10 \mu\text{m}$ and $30 \mu\text{m}$, respectively. The reconstruction efficiency for prompt charged particles in $t\bar{t}$ events with $p_T = 10 \text{ GeV}$ and for a mean number of interactions per crossing around 10 is 94% in barrel and 85% in endcaps and the resolution is 1.5% , $25 \mu\text{m}$ and $45 \mu\text{m}$ for the same quantities as previously [57].

2.2.3 Calorimeters

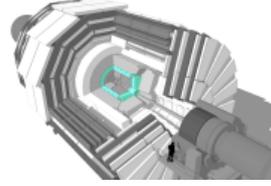
Calorimeters allow for the measurement of particle energies and also, to a lesser extent, measurement of the trajectory. The operation of the calorimeters relies on the phenomenon called *particle shower*, i.e. production of multiple secondary particles with less energy as the result of interaction of particle with matter. The energy deposits can be measured using scintillating properties of some materials and by calculating emitted photons intensity, which is proportional to the particle energy. The information about particle position is obtained by exploiting the segmentation of scintillation detectors. The increase of relative resolution with the particle energy is a distinctive feature of calorimeters. These are usually placed between the tracker and muon chambers.

The construction of efficient calorimetric systems is mostly a matter of materials engineering. There are calorimeters optimized only to particles interacting by electromagnetic force (*Electromagnetic CALorimeters/ECALs*) or by strong nuclear force (*Hadronic CALorimeters/HCALs*) or general purpose calorimeters. The scintillator simultaneously may be an absorber (*Homogeneous calorimeters*) or may be interconnected with some dense material, e.g. lead (*Sampling calorimeters*).



Electromagnetic calorimeter

The Electromagnetic CALorimeter (ECAL) of the CMS is a hermetic and homogeneous detector made out of 61200 lead tungstate (PbWO_4) scintillating crystals. It is placed outside the tracker and divided into a barrel and two endcaps.



The barrel covers pseudorapidity region below 1.479, has an inner diameter of 258 cm and is divided into 36 identical *supermodules*, 18 on each half of the barrel. The supermodule weights around 3 tonnes and covers 20° in ϕ . The short radiation length of 0.89 cm and small Molière radius of 2.2 cm of lead tungstate made the construction of compact and finely granulated calorimeter possible. A single crystal is an inverted truncated pyramid with a square $2.2 \times 2.2 \text{ cm}^2$ base⁴ (front face), height of 23 cm (corresponding to 26 radiation lengths) and covers 0.0174 in $\Delta\phi$ and $\Delta\eta$; hence the granularity of the barrel is 360-fold in ϕ i 170-fold in η .

The endcaps cover pseudorapidity of $1.479 < \eta < 3.0$ and are placed 314.4 cm (with magnet on) away from the primary vertex. Each endcap is subdivided into two halves comprising of 3662 crystals. Each crystal is a polyhedron with approximately $2.9 \times 2.9 \text{ cm}^2$ base (front face) and height of 22 cm.

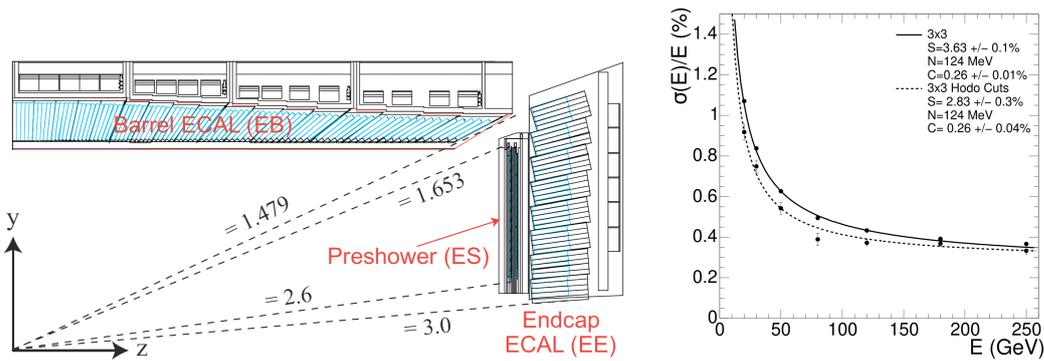


Figure 2.3: Left: Longitudinal view of the ECAL. Right: ECAL supermodule energy resolution, σ_E/E , as a function of electron energy measured from a beam test. The upper series of points correspond to events taken with a $2 \times 2 \text{ cm}^2$ trigger. The lower series of points correspond to events selected to fall within a $0.4 \times 0.4 \text{ cm}^2$ region. The energy was measured in an array of 3×3 crystals with electrons impacting the central crystal [53].

The two layer sampling calorimeter (preshower device or ES) is installed in front of the fiducial region of the endcaps ($1.653 < \eta < 2.6$), and is designed primarily to identify neutral pions and to help in identification of electrons. In ES the lead is used to initiate particle showers from photons/electrons and silicon strip sensors measure the deposited energy.

The CMS ECAL energy resolution can be measured with a test beam⁵ (Fig. 2.3) and parameterized with the following function:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2,$$

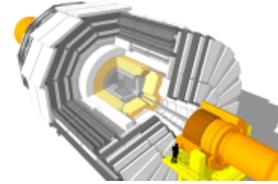
⁴The equality of this values with Molière radius of PbWO_4 allows to capture full cascade with matrix of 2×2 crystals.

⁵The electromagnetic calorimeter resolution was also confirmed with 5 fb^{-1} of $\sqrt{s} = 7 \text{ TeV}$ 2010/11 data [58]. The obtained energy resolution was 2% for barrel and 2-5% for endcaps in the Z boson decay events.

where E is a energy of incident electron/photon energy, S is the stochastic term, which depends on photodetector gain, photo-statistics and event-to-event fluctuations, N represents the noise term, which depends on event pile-up and electronic noise, C is a constant term, which depends on additional effects like leakage of energy from the rear face of the crystal, the accuracy of the inter-calibration constants or non-uniformity of the longitudinal light collection [59]. For the representative result shown in Fig. 2.3, the values are: 3.63 or 2.83 for S (depending on the trigger), 124 MeV for N and 0.26 for C .

Hadronic calorimeter

The volume between ECAL and the magnet coil is occupied by the Hadronic Calorimeter (HCAL). It is primarily designed to absorb all remaining particles from the interaction except muons and neutrinos. Hence, the selection of brass for the absorber, as it is non-magnetic, easy to machine and most importantly has a relatively short interaction length⁶. The brass is complemented with stainless steel in the innermost and outermost layers for the sake of structural strength. The plastic scintillator with embedded wavelength-shifting (WLS) fibres (tile/fibre technology) plays the role of an active medium. The signal is carried out outside of the scintillator by high-attenuation-length fibres to multi-channel hybrid photodiodes.



Similarly to the ECAL, the HCAL is divided into a hadron barrel part (HB) and two hadron endcap parts (HE). Additionally, two specialized subdetectors form parts of HCAL: 1) hadron outer detector (HO) which increases HCAL to over 10 interaction lengths serving as a “tail-catcher” located in the return yoke of the barrel and 2) hadron forward detector (HF) which increases HCAL pseudorapidity coverage to 5.0.

The HB consists of 32 towers and lies from 177.7 cm to 287.65 cm from the beam line and is assembled of two half barrels, each covering pseudorapidity region below 1.4 and constructed of 18 20° wedges in ϕ . The wedge contains 17 layers which are made of plastic scintillator tiles stacked with brass or stainless steel. The double thick first layer of scintillator is dedicated to measurements of low energy showering particles. Each tile has a size of $\Delta\phi \times \Delta\eta = 0.087 \times 0.087$ and is instrumented with single WLS.

The 14 towers of HE cover pseudorapidity region $1.3 < |\eta| < 3.0$. The baseline segmentation of HE of 5° in azimuthal direction and 0.087 in η is modified for $\eta > 1.74$ in order to accommodate the bending radius of the fiber readout and is equal to 10° in ϕ and 0.09 – 0.35 in pseudorapidity. In total, HE consists of 2304 towers.

The HO is located in front of the barrel muon system and covers $|\eta| < 1.26$. It is built with one layer 1 cm thick scintillator at 409.7 cm radial distance everywhere, except for the most central region of 2.54 m along the z -axis, where another active layer and 18 cm thick iron absorber is placed.

The hadron forward parts of HCAL are located behind the HEs, between 11.2 m and 12.85 m from the interaction point. The HF is a steel/quartz fibre detector which is suitable for the congested forward region environment due to short and narrow hadronic showers. The Cherenkov light originated from passing particle is collected in quartz fibers aligned parallel to the beam axis and positioned at 0.5 cm intervals

⁶The brass used is cartridge brass #260 composed of 70% Copper and 30% Zinc, characterized by the density of 8.83 g/cm^3 and nuclear interaction length of 16.42 cm.

in 0.1 cm grooves in surrounding steel plates. The HF is divided into 13 towers in η and 35 towers in ϕ totaling to 910 towers and over 1800 channels in both HF parts.

The HCAL and ECAL barrel hadronic energy resolution may be parameterized as $\sigma/E = a/\sqrt{E} \oplus b$ (E in GeV), with stochastic term $a = 0.847 \pm 0.016$ GeV and constant term $b = 0.074 \pm 0.008$. The endcap energy resolution is similar to that in the barrel [60].

2.2.4 Muon system

Muons do not interact by strong force and emit only minor bremsstrahlung so the vast majority of muons produced in LHC can survive the travel through tracker and calorimeters of the CMS. In fact, the design assumption of the CMS is that if neutrinos and non-SM particles are neglected, the only particles going outside the magnet coil (or HO) from interaction vertex are muons. The negligible background for muons in outermost regions of the CMS allowed for construction of very efficient muon track finding device, and the 1.8 T magnetic field inside large return yoke helps in reliable momentum recognition even for high- p_T muons.

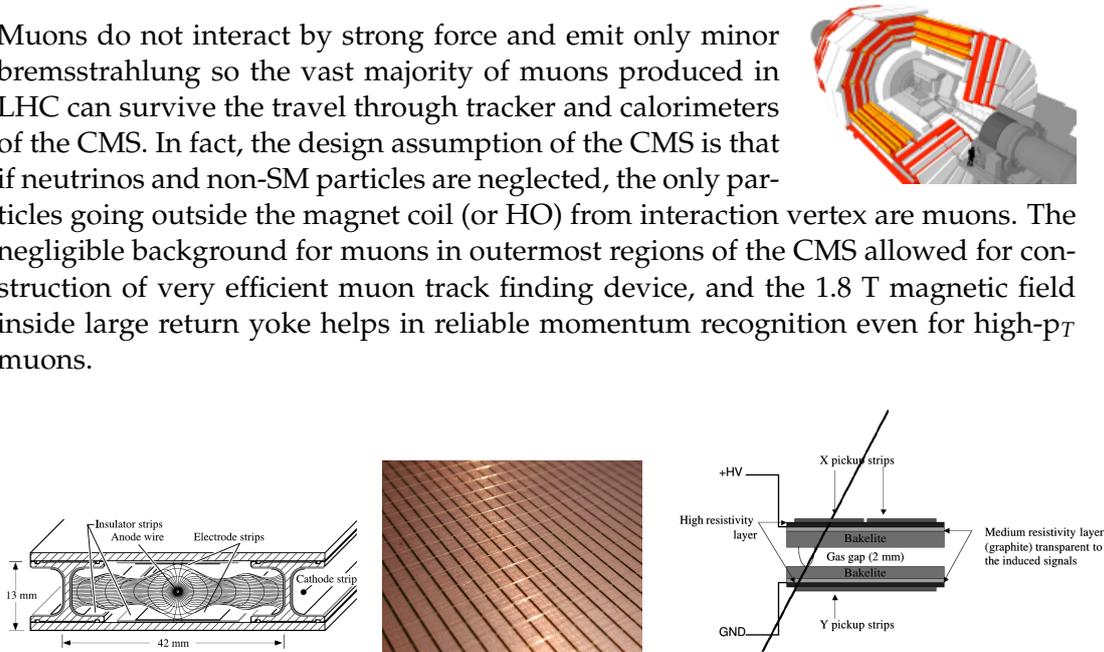


Figure 2.4: Left: Scheme of the DT chamber; Center: Photo of the CSC with visible chambers; Right: Scheme of the double gap RPC used in CMS [61].

The muon system is made exclusively of gaseous detectors, but using different chamber technology in different regions of the CMS (see Fig. 2.4) to ensure robust and redundant operation to fully explore muon physics potential.

The Drift Tube (DT) is a proportional chamber with a cell pitch of 4.2 cm (2 cm drift length), characterized by a high spatial resolution and at relatively low price, but also low temporal resolution, high deadtime and susceptibility on external electromagnetic field. The Drift Tubes are used in the barrel region, where a very large surface has to be covered and the residual magnetic field and the particle flux are low. Single DT station is designed to produce a spacial muon momentum vector, with resolution of $100 \mu\text{m}$ in position and 1 mrad in the direction of ϕ .

The derivative of Multiwire Proportional Chamber called Cathode Strip Chamber (CSC) is deployed in the endcaps of the muon system ($0.9 < |\eta| < 2.4$), where neutron induced background rate, muon rate and the magnetic field are all high. The geometry and thereby the production process is much more complicated for CSCs than for the DTs. In return, CSCs have large bandwidth and are stable in high magnetic field environment. At the CMS, a single CSC chamber consists of 7 trapezoidal panels which create 6 gas gaps. Every gap has a net of anodes and cathodes running perpendicular to each other. The anodes collect the electron from avalanche induced by the passing of the particle (cathodes collect an image charge). The CSCs have a typical spatial resolution of $200 \mu\text{m}$ and an angular resolution of 10 mrad in ϕ .

Resistive Plate Chambers (RPCs) are installed on top of the DTs and CSCs for $|\eta| < 1.6$. They have effective temporal resolution of the order of several nanoseconds and are dedicated to providing fast information for the Level-1 trigger. The CMS RPCs are double 0.2 cm gas gap chambers operating in avalanche mode to ensure stable operation at high signal rates.

The Barrel Muon Detector of the CMS experiment covers pseudorapidity below 1.2 and is divided into 5 wheels along z -axis, into 4 stations in radial direction and into 12 sectors in azimuthal angle. DTs detectors are sandwiched by RPCs in two innermost layers, and paired with RPCs in remaining four. Therefore, a single muon can cross up to four DT chambers and six RPCs, leaving 44 measurement points in the DT system and 6 in RPC system. In total, there are 250 DT chambers and 480 RPCs in the barrel muon system.

The Endcap Detector is made of 4 disks mounted perpendicular to the beam at both sides of the barrel and covers pseudorapidity between ± 0.9 and ± 2.4 . The majority of the 468 CSCs placed in the disks are overlapped in ϕ in order to avoid gaps in the system. CSCs for $\eta < 1.61$ are accompanied by 576 RPCs in all four stations (the 144 in fourth station were added during LS1).

The segmentation of the hosted in the magnet return yoke muon system is shown in Fig. 2.5.

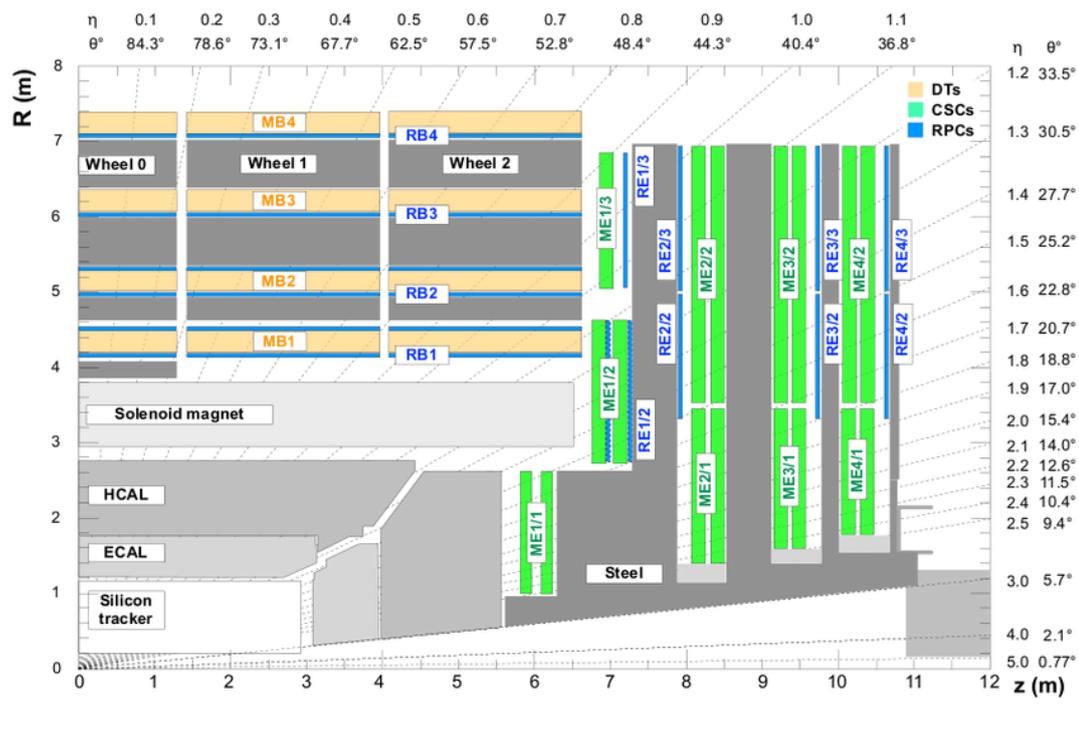


Figure 2.5: Top right quarter view of the CMS muon system in rz -plane [53]. In 2016 four layers of RPC detectors were installed in the endcaps.

2.2.5 Trigger system

With a new bunch crossing coming every 25 ns it is not technically possible to process all events online and with full available granularity nor save them all to the storage. To go around this problem, the CMS detector is equipped with a trigger system, which reduces beam-beam collision rate down to 1 kHz. Such a system has

to be designed in a manner that enables to keep the latency, timing and the rate controlled simultaneously ensuring good physics performance. In the CMS, the trigger system is divided into Level 1 trigger (L1) and High Level Trigger (HLT).

Level 1 trigger (Fig. 2.6) utilizes simplified algorithms implemented on a custom hardware (FPGAs and ASICs) in order to reduce the collision rate to below 100 kHz. L1 is synchronized with 40 MHz LHC clock and generates a L1 accept decision with a fixed latency of about 4 μ s. Level 1 trigger comprises of two independent sub-systems, namely of the calorimeter trigger and the muon trigger (no tracker data is used at L1). The former uses signals coming from calorimeters to create simplified high-level objects: electrons, photons, jets, tau candidates and Missing Transverse Energy (MET). The latter correlates hits in muon chambers to reconstruct muons.

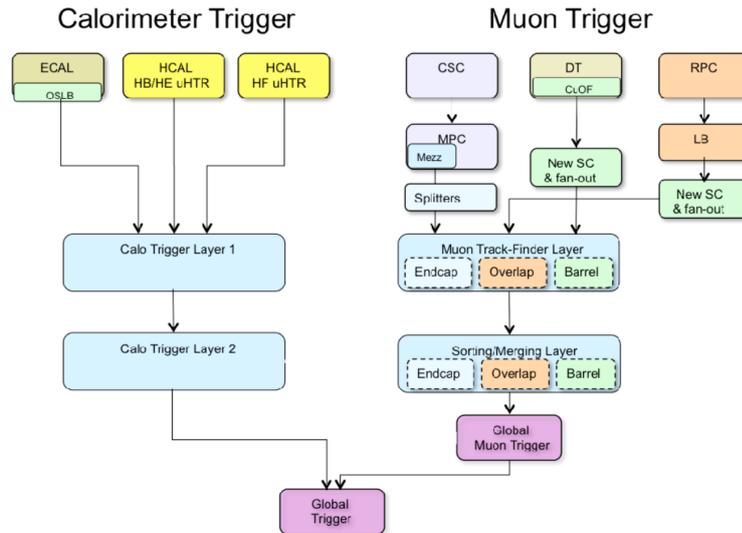


Figure 2.6: Overview of the CMS L1 trigger system for Phase-I upgrade (used in Run 2 after 2015/16) [62].

All calorimetric trigger algorithms are streamlined and follow the order:

- The particle transverse energy from ECAL, HCAL and HF in the form of primitives are unpacked, linearized and undergo some energy threshold requirements.
- The dynamic clustering algorithm is used to reconstruct individual clusters. The good electron energy containment and energy resolution is obtained with this method, which helps to minimize the effects of pile-up. The individual clusters are subsequently combined, filtered and summed.
- The electron/photon-, tau lepton-, jet- and MET- finders are used to generate high level objects. The identification is based on the energy deposit values, shower profile, cluster shape and isolation criteria. The jets finder employs the 9x9 trigger towers sliding window centered around a local maxima to approximate $R = 0.4$ cone radius used in offline anti- k_T algorithm [63].
- The calibration and isolation energy are evaluated.
- Sorted reconstructed objects are directed to the links out to the Global Trigger (μ GT).

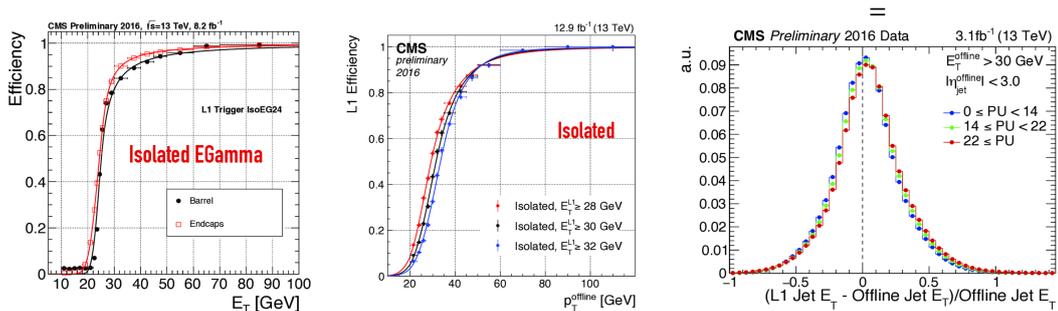


Figure 2.7: (Left:) Efficiency for a single e/γ trigger reconstruction vs offline E_T in the barrel (black curve) and in the endcaps (red curve). Threshold of $E_T > 40$ GeV applied. (Center:) The same for a single tau vs offline p_T . Curves plotted for taus with E_T above 28 (red curve), above 30 (black curve) and above 32 (blue curve). (Right:) Normalized distribution of the difference in jet transverse energy between Level 1 trigger and offline reconstruction. Different colors label results for different pile-up vertices count. The $E_T^{offline} > 30$ GeV and pseudorapidity below 3.0 thresholds are applied [63].

The performance of the calorimetric trigger is presented in Fig. 2.7.

The L1 muon trigger consists of three separate algorithms for three pseudorapidity regions of the detector: Barrel Muon Track Finder (BMTF, $|\eta| < 0.83$), Overlap Muon Track Finder (OMTF, $0.83 < |\eta| < 1.24$) and Endcap Muon Track Finder (EMTF, $1.24 < |\eta| < 2.4$). The Global Muon Trigger (μ GMT) cleans the duplicates of reconstructed muon candidates generated in all systems above. The muon candidates from μ GMT are sent to Global Trigger (μ GT) where L1 decision (to hold or remove event) is undertaken (taking into account calorimeter data).

The BMTF track finding algorithm is a three-stage algorithm. In the first stage the algorithm combines track primitives⁷ from different stations. It is done by extrapolating the muon candidate to the following station and checking if it is within an acceptable window. In the second stage the acceptable extrapolations create tracks with corresponding quality. In the third stage the tracks are issued physical parameters like transverse momentum and pseudorapidity.

OMTF is based on the idea of the comparison of the hits from DT, RPC and CSC (on an equal footing), with the so-called "Golden Patterns". Golden Pattern holds information about bending distribution in each layer for a muon of a given p_T [64]. Different p_T hypotheses are probed and the result muons undergo *ghostbusting* i.e. are cleaned from duplicates, before being delivered to the μ GMT.

EMTF uses pattern-based approach to the muon finding. Track segments from the CSCs together with RPC clusters are matched with patterns in the azimuthal direction to form tracks. Here, the RPC hits are used only in case of lacking CSC data. Furthermore, the low quality (very scattered) and duplicating tracks are removed. Finally, the best three tracks are assigned transverse momentum, azimuthal and bending angles, and charge.

The efficiency plots for the Muon Track Finders are presented in Fig. 2.8.

Events accepted by Level 1 trigger are read fully (i.e. all signals from subdetectors) and passed on for a further analysis in the HLT. HLT is built on a commercial hardware (26 000 CPU cores in 2017) to run online event reconstruction using algorithms closely related to the ones used in the offline (final) analysis. Therefore,

⁷Track primitives give muon coordinates, bending angle as well as quality bits. In case of BMTF the track primitives are called *super-primitives*, because they use combined information from DT super-layer and RPC hits. The use of super-primitives allows to improve efficiency and obtain better bunch crossing determination.

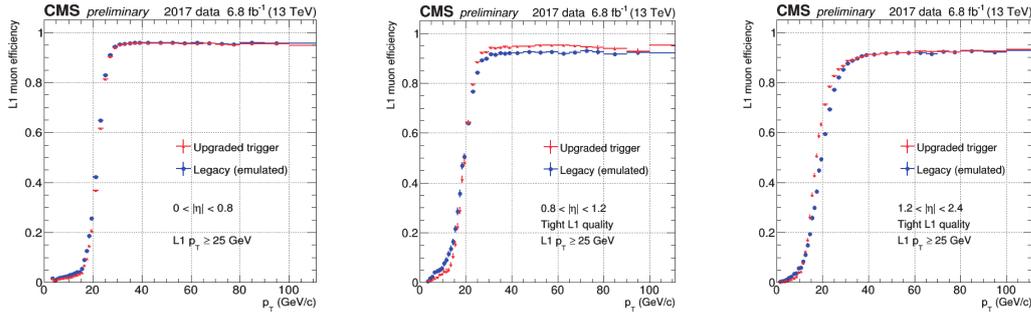


Figure 2.8: Efficiency distributions as a function of reconstructed muon p_T for EMTF (Left), OMFT (Center), EMTF (Right). The $p_T = 25\text{GeV}$ threshold applied [64].

HLT takes advantage of the full granularity of the CMS and thanks to performance refinements generates a decision in average time of (26k cores / 100 kHz =) 260 ms per event. As a result, the maximum input rate of 100 kHz rate is further reduced to about 1 kHz.⁸ HLT is designed as a large number (hundreds) of independent sequences of reconstructing modules seeded by L1 triggers called "HLT paths". Additional filtering modules are implemented to discard uninteresting events at the early stages of reconstruction. The common sequences are interchanged across different HLT paths and a single event can be processed within different streams and datasets. Also, special LHC runs like cosmic, low pile-up and heavy ions have special HLT menus. The highest-rate-paths in the CMS are isolated single muon/electron paths due to large production rate of $W \rightarrow e/\mu\nu$ in the LHC. The rates for common HLT paths used at CMS at 13 TeV are shown in Table 2.1.

Description	Condition	Rate
Isolated single muon	$p_T(\mu) > 24 \text{ GeV}$	235 Hz
Isolated single electron	$p_T(e) > 32 \text{ GeV}$	165 Hz
Non isolated single muon	$p_T(\mu) > 50 \text{ GeV}$	46 Hz
Non isolated single electron	$p_T(e) > 115 \text{ GeV}$	17 Hz
Isolated diphoton	$p_T(\gamma) > 30/22 \text{ GeV}, M(\gamma\gamma) > 90 \text{ GeV}$	40 Hz
Isolated ditau	$p_T(\tau) > 35/35 \text{ GeV}, \eta(\tau) < 2.1/2.1$	40 Hz
Isolated dielectron	$p_T(e) > 23/12 \text{ GeV}$	25 Hz
Isolated dimuon	$p_T(\mu) > 17/8 \text{ GeV}, M(\mu\mu) > 3.8 \text{ GeV}$	28 Hz
Isolated electron-muon	$p_T(e) > 23(12) \text{ GeV}, p_T(\mu) > 8(23) \text{ GeV}$	7.5 (4) Hz
Single jet	$p_T(j) > 500 \text{ GeV}$	11 Hz
Hadronic transverse energy	$H_T > 1050 \text{ GeV}$	10 Hz
Missing transverse energy	$\text{PFMET} > 120 \text{ GeV}, \text{PFMHT} > 120 \text{ GeV}$	33 Hz
Hadronic $t\bar{t}$	$H_T > 380 \text{ GeV}, \geq 6 \text{ jets } (p_T > 32 \text{ GeV}), 2 \text{ b-tagged jets}$	9 Hz
Boosted heavy jets	$p_T(j) > 400 \text{ GeV}, M(j) > 30 \text{ GeV}$	27 Hz
Isolated single photon	$p_T(\gamma) > 110 \text{ GeV}, \eta(\gamma) < 1.479$	12 Hz
Non isolated single photon	$p_T(\gamma) > 200 \text{ GeV}$	13 Hz
Triple muon	$p_T(\mu) > 5/3/3 \text{ GeV}, M(\mu\mu) > 3.8 \text{ GeV}$	9 Hz
isolated dimuon+electron	$p_T(\mu) > 4 \text{ GeV}, p_T(e) > 9 \text{ GeV}$	4.5 Hz
Displaced $J/\psi \rightarrow \mu\mu$	$p_T(\mu) > 4/4 \text{ GeV}, 2.9 < M(\mu\mu) < 3.3 \text{ GeV} + \text{displaced vertex}$	33 Hz

Table 2.1: HLT paths with their rate for an instantaneous luminosity of $1.8 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$ at 13 TeV and a pile-up of 50. The rate uncertainties are of the order of few Hz [65].

⁸The 1 kHz is a typical HLT rate. However space on tapes that events are saved to allow to record up to 2 kHz data and DAQ have an even higher limit of about 6 kHz, while the limiting factor is performance of the offline reconstruction. Therefore, the 1 kHz limit can be extended by saving only reconstructed objects information (no raw data, "data scouting") or by saving full events without the offline reconstruction to be reconstructed during technical stops ("parking").

The CMS tracker is the source of trigger primitives used at the HLT. The tracker hits and/or reconstructed track can validate the calorimeter or muon object, improve precision of p_T and isolation measurements or help to discriminate pile-up events by giving primary vertex location. All this results in reduction of rates and backgrounds at the output of the HLT.

Chapter 3

Offline object reconstruction

The CMS is an exemplary "hermetic detector" (or " 4π detector"), which means that it is designed to observe all decay products (except neutrinos) of beam-beam interaction in a collider. The particle reconstruction and identification strategy is based on the feature that particles of different type interact differently in consecutive subcomponents of the detector. In Fig. 3.1 sketch of the interactions with detector matter is presented for basic particle types.

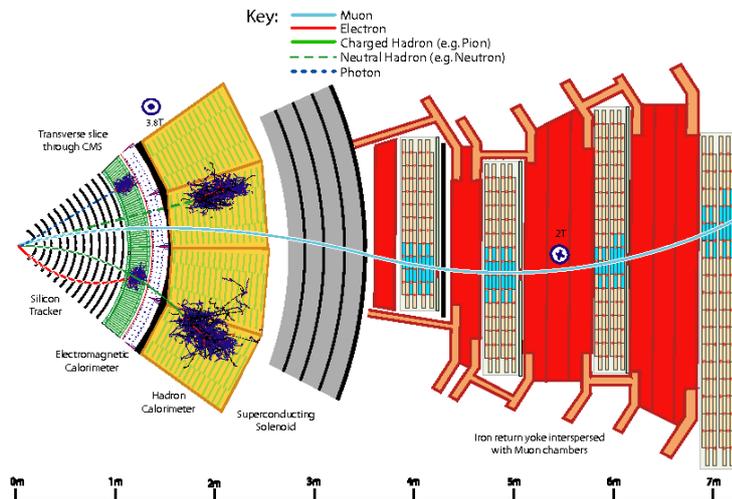


Figure 3.1: The transverse view of the CMS detector with specific particle interactions. Antimuon (μ^+), charged pion (π^+), electron (e), photon (γ) and neutron (n) are shown in this example [66].

The algorithmic representation of this idea, which the CMS collaboration has decided to implement a few years before the first LHC run, is called the Particle Flow algorithm (PF). In this approach, the global event description is provided with a comprehensive list of final-state particles for each event. The outstanding CMS performance in terms of jet and hadronic tau reconstruction, missing transverse momentum (\vec{MET}) determination¹ and lepton identification stems directly from utilization of the PF algorithm. The quality of the detector is a necessary requirement to be able to use the PF algorithm. As we will later see, the PF relies on reconstructing particles from the so-called "blocks", which are created on the topological basis and

¹In the CMS, for every event, we are reconstructing missing transverse momentum as a vector with two nonzero components (perpendicular to the beam axis, MET_x and MET_y). The absolute value of this vector is called *missing transverse energy* (MET).

are very sensitive to the tracking inefficiencies, requiring hermetic and fine-grained calorimeters and a highly segmented muon system.

3.1 Particle Flow algorithm

The Particle Flow algorithm provides a list of individual particles, namely: electrons, muons, photons, charged and neutral hadrons. Those can later be used to build jets, determine MET, reconstruct and identify hadronic taus, determine isolation, etc. The adopted workflow is the following:

- The tracks are reconstructed and the ECAL and HCAL clusters are built (separately).
- The "blocks" are defined by linking tracks and clusters.
- The PF-particles are created from blocks.

In the particular case of muons, their identification is based on the signals from the CMS muon detectors and relies on the tracks matching between the tracker and the muon system.

Tracker tracks reconstruction

The tracks reconstruction process begins with reconstruction of *hits* in pixel and strip detectors. In both cases, the zero-suppressed signals from data acquisition system are clustered and the cluster position and its uncertainty is calculated in a local coordinate system of the sensor. Translation of the hit into the global coordinate system is combined with introducing corrections to the actual location of detector elements and surface deformation found in the alignment process.

The tracker hits (local reconstruction) are an input to the track reconstruction (global reconstruction), which returns particle position and momentum estimates. Two basic ideas are utilized in this process: 1) CMS version of combinatorial Kalman filter called *Combinatorial Track Finder (CTF)* [67] and 2) the *iterative tracking* [57] method.

CTF is an algorithm designed to reconstruct tracks from the hits in three steps. In the first step, the very short trajectories of charged particles called *seeds* are generated using a few pixel hits² (in simplest case). Seeds are required to satisfy some restrictions (e.g. on momentum or position) to limit the total number of created seeds. In the second step, the track candidates are built by gathering the hits from successive layers of the tracker. For every track candidate this procedure starts with seed and locates the compatible hits using the projection of initial trajectory. At every layer the track candidate parameters are updated. In the third step of CTF, the final track candidate is used to determine origin, transverse momentum and direction of charged particle.

The sole use of CTF would end up in either low efficient tracking (by requiring high- p_T tracks and/or with low number of missing hits) or large fake-rate (without any special requirements the algorithm will create a lot of tracks by correlating

²The seeds are build from inside to outside of the tracker. This particular choice is backed up by a few reasons: 1) the pixel tracker has a very good efficiency (85% of simulated $t\bar{t}$ events leave hits in at least 3 pixel layers [57]); 2) the occupancy of pixels is typically much lower than strips due to large granularity; 3) the bremsstrahlung radiation and inelastic interactions of particles do not affect the reconstruction efficiency.

hits originating from different particles). To go around this problem, the CTF algorithm is applied multiple times with a different parametrization (seeds/tracks requirements) and to a different set of hits. More specifically, the very tough starting quality requirements on tracks are loosened from iteration to iteration. Simultaneously the initially complete set of hits shrinks with every iteration by masking the hits used previously. This way it is possible to build up efficiency with consecutive iterations keeping overall fake-rate low, i.e. high purity. The seeding configuration in the CMS iterative tracking method is presented in Table 3.1.

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

Table 3.1: Tracking iterations with seeding configuration and targeted tracks as used in PF reconstruction during LHC Run II. R is the distance between the track production position and the beam axis [66].

Electron tracking

Since electrons are charged and stable particles we could imagine that their tracks are perfectly well reconstructed with default iterative tracking method. This is indeed true for nonradiating electrons for which the tracking efficiency is the same as for muons. However, electrons have a large cross-section on the bremsstrahlung photon radiation inside the tracker. If the emitted photon is energetic, the Kalman filter based pattern recognition can fail because a kink in the electron trajectory occurred at the emission point. Therefore, the CMS Collaboration developed a dedicated electron reconstruction algorithm based on the Gaussian Sum Filter (GSF) track fitter, where the electrons are seeded using both ECAL measurements (*ECAL-based approach*) and output of iterative tracking (*tracker-based approach*). As a result, the dedicated collection called GSF electrons tracks (see Fig. 3.2) is introduced into the CMS event reconstruction.

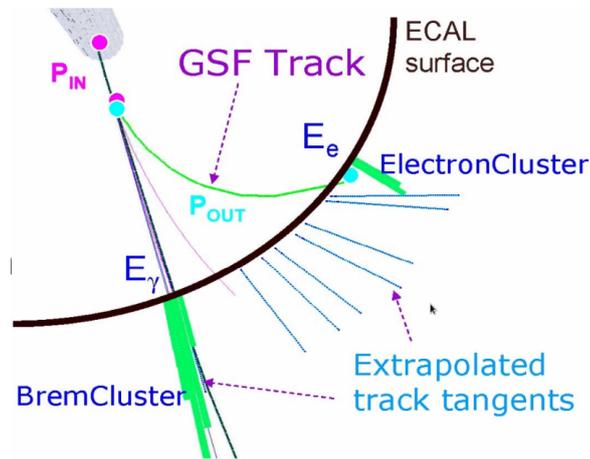


Figure 3.2: Illustration of GSF reconstruction [68].

In the ECAL-based approach the electron seed is given by the crystal containing most of the energy deposited in the considered region. If the electron seed is found, the cluster energy and position are used to infer the position of the hits in the tracker layers. The ECAL-based approach is effective for energetic and well-isolated electrons and can work fine even for cases where bremsstrahlung photons are emitted

(if they fit into ECAL cluster window). ECAL-based approach is not effective for nonisolated electrons (energy and position of electron are biased by contributions from other particles in the jet) and for low-energy ones (the electron energy is underestimated as it radiates outside the cluster window due to large bending in the magnetic field of the CMS). The electrons missed by the ECAL-based method are often recovered using tracker-based approach where the electron seed is defined by the tracker track. The track can be refitted with GSF filter if it has many missing hits and its momentum is not compatible with ECAL cluster energy.

Muon tracking

The muon track can be reconstructed in the CMS inner tracker and/or outer muon spectrometer. The former has an advantage of very precise muon momentum measurements, the latter of high reconstruction purity thanks to calorimeters in front of it absorbing the vast majority of particles other than muons (and neutrinos). In the CMS the collection of muon objects is composed of three muon types: 1) tracker muons, 2) standalone muons and 3) global muons. *Tracker muons* are related to the reconstructed tracker tracks having extrapolation matched with at least one active muon segment. Tracker muons are efficient in the reconstruction of low- p_T muons for which multiple scattering in the return yoke affects muon system resolution. *Standalone muons* are the muons reconstructed only with hits in the muon system. Standalone muons usually have a worse p_T resolution and a higher cosmic muon contamination versus tracker and global muons. *Global muons* are the muons reconstructed by refitting hits from two compatible tracks from tracker and muon system. Global muon reconstruction is especially efficient for high- p_T muons as it requires at least two coincided layer hits in muon spectrometer.

Calorimeter clusters

The clusterization is a process designed to generate information about the energy and direction of the particles interacting within the calorimeters volume. It is deployed separately in each calorimeter subdetector and comprises of four steps:

- Identification of the cluster seeds. From all calorimetric cells surpassing energy (and transverse energy in ECAL due to significant noise) threshold(s) the ones with the largest energy deposit among the neighboring cells are taken as seeds.
- Identification of the topological clusters. In this procedure the cells surrounding the cluster are aggregated starting from the seeds. The cell energy needs to pass certain threshold requirements and has to have at least one corner in common with the cluster.
- Identification of the clusters. The clusters are searched inside topological clusters using a Gaussian-mixture model. The basic idea here is to assume that seeds in the frame of topological clusters are the source of Gaussian energy deposits and that only those deposits are set off in all the cells within the topological cluster. The model parameters are fitted using maximum-likelihood method iteratively until convergence, where the seeds position and energy are used for initial model evaluation.
- The clusters parameters (positions and energies) are taken directly from the final shape of the Gaussian functions.

The energy calibration procedure plays a crucial role in the clusterization with regard to PF reconstructing. This is mainly because the PF particle identification relies on comparison of cluster energy with the compatible tracker track momentum. The possible discrepancies can be interpreted as signal from neutral particles. The calorimeter cluster calibration procedure uses test beams, radioactive sources, cosmics and collision data (using photons and kaons). It is executed separately for the ECAL and HCAL and takes into account threshold conditions, energy and pseudorapidity dependence and, in case of HCAL, also the already performed ECAL calibration (for more details see [66]).

Link procedure and PF-particles reconstruction/identification

The reconstructed tracks and clusters are elements of *blocks*. The creation of blocks takes place for neighboring elements on the (ϕ, η) (or (x, y)) plane for barrel (or end-cap) and consists of linking the elements with the following procedure:

- *The tracker track to calorimeter cluster link.* The cluster is linked to a track if it lays on the track prolongation. The track (last measured hit) is extrapolated: to the two first layers for the preshower; to the expected maximum of the longitudinal electron shower for the ECAL; to the depth of one interaction length into the HCAL. In case of multiple linking (many HCAL clusters to a single track or many tracks to a single ECAL cluster), only the closest is kept.
- *The GSF tracks to ECAL cluster (bremsstrahlung photon) link.* The cluster is linked to the GSF track if it lays on any of the track extrapolations, made as straight lines (tangents) originating from the GSF track and tracker layer intersection (as shown in Fig. 3.2).
- *HCAL to ECAL and ECAL to preshower clusters links.* The cluster of preshower (ECAL) is linked to the ECAL (HCAL) if it lies in the cluster envelope of ECAL (HCAL). In case of multiple linking (many HCAL clusters to a single ECAL cluster or many ECAL clusters to a single preshower cluster), only the closest is kept.
- *Tracker tracks to tracker track links.* The tracks can be linked through common secondary vertex (SV). It is required that there are at least 3 tracks connected to SV from which only one leads to PV and that the outgoing tracks have invariant mass bigger than 0.2 GeV. This linking is meant to take into account nuclear interactions within tracker volume.
- *Tracker track to muon track link.* This link is established during the creation of global muons.

This algorithm does not guarantee to always link elements from one particle into a single block, as this is limited by the granularity of subdetectors.

Every block is a starting point for the same PF-particles reconstruction and identification analysis. At first, the global muons (with reconstructed momentum which is in agreement with momentum of compatible tracker muon) are identified as PF muons. The relevant tracks and ECAL energy clusters compatible with the muon hypothesis are removed from the block. In the second step, the PF electrons are identified using multivariate analysis techniques, where GSF tracks and ECAL clusters parameters serve as input parameters. The relevant tracks and ECAL clusters are again removed. The third step is to search for fake tracks: the tracks with relative p_T uncertainty larger than relative calorimetric energy resolution are removed.

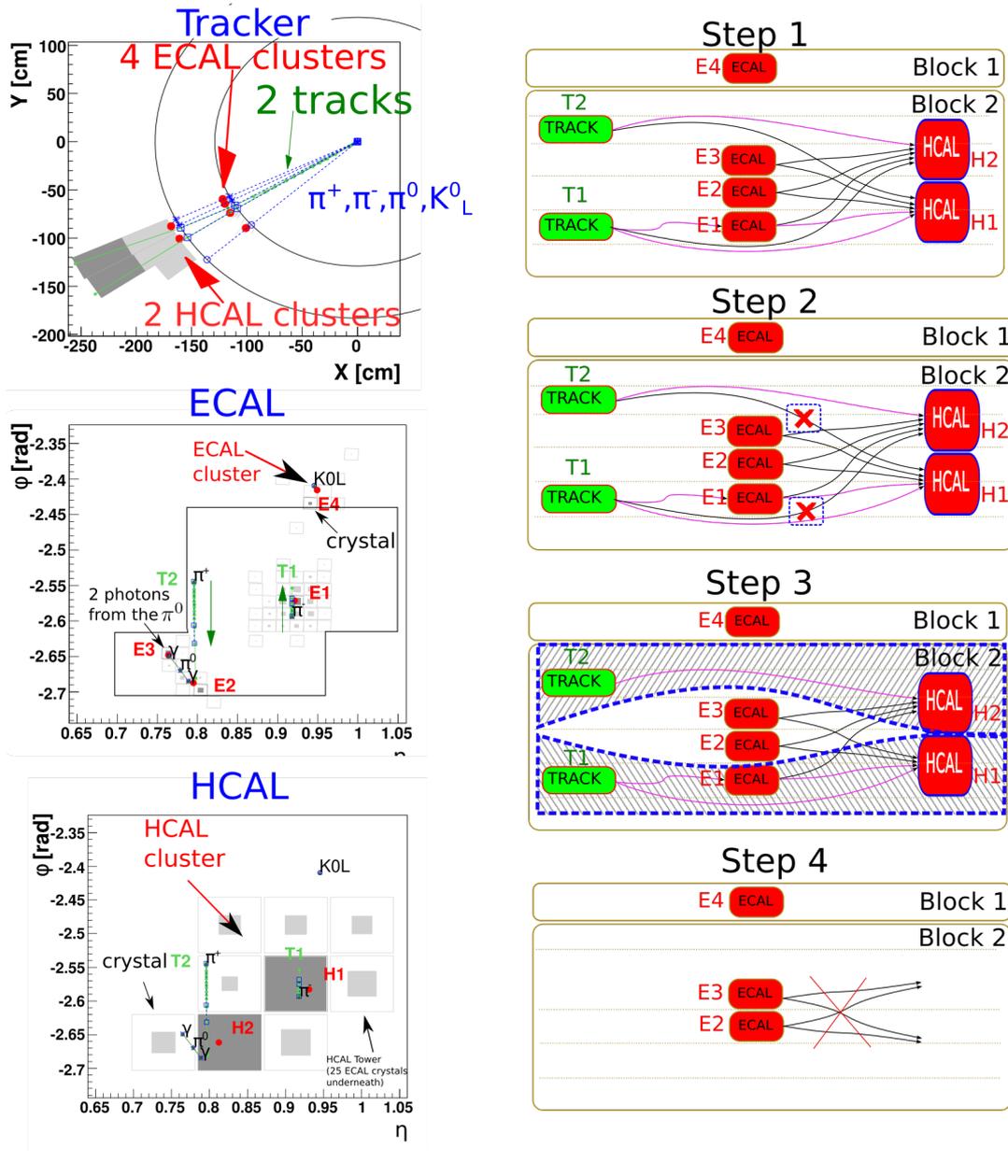


Figure 3.3: An exemplary event reconstructed with Particle Flow algorithm. The simulated particles are: neutral kaon (K_L^0), neutral pion (π^0) and two oppositely charged pions (π^\pm). On the left-hand side different CMS planes are presented. The *top-left* plot is an (x, y) view, where two circles represent the ECAL and HCAL surfaces. The *middle-left* and *bottom-left* plots are a (ϕ, η) planes of the ECAL and HCAL surfaces, respectively. The gray border on the middle-left plot marks the HCAL topological cluster envelope projection. On all plots the green lines represent (the projections of) the reconstructed tracks, the gray areas represent the energy deposits in crystals (more gray means more energy is deposited) and the red dots indicate the position of the reconstructed clusters. On the right-hand side, the PF reconstruction steps are outlined. In the first step, the blocks are defined by linking tracks with clusters. In the second step, tracker-HCAL links are cleaned. Then the charged hadrons (π^\pm) are reconstructed from block 2. In the last step two block 2 clusters are identified as photons (from $\pi^0 \rightarrow \gamma\gamma$ conversion) and the block 1 cluster is identified as a neutral kaon.

Left hand side figures are adopted from [69].

Next, all tracker to HCAL links are removed from the block except the closest ones. Then, the connected tracks and ECAL clusters are considered. The sum of energies of the ECAL clusters linked to a single track is calculated, starting from the closest cluster. All links are kept till the sum of the energy does not exceed the track transverse momentum and the remaining links are removed. If after the previous step the sum of the ECAL clusters energy connected to the tracker track is considerably smaller than the tracks p_T , then additional muon and fake track search is triggered. In the last two steps the remaining block elements are used to identify hadrons and photons. First the PF charged hadron, PF neutral hadron or PF photon are extracted by comparing the (summed) track momentum and the calibrated calorimeter energy. At last, the PF neutral hadrons and PF photons are created using remaining clusters (not linked to any track). The reconstructed PF particles let us to determine the missing transverse energy $\vec{E}_{T,PF}$ as a negative vectorial sum of all PF particles transverse momenta $\vec{p}_{T,i}$:

$$\vec{E}_{T,PF}(\text{raw}) = - \sum_{i=1}^{N_{\text{particles}}} \vec{p}_{T,i}. \quad (3.1)$$

A more detailed description of the procedure above and the Particle Flow algorithm performance plots can be found in Ref. [66]. The exemplary PF event reconstruction is shown in Fig. 3.3.

3.2 Jet reconstruction

The PF particles are clustered to reconstruct jets. The standard method employed in the CMS experiment to achieve this starts with defining *protojets*. A protojet can be an individual particle or a bunch of overlapping, particles and is characterized by its transverse momentum \vec{k}_i , transverse energy $E_i = |\vec{k}_i|$, rapidity y_i and azimuth ϕ_i , where i is a protojet enumerator. Having a set of protojets we can define the distances: 1) $d_{iB} = k_i^{2p}$ - the distance between a single protojet and the beam and 2) d_{ij} - the distance between the protojets defined as:

$$d_{ij} = \min(k_i^{2p}, k_j^{2p}) \frac{(y_i - y_j)^2 + (\phi_i - \phi_j)^2}{R^2}, \quad (3.2)$$

where the parameter R is the cone size parameter equivalent and should be of order 1, and p is an arbitrary integer governing the relative power of the energy versus geometrical scales³. Now the jets can be extracted recursively by finding the minimum of all the protojets distances. If the minimum found is a d_{iB} -type, then protojet i is a jet and is removed from the protojets list. If the minimum found is a d_{ij} -type, the protojets i and j are merged to create a single protojet, which is added to the protojets list in place of the original ones. This procedure is repeated until all jets are found and the protojets list is empty.

It can be shown that natural $p = 1$ choice is not optimal and that algorithm $p = -1$ gives a fast, collinear and an infrared safe⁴ way to extract the hadronic jets [71]. This latter choice is the default CMS jet clustering algorithm and is called *anti- k_T algorithm*. Both cases are compared for a simple event scenario in Fig. 3.4.

³Typically $R=0.4$, as was mentioned during the L1 calorimetric trigger description.

⁴The algorithm is infrared safe if any infrared singularities do not appear in the perturbative calculations and found solutions are insensitive to soft radiation in the event [70].

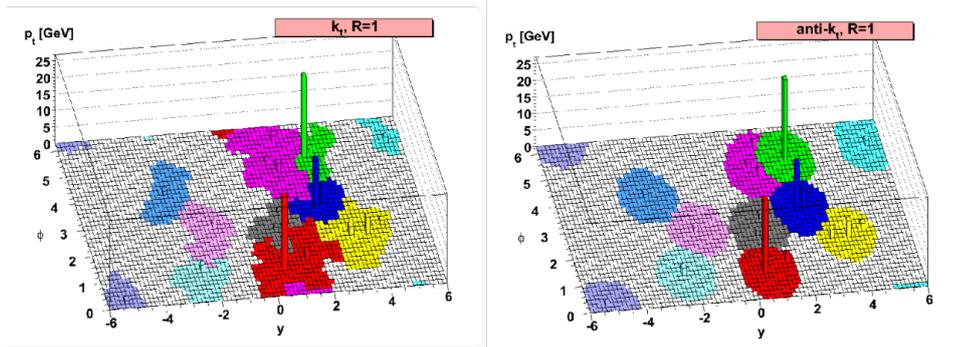


Figure 3.4: A comparison between k_T (left) and anti- k_T (right) algorithms reconstruction behavior on a sample event [71]. The color regions indicate which random soft ghost particles (out of $\sim 10^4$ present in the event) are clustered into a jet. k_T algorithm is somewhat more dependent on the specific set of ghost.

The jet energy found turns out to be biased versus the real/generated jet energy. Therefore, after identification every jet undergoes the *Jet Energy Correction (JEC)* procedure, which in its basic form is composed of three levels [72]:

- $L1^5$: *correction for average offset-energy from pile-up*. The pile-up, i.e. presence of multiple proton-proton collisions within a single bunch crossing, is a source of tracks and calorimeter deposits. The L1 correction is introduced to mitigate these contributions in jet reconstruction and proceeds in steps designed to take care of different types of signals. The calorimeter deposits echoed after prior collisions and existing because of finite temporal calorimeter resolution are removed with calorimeter signal processing. The pile-up charged-hadrons are subtracted by exploiting the tracker resolution and tagging particles not originating from primary hard interaction. Then the pile-up jet identification (PUJetID) is used to remove those jets. Finally, the pile-up neutral particles diffuse energy is estimated per event and subtracted from the jets.
- $L2L3$: *correction for (p_T, η) dependence of jet response*. Since the CMS detector is a cylinder and not an isotropic ball, it will respond differently for the same jet sent in two different directions. In L2L3 correction the simulated particle response is defined as:

$$R_g(\langle p_T \rangle, \eta) = \frac{\langle p_T \rangle}{\langle p_{T,g} \rangle} [p_{T,g}, \eta] \quad (3.3)$$

where p_T and η are defined for anti- k_T reconstructed L1-corrected jets and g indicates generated particle-level jets. The R_g is evaluated in $(p_{T,g}, \eta)$ bins (see Fig. 3.5). The L2L3 correction relies exclusively on simulated events as it simplifies the coverage of the phase space corners (very low- and very high-transverse momenta and pile-up events).

- $L2L3Residuals$: *correction for residual data-simulation differences (applied to data only)*. This correction complements the previous one and is devoted to effects which are not precisely simulated. The L2L3Residual corrections are also defined in (p_T, η) bins and modeled with tag-and-probe on the dijet events. The

⁵L1 correction has nothing to do with L1 trigger and the similarity of names is coincidental.

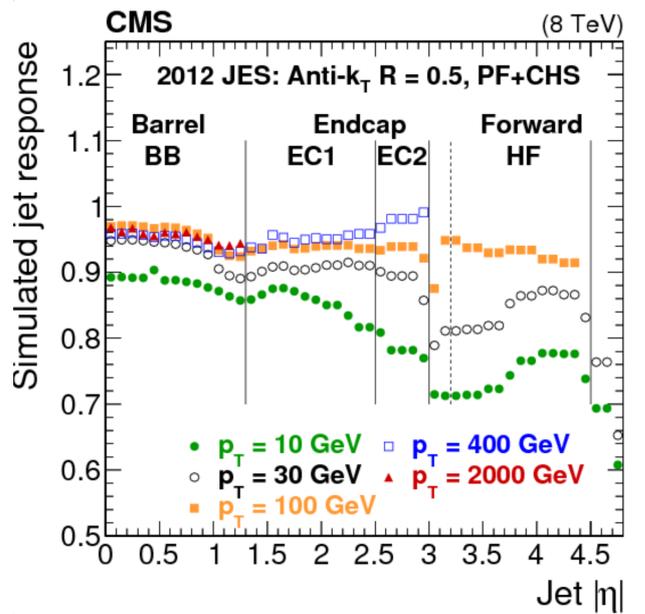


Figure 3.5: The simulated jet response R_g (after JEC applied) as a function of the reconstructed jet pseudorapidity η for various values of jet transverse momentum [72].

jet energy response is figured out by studying transverse momentum imbalance between reference (tag) jet and the jet to be calibrated (probe).

3.3 Tau reconstruction

The branching ratio for the tau decaying into hadrons is about 65%, with final state predominantly containing either one or three charged mesons and up to two neutral pions.

The electrons and muons coming from the leptonically decaying taus are reconstructed by standard techniques [73, 74, 75, 76], and taken as they are. The reconstruction of hadronically decaying taus⁶ (τ_h) is more involving and starts with anti- k_T jets, which are used as seeds for dedicated Hadron-Plus-Strip (HPS) algorithm [77]. As π^0 decays virtually instantaneously into γ particles, which in turn can convert into an e^+e^- pair, the HPS algorithm tries to reconstruct π^0 s from photon and electron constituents of the jet first. All energy deposits in electromagnetic calorimeter within a $0.05 [\eta] \times 0.20 [\phi]$ window are clustered into a "strip" (see Fig. 3.6). The strip is enlarged in ϕ -direction to account for the e^+e^- pair bending in the magnetic field of the detector. The special treatment is applied for low- p_T electrons, for which the e^+e^- pair can go outside the strip and produce an energy deposit in the isolation region of the tau jet, causing isolation cuts (< 2 GeV) to fail.⁷ For such cases the original strip window size can be dynamically changed (enlarged) with electron p_T .

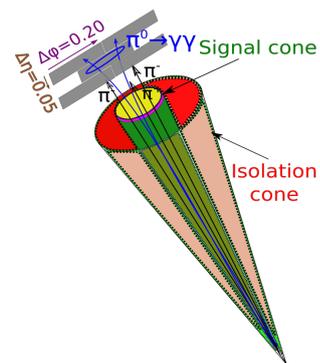


Figure 3.6: Tau jet

⁶Such decays with hadrons will be called *hadronic*, even if this is not strictly correct because of existence of neutrons in final state (weak decay).

⁷This effect is more pronounced at higher- p_T taus, as the decay product usually has higher transverse momentum.

For each jet charged hadrons (h) and ECAL strips are combined to reproduce the tau decay mode. The following topologies are considered: (1) single hadron (relevant to $\tau \rightarrow h\nu_\tau$ decays); (2) hadron plus strips (relevant to $\tau \rightarrow h\pi^0\nu_\tau$ and $\tau \rightarrow h\pi^0\pi^0\nu_\tau$ decays); (3) three hadrons (relevant to $\tau \rightarrow hhh\nu_\tau$ decays); (4) three hadrons plus strip (relevant to $\tau \rightarrow hhh\pi^0\nu_\tau$); (5) two hadrons (with or without additional strips).

Lp.	Decay Mode
0	1prong
1	1prong + π^0
2	1prong + $2\pi^0$
5	2prong
6	2prong + π^0
7	2prong + $2\pi^0$
10	3prong
11	1prong + π^0

Table 3.2: Hadronic tau decay modes

The 2-prong topology (with or without π^0 s) is used to care of tracking inefficiencies of the detector. The fourth and fifth topologies add up to about 3–4% of overall reconstruction efficiency and are especially important for high energy taus for which the jet to tau fake rate is low.

For every reconstructed tau the HPS algorithm attributes the *decay mode* in form of integer between 0 and 11. The decay mode number is defined as $5 \times (N_c - 1) + N_p$, where N_c is the number of signal charged hadrons and N_p is the number of signal π^0 s (strips). Table 3.2 presents available decay modes.

The tau reconstruction is followed by identification designed to separate hadronic taus from quark and gluon jets and from other flavor leptons [78]. In this step, it is required for the tau to pass cut-based isolation defined by the formula:

$$I_\tau = \sum P_T^{\text{charged}}(d_Z < 0.2 \text{ cm}) + \max(P_T^\gamma - \Delta\beta \sum P_T^{\text{charged}}(d_Z > 0.2 \text{ cm}), 0),$$

where the two right-hand side terms are the transverse momentum sums of charged particles (originating from primary vertex within a distance of 0.2 cm and within a $\Delta R = 0.5$ isolation cone, excluding hadrons used for reconstruction) and photons, respectively. The $\Delta\beta$ correction factor, equal to 0.2, accounts for overestimation of neutral energy deposits due to pile-up. Its value corresponds to the ratio of neutral to charged pion production rate (0.5), corrected for the adopted isolation cone size difference for charged hadrons originating from the primary vertex (0.5) and from pile-up (0.8). Loose, medium and tight isolation working-points are defined as I_τ cuts of 2.5, 1.5 and 0.8 GeV [79]. Additionally to the isolation sum discriminant, every tau is probed with a multivariate (MVA-based) discriminants combining isolation with available life time and tau shape information. The expected jet to tau misidentification probability as a function of expected hadronic tau identification efficiency is presented in Fig. 3.7.

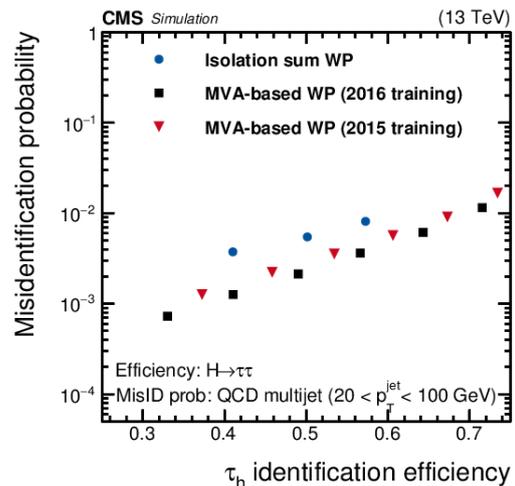


Figure 3.7: Misidentification probabilities for hadronic tau as a function of their identification efficiency, evaluated using $H \rightarrow \tau\tau$ and QCD multijet MC events [79].

3.4 Higgs mass reconstruction in decays into a pair of taus

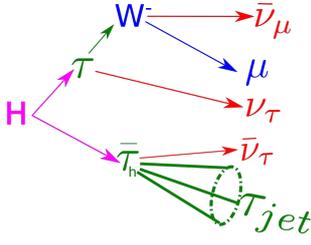


Figure 3.8: $H \rightarrow \tau_\mu \tau_h$

The calculation of the mass for the two-body Higgs decay into taus is in principle straightforward. For center-of-mass reference frame the mass of Higgs is simply the invariant mass of the two taus:

$$M_H^{CM} = E_{\tau_1} + E_{\tau_2} = 2(\sqrt{m + p}). \quad (3.4)$$

The problem can be therefore narrowed down to the determination of the momentum of the tau leptons, i.e. momentum of the lepton + two neutrinos and/or hadronic tau jet and one neutrino (depending on the tau decay mode, e.g. see Fig. 3.8). The momenta of the lepton or hadronic tau are given as an output of the PF or HPS algorithms. However, the momenta carried out by two to four neutrinos are clearly underconstrained: only E_x and E_y variables are known. This problem is addressed by the SVfit algorithm, which tries to estimate it using statistical maximum likelihood estimation (MLE) method. The MLE allows to find unknown parameters x_u for a given model by using (conditional) probability density function f to define the likelihood function \mathcal{L} , and then maximizing it. For a given ditau mass $M_{\tau\tau}$ we can write (δ stands for Dirac delta function):

$$\mathcal{L}(M_{\tau\tau}) = \int \frac{df(x_u|x_m)}{dx_u} \delta(M_{\tau\tau} - M_{\tau\tau}(x_u, x_m)) dx_u. \quad (3.5)$$

This equation simply means that the likelihood function is a weighted sum over all parameters compatible with measured x_m values and assumed mass value. In the SVFit the calculations are repeated for a range of mass values $M_{\tau\tau}^i$ and the one corresponding to the maximum of $\mathcal{L}(M_{\tau\tau}^i)$ is the Higgs mass we are looking for.

To calculate $\mathcal{L}(M_{\tau\tau})$ the parameters (x_u, x_m) and the integration range has to be known.

Let us consider the kinematics of the tau. If we split the tau energy E_τ into visible and invisible part and introduce m_{inv} as the mass of the neutrino pair produced in leptonic tau decays (m_{inv} is equal to zero for hadronic tau decays by convention) we get:

$$E_\tau = E_{vis} + E_{inv} = E_{vis} + \sqrt{|\vec{p}_\tau^2 - \vec{p}_{vis}^2| + m_{inv}^2}$$

hence:

$$m_\tau^2 + \vec{p}_\tau^2 = E_\tau^2 = (E_{vis} + E_{inv})^2 = \left(E_{vis} + \sqrt{\vec{p}_\tau^2 - 2|\vec{p}_\tau||\vec{p}_{vis}|\cos\theta + \vec{p}_{vis}^2 + m_{inv}^2} \right)^2$$

Solving the equation above with respect to the angle θ (known in literature as the *Gottfried-Jackson angle*) we obtain:

$$\theta = \arccos \frac{m_{inv}^2 - m_\tau^2 - m_{vis}^2 \pm 2E_\tau E_{vis}}{2|\vec{p}_{vis}||\vec{p}_\tau|}, \quad (3.6)$$

where $m_{vis}^2 = E_{vis}^2 - p_{vis}^2$. Here, tau lepton energy E_τ is given by formula $E_\tau = \frac{E_{vis}}{X}$, where X is the fraction of tau energy carried out by the visible decay products. The variables X and m_{inv} constrain the tau momentum vector to lie on the surface of the cone (Fig. 3.9).

To determine relative position between \vec{p}_{vis} and \vec{p}_τ in the laboratory frame two angles are needed: ϕ (which is a free parameter that gives the angular coordinate of the tau lepton momentum vector around the cone axis) and θ (which is given by Eq. 3.6). Since all remaining variables are known we have: $x_m = \{m_\tau, m_{vis}, E_{vis}, \vec{p}_{vis}, \theta\}$ and $x_u = \{X, \phi, m_{inv}\}$.

In the SVfit algorithm the probability density function is factorized:

$$\frac{df(x_u|x_m)}{dx_u} = \frac{d\Gamma_1(x_u^1|x_m^1)}{dx_u^1} \times \frac{d\Gamma_2(x_u^2|x_m^2)}{dx_u^2} \times f_{MET},$$

where the first two terms are proportional to the matrix elements for two taus decays and the third quantifies the compatibility of parameters with measured E_x and E_y values. In the case of unpolarized taus:

$$\frac{d\Gamma(x_u|x_m)}{dx_u} \propto \frac{1}{2\pi} \left(\frac{1}{1 - \frac{m_{vis}^2}{m_\tau^2}} \right),$$

$$\frac{d\Gamma(x_u|x_m)}{dx_u} \propto \frac{m_{inv}}{4m_\tau^2} ((m_\tau^2 + 2m_{inv}^2)(m_\tau^2 + m_{inv}^2)),$$

$$f_{MET} = \frac{1}{2\pi\sqrt{|V|}} \exp \left\{ -\frac{1}{2} \begin{pmatrix} E_x - \sum p_x^v \\ E_y - \sum p_y^v \end{pmatrix}^T \cdot V^{-1} \cdot \begin{pmatrix} E_x - \sum p_x^v \\ E_y - \sum p_y^v \end{pmatrix} \right\},$$

where the top term pertains to hadronically decaying taus, the middle term concerns leptonically decaying taus and in the bottom term V is a MET covariance matrix. The exhausting description of probability density function (e.g. for polarized decay case) can be found in the literature [80].

The boundary conditions for the integration are: $m_{vis}^2/m_\tau^2 \leq X \leq 1$ and $m_{inv} = 0$ for hadronic tau decays, $0 \leq X \leq 1$ and $0 \leq m_{inv} < m_\tau\sqrt{1-X}$ for leptonic tau decays. For angles we have: $-1 \leq \cos\theta \leq 1$ and $0 \leq \phi \leq 2\pi$.

The relative $m_{\tau\tau}$ resolution obtained with SVfit varies between 10% and 20% depending on considered tau decay modes (and boost of $\tau\tau$ system), and by itself improves the final expected significance of the order of 40% when compared to visible mass usage.

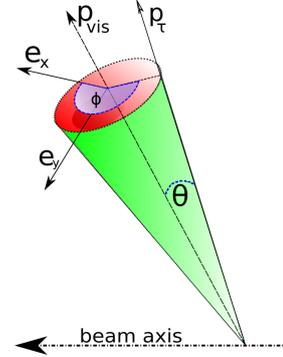


Figure 3.9: Tau cone

Chapter 4

The Higgs boson analysis in the ditau decay channel

In the framework of the Standard Model the Higgs boson can be produced in a hadron collider like the LHC mainly via gluon-gluon fusion (GGF), vector boson fusion (VBF) and by associated production with a weak-boson. The observation of the particle is performed in a statistical way throughout identification of its decay products and their kinematics, and is strictly correlated with our ability of background rejection. The Higgs boson was first observed in $H \rightarrow ZZ$ and $H \rightarrow \gamma\gamma$ decay channels [81], which are well reconstructed and do not have particles with color charge present in the final state.

Sensitivity in the double tau lepton channel is suppressed in a hadron collider although its branching fraction of about 6% is second highest for a 125 GeV SM Higgs boson. It is caused by a large QCD-background, which hinders the identification of hadronically decaying taus, and by neutrinos in the final state worsening the reconstructed mass resolution. Nevertheless, this particular channel is a valuable source of data for Higgs Physics studies for several reasons:

- decay into taus gives a chance to directly measure Yukawa coupling of the Higgs boson;
- the angular distributions of decay products of tau leptons can be used as a probe of CP nature of Higgs-to-tau couplings;
- in some Beyond SM scenarios, in particular large- $\tan\beta$ Minimal Supersymmetric Standard Model (MSSM), the branching ratio of Higgs resonances into taus can be significantly enhanced.

A complex $H \rightarrow \tau\tau$ analysis is also perfectly suited as a playground for various machine learning techniques. In this chapter specifically, the $H \rightarrow \tau_\mu\tau_l$ analysis will be presented.

Author's note

The official CMS analysis was lead and implemented by the CMS groups at University of Wisconsin Madison (USCMS) and Deutsches Elektronen-Synchrotron (DESY). Since the analysis is very complex, the groups agreed to share partial results and analysis ideas rather than actual implementation and the details of implementations are not available for viewing. Thus, the author made his own version of the analysis from scratch using only available generic CMS tools. The content of this chapter closely follows the author analysis implementation and does it in a manner that should in principle allow to reproduce all results. Hence, both order and language

of this chapter are more technical and less pedagogical than those adopted in the reference publication [82]. All plots sourced here from this official work are marked with bold **CMS** word on the top-left.

In this chapter all *emphasized words/phrases* will be given their clarification in Technical Appendix B. Also from now on by tau we usually mean hadronic tau and by lepton we mean an electron or a muon.

Analysis overview

The $H \rightarrow \tau\tau$ analysis follows the methodology presented in Section 1.3.3. There are three main steps in the analysis. First, pairs of high level objects (particles) are reconstructed (created), where every object corresponds to the single tau lepton final state. The Higgs boson signal extraction consists of reconstructing the ditau system mass which plays the role of discriminating variable. The overall low signal is enhanced by introducing a set of carefully selected kinematical cuts which mitigate the background. Almost all steps of this analysis can be attributed into one of six common categories:

- **Objects creation:** the data samples used are created centrally by the CMS Collaboration and are meant to serve multiple analyses (not only Higgs boson-related ones). As such, they do not contain all needed physical objects or variables in a form of ready-to-use data. Examples of the objects/variables we reconstruct from the data in the flow of analysis are Higgs boson candidate collection (p. 4.2), loose jet ID (p. 46), electron ID, generator matching information (p. 47), medium 2016 muon ID, event vetoes or muon isolation (p. 51). In this chapter we adopted a simple rule of thumb to write extensively only about objects created purposefully for this analysis.
- **Objects/events selection:** the opposite situation to the above one also happens, i.e. not all information available is required. The selection is three-fold: 1) we choose only samples with non-negligible contribution to signal phase space; 2) we reject uninteresting events from selected samples; 3) we remove not used information from the events. The samples selection is discussed in Section 4.1. Uninteresting events are removed due to many reasons, e.g. they do not contain any Higgs boson candidate or do not pass baseline selection (p. 50). The removal of information from the event (e.g. photons collection) is motivated by implementation reasons (less storage space occupied, faster code execution on smaller datasets) and done at the end of the first part of the analysis (see below).
- **Categorization:** the categorization is basically the superposition of events selection and clustering. The majority of events left after baseline selection are further rejected by categorization cuts. The remaining ones are subdivided into three detached sets called 0jet, boosted and vbf categories. The categorization is meant to boost signal-to-background ratio by fine-tuning the final selection in given category using information about anticipated Higgs production process (p. 53).

- Simulated data (MC) corrections: the Monte Carlo simulation is not perfect and must be corrected¹. Examples are tau energy scale (p. 49), recoil (p. 51), pile-up or lepton identification corrections (Section 4.4).
- Simulated (MC) to collision (Data) data matching. MC samples need to be normalized to the collected collision data luminosity (p. 57). They are then used for the background estimation (Section 4.6) and to create stacked Data-to-MC control plots (Appendix C).
- Systematic uncertainties calculation: the implementation of systematic uncertainties requires creating rescaled (up and down) distributions of discriminating variables which are input to the global fit. This is a complicated task and since it goes beyond the goals of this thesis is not included. The main sources of systematic uncertainties are however enumerated at the end of this chapter.

The whole analysis is split into two parts. In the first part (corresponding roughly to the first two points above, described in Section 4.2) samples are processed within the CMS SoftWare Framework (CMSSW) version 8_0_26, which allows for easy data files retrieval and CMS Event Data Model (EDM) handling, fast analysis modules (analyzers / producers / filters) creation, native and extended ROOT framework support, build in access to the Worldwide LHC Computing Grid (WLCG) via CMS Remote Analysis Builder (CRAB) utilities and more. The CMSSW – although being a capable and beautify designed system – is constrained by the computing environment as its installation outside CERN LXPLUS cluster is cumbersome, computing resources demanding and operationally slow, and is very limited in terms of compatibility with third-party software (including machine learning libraries). In the second part of the analysis (categorization, MC corrections and data matching) we go around these issues by making a custom code written independently of the CERN/CMS infrastructure.

4.1 Samples

The *data samples* used in this analysis are a subset of the proton–proton data collected in 2016 at center of mass energy of 13 TeV. Only *certified (high quality) runs* from LHC data taking periods 2016B to 2016H are used with total luminosity of 35.9 fb^{-1} . The average pile-up was about 27 interactions per bunch crossing.

The simulated signal Monte Carlo (MC) samples for gluon fusion (ggH), vector boson fusion (VBF) and with associated W^\pm or Z bosons production modes are generated with POWHEG 2.0 BOX program [83], where the NLO QCD calculations can be matched with parton shower generators. As mentioned in Chapter 1 the $t\bar{t}H$ process samples are skipped. The cross sections and branching fractions for the Higgs boson production are taken from [25] and [84].

The simulated background MC samples are generated with MADGRAPH [85] or POWHEG BOX programs. The former is used at LO for Z + jets and W + jets processes and at NLO for diboson production. The latter for $t\bar{t}$ (version 2.0) and single top (version 1.0) quark production. Please note that the dedicated Drell-Yan samples with low (below 50 GeV) and high (above 150 GeV) invariant *mass of Z boson* are not used because of negligible contribution to the signal region. Parton showers, hadronization and hadronic taus decays are modeled with PYTHIA8 [86] using underlying event tune CUETP8M1 [87].

¹We introduce new corrections only to MC events. The only corrections discussed in context of collision data are Jets (p. 46) and Type-I MET (p. 47) corrections, which are reapplied.

Simulated events are reconstructed in the same way as events from collisions with exception of being additionally processed with CMS detector simulator based on GEANT4 [88]. Both Data and MC datasets, together with cross sections, are summarized in Appendix A.

A Global Tag is a collection of records of data needed by the reconstruction and analysis software but not included in data samples, like alignment, calibration, temperature or parameters for the simulation software. The Global Tag is defined for collision and simulated data separately and for each data-taking period. Most physics objects reconstructed in data are already calibrated and ready-to-use and do not require additional corrections. The notable exceptions are jet energy corrections and trigger configuration information which are sensitive to the Global Tag version [89]. The *global tags* used in this analysis are given in Appendix B.

4.2 Data pre-processing

Pairs creation

The analysis starts with *defining the Higgs boson candidates* as pairs of *muons and taus*. The muons are required to pass $p_{T,\mu} > 20$, $|\eta_\mu| < 2.1$ and taus to pass $p_{T,\tau_h} > 27$, $|\eta_\tau| < 2.3$ and have absolute value of charge equal to unity ($|q_\tau| = 1$). Moreover, a muon and a tau in the pair have to be separated in the $(\eta - \phi)$ -plane:

$$dR := \sqrt{(\eta_{\tau_h} - \eta_\mu)^2 + (\phi_{\tau_h} - \phi_\mu)^2} > 0.5,$$

where ϕ in radians. Please note that there can be zero, one or more pairs per event. Also note that the Higgs boson candidate is created from the muon collection; in other words there is no dedicated collection of taus decaying into muons (or electrons to be precise).

Jets corrections and selection

Jets collection available in data contain PF-jets reconstructed with anti- k_T algorithm with $\Delta R = 0.4$ cone size. Although the jet energy corrections are by default applied to collision and MC events, we *re-apply them* for data taking periods 2016B - 2016G samples due to updated Global Tag (see Appendix B).

Only jets with $|\eta| < 4.7$, $p_T > 30$ GeV and passing so called *loose jet ID* Working Point (WP, criteria are presented in Table 4.1) are kept for further analysis. Loose jet ID is meant to reject badly reconstructed, fake or noise jets. The fake-rate in jets passing *loose jet ID* is 1-2%.

Variable	$ \eta_j \leq 2.4$	$2.4 < \eta_j \leq 2.7$	$2.7 < \eta_j \leq 3.0$	$ \eta_j > 3.0$
Neutral Hadron Energy Fraction	< 0.99	< 0.99	–	–
Neutral Electromagnetic Energy Fraction	< 0.99	< 0.99	< 0.9	< 0.9
Charged Hadron Energy Fraction	> 0	–	–	–
Charged Multiplicity	> 0	–	–	–
Charged Electromagnetic Energy Fraction	< 0.99	–	–	–
Neutral Multiplicity	–	–	> 2	> 10
Number of Constituents	> 1	> 1	–	–

Table 4.1: Loose jet ID criteria. In columns the conditions for different jet pseudorapidities ($|\eta_j|$) ranges are given.

MET corrections

MET definition given in Eq. 3.1 does not take into account the results of the jet energy correction process. The necessary correction which is applied is called Type-I correction and consists of the propagation of the jet energy corrections onto missing transverse energy by replacing the vector sum of the transverse momenta of the uncorrected jets with the same for corrected (JEC applied) jets. The Type-I correction is the most widely used MET correction in CMS.

The MET can be decomposed with respect to the origin of the energy as in the following expression:

$$\vec{E}_T^{\text{uncorr.}} = - \sum_i \vec{p}_{T,i} = - \sum_{i \in \text{jets}} \vec{p}_{T,i} - \sum_{i \notin \text{jets}} \vec{p}_{T,i} = - \sum_{\text{jet}} \vec{p}_{T,\text{jet}}^{\text{uncorr.}} - \sum_{i \notin \text{jets}} \vec{p}_{T,i}, \quad (4.1)$$

where we have split the sum of the PF particles transverse momentum into clustered objects (jets) and unclustered objects parts. The jets can be further classified into classes with a threshold at the corrected transverse momentum:

$$\vec{E}_T^{\text{uncorr.}} = - \sum_{\substack{\text{jet} \\ \vec{p}_{T,\text{jet}}^{\text{L123}} > 10\text{GeV}}} \vec{p}_{T,\text{jet}}^{\text{uncorr.}} - \sum_{\substack{\text{jet} \\ \vec{p}_{T,\text{jet}}^{\text{L123}} < 10\text{GeV}}} \vec{p}_{T,\text{jet}}^{\text{uncorr.}} - \sum_{i \notin \text{jets}} \vec{p}_{T,i}, \quad (4.2)$$

where L123 indicates L1, L2 and L3 levels of JEC (see Section 3.2).

Type-I correction can be written as $(\vec{p}_{T,\text{jet}}^{\text{uncorr.}} - \vec{p}_{T,\text{jet}}^{\text{L123}})$, so we can modify Eq. 4.2 to the form:

$$\vec{E}_T^{\text{uncorr.}} = - \sum_{\substack{\text{jet} \\ \vec{p}_{T,\text{jet}}^{\text{L123}} > 10\text{GeV}}} \vec{p}_{T,\text{jet}}^{\text{L123}} - \sum_{\substack{\text{jet} \\ \vec{p}_{T,\text{jet}}^{\text{L123}} > 10\text{GeV}}} (\vec{p}_{T,\text{jet}}^{\text{uncorr.}} - \vec{p}_{T,\text{jet}}^{\text{L123}}) - \sum_{\substack{\text{jet} \\ \vec{p}_{T,\text{jet}}^{\text{L123}} < 10\text{GeV}}} \vec{p}_{T,\text{jet}}^{\text{uncorr.}} - \sum_{i \notin \text{jets}} \vec{p}_{T,i}.$$

The Type-I corrections are applied for *MET collection* available in collision and MC data samples, but without MET covariance matrix filled. As it is required for SVfit algorithm (see Section 3.4) we have recalculated MET using updated JEC (p. 46).

Electron identification

Electrons are reconstructed from ECAL clusters and GSF tracks (Chapter 3). The identification of electrons is based on the information about reconstructed electron electromagnetic shower shape and track parameters, e.g. cluster pseudorapidity, electron transverse momentum, cluster width (in η and ϕ directions), active tracker layers number, number of missing hits, track-to-cluster displacement and so on. To avoid creating a multitude of separated identification categories in these strongly (p_T, η) -dependent calculations, we used multivariate electron identification method (MVA ID) where single discriminator variable is computed in a way providing the best separation between the signal (prompt electrons) and backgrounds (fake electrons, mostly due to misidentified jets). The different WPs are defined for different signal efficiency as *cuts on discriminator value*. Working points are trained on DY + Jets MC samples (with prompt electrons as signal and non-prompt and fake/unmatched electrons as background) corresponding to data collected in 2016 and tuned for the $p_T > 10$ GeV electrons.

Generator matching

Many corrections applied to MC samples depend on the origin of the reconstructed hadronic tau, which is taken as the type of generated particle that matches the tau particle most closely in the $(\eta - \phi)$ -plane (dR). In the process of *generator matching* six categories are introduced, and they are presented in Table 4.2.

Cat	Type	Nearest gen level object
1	Prompt electron	<i>Prompt electron</i> , $p_T > 8$ GeV
2	Prompt muon	<i>Prompt muon</i> , $p_T > 8$ GeV
3	$\tau_h \rightarrow e$	<i>Direct prompt electron</i> , $p_T > 8$ GeV
4	$\tau_h \rightarrow \mu$	<i>Direct prompt muon</i> , $p_T > 8$ GeV
5	$\tau_h \rightarrow \tau_h$	Generator level τ_h jet, $p_T > 15$ GeV (summed four-momenta of the visible (prompt) generator τ_h decay products)
6	Jet/pile-up fake	Anything that does not fall in any of the above categories i.e. no match within $dR < 0.2$

Table 4.2: Hadronic tau generator matching categories

The $p_T > 8$ GeV cut in Table 4.2 is introduced in order to eliminate electrons and muons from decays of hadronic resonances and photons converting into a lepton pair (significant for $p_T^{e,\mu}$ below 6 GeV) and is set to be at least 1 GeV below the analysis thresholds on the reconstructed lepton transverse momenta. The $p_{T,\tau_h} > 15$ cut is inspired by the transverse momentum requirement for seeding a jet in hadronic tau reconstruction ($p_T^{\text{seed}} > 14$ GeV).

The generator matching is crucial from the point of view of MC corrections/rescaling, because they depend on the generator matching category. Hence it strongly influences both shapes and yields in final discriminating variable(s) distributions. Above is the reason for the decision to synchronize the *generator matching recipe* between the CMS groups involved in the $H \rightarrow \tau\tau$ analysis.

Lepton isolation

The relative (with respect to the lepton transverse momentum) lepton isolation is given by the formula:

$$R_{iso}^l = \left(\sum p_T^\pm + \max \left[0, \sum p_T^0 + \sum p_T^\gamma - \frac{1}{2} \sum p_T^{\text{pu}} \right] \right) / p_T^l, \quad (4.3)$$

where p_T^\pm is transverse momentum of charged hadron originating from the primary vertex and $p_T^{0,\gamma}$ is the same for neutral particles (neutral hadron and photon). The p_T^{pu} is a neutral particles pile-up correction using: 1) a jet area method [90] for electrons; 2) the half of the sum of pile-up charged particles (i.e. transverse momenta for particles not originating from the primary vertex) for muons. Only half of the sum is taken as this corresponds to the ratio of neutral to charged hadrons production in inelastic proton-proton collisions. The "max"-term is put to ensure that only the positive pile-up corrected contributions from neutral particles are taken into account. Negative corrections can occur due to statistical character of pile-up contamination. All summations in Eq. 4.3 are over PF-particles in the cone around lepton with size of $\Delta R = 0.3$ for electrons and $\Delta R = 0.4$ for muons. The numerator of Eq. 4.3 is called the absolute lepton isolation.

Tau isolation

The hadronic tau isolation is derived for Run II data with MVA method using e.g. energy deposits in calorimeters around tau candidate and tau lifetime information. The resulting discriminator (tau MVA isolation ID) is available for every tau candidate in the data tau collection (see Section 3.3). It comes in two forms: as decimal number in range from zero to one (raw tau MVA isolation ID) or as one of six WPs². In the latter case, the discriminator either takes zero value (does not pass WP, i.e. tau is not isolated) or one (tau is isolated on given WP level).

Muon identification

The muon identification (ID) helps to reject fake and non-prompt muons, e.g. from decays in-flight. *Three WPs* are available in data (muon collection):

- Loose muon ID: it requires PF-muon which is also reconstructed as either global-muon or tracker-muon, without any further additional track-quality and muon-quality requirements. It is highly efficient for prompt muons and muons from quark decays.
- Medium muon ID: it is attributed to the muon particle if it passes loose muon ID together with additional track-quality and muon-quality requirements related to tracker to muon system matching (segment compatibility, global χ^2), number of hits in the muon system, tracker variables (number of missing hits, tracker track χ^2) and calorimeter energy deposits compatibility. Medium muon ID is optimized to reach 99.5% efficiency on prompt muons with transverse momentum above 20 GeV and absolute pseudorapidity below 1.2 (in barrel part of the muon system).
- Tight muon ID: it requires global muon with additional muon-quality requirements. It has the best purity but is 2-3% less efficient than medium muon ID.

Although available in the data, the medium muon ID had to be recalculated for collision data taking periods 2016B to 2016F in order to match the performance of the ID obtained in data taking periods 2016G and 2016H³. The criteria for so called *medium2016 muon ID* are given in Appendix B.

Tau energy scale corrections

Tau energy scale (TES) corrections are applied to the hadronic taus with transverse momentum between 20 and 400 GeV⁴, with decay mode 0, 1 and 10 (see definition in Section 3.4) and with respect to the generator matching category (see Table 4.2). The energy of the tau can be corrected in two ways: 1) by re-scaling the (p_x, p_y, p_z) -vector by some factor and recalculating tau energy or 2) by re-scaling the whole four-vector of the tau.

²VVL / VL / L / M / T / VT WP tau MVA isolation ID, where V - very, L - loose, M - medium and T - tight (see Fig. 3.7)

³The tracker issues were the reason for different performance. A decrease of signal over noise ratio associated to loss of tracking hits has been observed in late 2015 and part of 2016. The problem was initially believed to be due to heavily ionizing particles (HIPs), but finally traced to saturation effects in the pre-amplifier of the APV chip of tracker readout.

⁴The hadronic taus with $p_{T,\tau_h} < 20$ are not used and the taus with $p_{T,\tau_h} > 400$ are very rare and no correction is delivered.

Generator matching category	5	5	5	2 or 4	2 or 4	1 or 3
Tau Decay Mode	0	1	10	0	1	1
TES factor value	-1.8%	+1.0%	+0.4%	-0.2%	+1.5%	+9.5%

Table 4.3: Tau Energy Scale factors. Taus with combinations of generator matching category and decay mode that are not present in this table are not scaled. The uncertainty on all TES factor values is 0.6%.

For 1-prong tau decays the mass of the single hadron in jet is assumed to be charged pion mass, therefore the energy of taus with decay mode equal to zero is corrected using the first method to keep the mass unchanged. For other decay modes the second method is used. The re-scaling factors are presented in Table 4.3. TES influences the yields; taus can be pushed out or in of the analysis p_{T,τ_h} threshold, depending on the decay mode.

Baseline selection

Before selecting the best pair (Higgs boson candidate) in the event, pairs are verified for the constituent lepton quality (even if there is only one pair existing). Hadronic taus have to fulfill the following requirements:

- transverse momentum above 30 GeV ($p_{T,\tau_h} > 30$ GeV) and absolute pseudo-rapidity below 2.3 ($|\eta_{\tau_h}| < 2.3$). The first requirement is meant to increase the significance in the analysis and the second one is justified by tau object reconstruction constraints (related to available detector acceptance) and in agreement with value recommended by CMS Tau Particle Object group;
- longitudinal transverse impact parameter below 0.2 cm ($d_z < 0.2$ cm). It measures the distance of the track projection (distance of closest approach) to the primary vertex (PV) along the beam line direction. For hadronic tau this parameter is taken as transverse parameter of the leading charged hadron PF candidate used to HPS reconstruction;
- charge has to be equal to plus or minus one ($|q| = 1$);
- must be *reconstructed as 1- or 3-prong tau decay*.

Muon has to fulfill the following requirements:

- transverse momentum above 20 GeV ($p_{T,\mu} > 20$ GeV) and absolute pseudo-rapidity below 2.1 ($|\eta_\mu| < 2.1$). Both thresholds are driven by HLT trigger requirements.
- longitudinal transverse impact parameter below 0.2 cm ($d_z < 0.2$ cm) and transverse impact parameter below 0.045 cm ($d_{xy} < 0.045$ cm). Transverse impact parameter measures the distance of the track projection to PV in the $(x - y)$ -plane (perpendicular to the beam axis);
- has to pass medium (2016) muon ID.

The Higgs boson candidates which pass the lepton requirements are checked for compatibility with HLT trigger. If the transverse momentum of the muon in a pair is above 23 GeV, it has to match HLT filter trigger object within $\Delta R < 0.5$ for the single isolated muon trigger (with muon p_T^{HLT} threshold of 22 GeV). For Higgs boson candidates with muon in $20 < p_{T,\mu} \leq 23$ range, both muon and tau have to match appropriate HLT filter trigger object. The lowering of muon $p_{T,\mu}$ threshold from 23

to 20 GeV is introduced to the analysis in order to increase the signal acceptance (by about 25%) and sensitivity (by about 10%). Triggers and filters to match are given in Appendix B in Table B.1.

Best pair (Higgs boson candidate) selection

As already mentioned there can be more than one $\tau_\mu\tau_h$ pair in the event, which is undesirable. It was observed that the selection based on the lepton isolation had the highest efficiency for selecting the correct pair in signal events. If multiple pairs exist after HLT matching the following procedure is executed:

- For every pair, the relative muon isolation R_{iso}^μ is calculated. Only the pairs with smallest R_{iso}^μ are kept.
- If after the first step there is still more than one pair (for cases with possibly different muon candidates, but of genuinely the same isolation value), only the ones with the highest muon transverse momentum are kept.
- If after the second step there is still more than one pair (likely because two pairs share the same muon object), only the ones with the highest score of *MVA-based tau ID discriminant* are kept.
- If after third step there is still more than one pair with the same leptons isolation and muon p_T^μ , the one with the highest tau p_{T,τ_h} is kept.

The selected pair is put to one additional test. First, it is checked if muon is loosely isolated ($R_{iso}^\mu < 0.3$). Then to avoid events with electron or muon misidentified as hadronic tau, it is required for tau to satisfy anti-lepton discriminators. For the discriminator against muon the tight WP is chosen in order to efficiently reduce the large $Z \rightarrow \mu\mu$ background. For the discriminator against electron the very loose WP is selected because of a small yield for backgrounds with electrons in the final state and to keep a high efficiency for the signal. If the pair fails this test, the event is removed.

Event vetos

In order to keep different final states of the $H \rightarrow \tau\tau$ analysis exclusive and suppress the DY+jets background all events should contain exactly two leptons of a desired flavor. In particular if there are two good, isolated, opposite charge sign (OS) muons in the event it is probably a Drell-Yan process $Z/\gamma^* \rightarrow \mu\mu$ with an additional jet faking hadronic tau. Above scenarios are covered with three boolean veto variables presented in Table 4.4. If any of those variables is true, the event is rejected.

In Table 4.4 the conversion veto discriminator assures that the electron does not come from photon conversion ($\gamma \rightarrow e^+e^-$). It is based on the information from the tracker, like the number of missing hits, distance between (hypothetical) conversion tracks and angles between conversion tracks at conversion vertex.

Corrections of missing transverse energy

The missing transverse energy is assigned to the pair/event in one of the three forms, depending on the data sample: 1) without any modifications (for real data); 2) with MET updated for the TES (for MC samples without recoil corrections); 3) with MET updated for the TES and then recoil corrected (for MC samples with recoil corrections, see Appendix A).

Veto name	Lepton definition	Truth condition
dimuon	$p_T > 15$, $ \eta < 2.4$, is global, tracker and PF muon, $d_z < 0.2$, $d_{xy} < 0.045$, $R_{iso}^l < 0.3$	two opposite sign muons with $dR > 0.15$
extra muon	$p_T > 10$, $ \eta < 2.4$, is medium muon, $d_z < 0.2$, $d_{xy} < 0.045$, $R_{iso}^l < 0.3$	more than one muon
extra electron	$p_T > 10$, $ \eta < 2.5$, $d_z < 0.2$, $d_{xy} < 0.045$, $R_{iso}^l < 0.3$, passes conversion veto, track has maximum one missing inner track hit, passes electron MVA ID 90% WP	at least one electron

Table 4.4: Hadronic tau generator matching categories

The change in tau transverse momentum caused by TES needs to be reflected in MET value. The following formula for MET TES-corrected four-vector elements is used:

$$\begin{aligned} p_x^{\text{corr}} &= p_x - \Delta p_x^\tau = E_T \cdot \cos \phi - (p_x^{\tau, \text{corr}} - p_x^{\tau, \text{uncorr}}); \\ p_y^{\text{corr}} &= p_y - \Delta p_y^\tau = E_T \cdot \sin \phi - (p_y^{\tau, \text{corr}} - p_y^{\tau, \text{uncorr}}), \end{aligned} \quad (4.4)$$

where ϕ and $p_{x,y}$ are transverse momentum azimuthal angle and (x, y) -coordinates of momentum vector, respectively.

It may happen, that the measured MET value does not strictly correspond to the genuine MET in the event and carries out the admixture of instrumentally induced MET. To the unwanted induced part of the MET contribute many experimental and physical effects like JEC, fluctuating jet composition, detector noise, pile-up, restricted detector resolution and acceptance, etc. MET recoil correction is introduced as the negative value of the induced MET. In other words recoil correction is a way of fixing the mismodeling of missing transverse momentum in the simulated samples. Strictly speaking, the correction

$$\vec{U} = \vec{\cancel{E}}_T - \vec{p}_{t,\nu} \quad (4.5)$$

is the difference of reconstructed MET ($\vec{\cancel{E}}_T$) and total transverse momentum of neutrinos ($\vec{p}_{t,\nu}$). The \vec{U} value is figured out by analyzing $Z \rightarrow \mu\mu$ events⁵, where no genuine MET is expected and vector boson transverse momenta are well-understood and well-measured. Therefore, the \vec{U} is determined as an offset with respect to the zero MET value. The recoil is computed for data and MC separately and the MC result is corrected w.r.t. the data. The correction vector can be decomposed onto axis parallel ($U_{||}$) and perpendicular (U_{\perp}) to the boson transverse momentum direction. The $U_{||}$ is typically negative and decreases linearly with boson p_T due to increase in jet energy resolution, whereas U_{\perp} is dominated by noise and pile-up. Recoil corrections are prescribed for the Drell-Yan, W+Jets and Higgs production simulated samples (where boson transverse momentum is well defined) and applied depending on the generated *boson p_T value (full and visible) and jet multiplicity*. The boson transverse momentum is known and reconstructed from simulated (generator) particles.

After MET corrections are done, the mass of the Higgs boson candidate is calculated with SVfit algorithm (Section 3.4) and all samples are saved for processing outside CMSSW.

⁵The list of cuts required for dimuon events extraction is presented in the analysis note [82].

4.3 Analysis categories

So far we have assured that all reconstructed leptons in the pair lie within the detector acceptance, originate from the primary vertex and are loosely isolated. Now we split the analyzed events into categories, so it is possible to fine tune the final cuts and further increase the significance.

Three main categories are defined by choosing events with two, one and zero jets, which reflects the anticipated Higgs production process. Zero-jet (0jet) category targets events with gluon fusion Higgs boson production. Two-jet (vbf) category is intended to be enriched in vector boson fusion Higgs production and delivers the highest signal-to-background ratio, but also the lowest signal yield. The reduction of the SM background (e.g. $t\bar{t}$ events) in this category is achieved by requiring the leading two jets system to have a large invariant mass ($m_{jj} > 300$ GeV). One-jet (boosted) category contains events which did not fall into any of the previous ones. It contains mainly events from gluon fusion Higgs production with additional jet(s) ($\approx 80\%$ of signal events in this category), vector boson fusion Higgs production with one jet unreconstructed or with two jets with low dijet system mass ($\approx 12\%$ of signal events) and associated Higgs boson production where W/Z boson decays hadronically ($\approx 8\%$ signal events).

All categories require the tau lepton to be isolated by passing *tight WP tau MVA isolation ID*, have leptons of opposite sign (charge conservation) and with a low transverse mass (M_T^μ) of the muon lepton candidate defined as:

$$M_T^\mu := \sqrt{2p_{T,\mu} \cancel{E}_T (1 - \cos \Delta\phi(\mu, \cancel{E}_T))}, \quad (4.6)$$

where $\Delta\phi(\mu, \cancel{E}_T)$ is the azimuthal angle between muon and missing transverse energy. For heavy W boson leptonic decays ($W \rightarrow \mu\nu_\mu$) the muon and neutrino are more likely to be emitted back-to-back

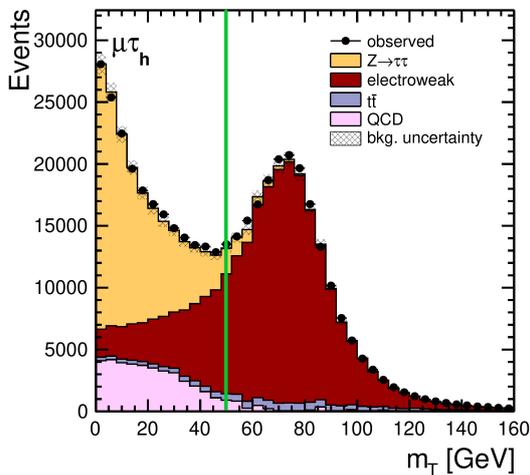


Figure 4.1: Distribution for M_T^μ in $H \rightarrow \tau_\mu \tau_h$ events for $L=19.4 \text{ fb}^{-1}$ of $\sqrt{s} = 8$ TeV CMS data. Green line indicates the 50 GeV cut [91].

and the Jacobi peak for $M_T^\mu \simeq m_W$, because transverse mass simulates the "invariant mass" of W boson made of available transverse momenta and with muon mass neglected. Therefore, for such events the transverse mass will be large. On the other hand, in DY + jets and $H \rightarrow \tau_\mu \tau_h$ events, the muon and neutrino are more likely to travel in the same direction due to the lower tau lepton mass compared to W boson. Furthermore, MET is only instrumental for QCD events. For all such cases the transverse mass tends to be small. The additional effects come from compensating neutrino momentum vectors in symmetrical decay scenarios, which

further decreases the reconstructed missing transverse energy and hence the transverse mass value. The 50 GeV cut is motivated by the placement of W + jets peak in the transverse mass distribution (Fig. 4.1), and allows to suppress significant backgrounds of W+jets and $t\bar{t}$ events.

The Higgs transverse momentum is defined as:

$$\vec{p}_T(H) := \vec{p}_T^{\tau_h} + \vec{p}_T^{\tau_\mu} + \vec{E}_T. \quad (4.7)$$

The distribution of this variable is shifted towards higher values for various Higgs boson production scenarios (VBF, ggH with jets, associated vector boson production), hence it presents discriminating power versus mainly QCD background. Finally, the jet pair system invariant mass m_{jj} is the mass of the sum of the four-momenta of the two leading (highest p_T) jets in the event. For the vbf category the large m_{jj} effectively reduces different backgrounds, e.g. $t\bar{t}$. Categorization cuts are summarized in Table 4.5.

Variable	0jet	boosted	vbf
No. of jets	0	1 or >2 and ($m_{jj} \leq 300$ or $p_{T,\tau_h} \leq 40$. or $p_T^H \leq 50$.)	≥ 2
Pair charge	-1	-1	-1
Muon transverse mass	≤ 50	≤ 50	≤ 50
Tau isolation	> 0.5	> 0.5	> 0.5
Tau transverse momentum	–	–	> 40
Muon isolation	≤ 0.15	≤ 0.15	≤ 0.15
SVFit ($m_{\tau\tau}$) mass	–	< 300	< 400
Higgs transverse momentum	–	–	> 50
Jet pair system mass	–	–	> 300

Table 4.5: The cuts defining the categories in $H \rightarrow \tau_\mu \tau_h$ decay channel.

4.4 Simulated samples corrections

Pile-up correction

The number of primary vertices in the MC is adjusted at the event-by-event basis using the *pileup distributions* calculated for data and MC separately. The distributions are generated using "true pile-up" method, i.e. contain the histograms of the mean number of interactions per crossing. For data samples this method uses total inelastic minimum bias cross section (of 69.2 mb) and average instantaneous luminosity to obtain proper value for a given luminosity section in the LHC run.⁶ For MC pile-up is generated (with PYTHIA) using minimum bias collision events.

Tau isolation correction

The efficiency on the hadronic tau identification varies for different isolation WPs. The necessary scale factor (SF) is obtained from $Z \rightarrow \tau_\mu \tau_h$ events by performing maximum likelihood fit for visible mass distribution for two categories where this SF is anti-correlated. The number of events in the fit are adjusted using $Z \rightarrow \mu\mu$ control region. For the tight WP tau MVA isolation ID used in this analysis a constant SF of 0.95 with uncertainty of 5% is applied for events where hadronic tau is a generator matched to hadronic tau (category no. 5).

⁶Another method is "observed pile-up", which returns the counted number of events seen in a given beam crossing (using the number of vertices in data and generator variables in MC).

Electron/muon fake rate correction

Similarly to the tau isolation correction, a muon to tau fake rate and electron to tau fake rate scale factors can be measured. They are applied depending on the anti-lepton discriminator ID, tau generator matching category and pseudorapidity range. The muon/electron to tau fake rate SFs are presented in Table 4.6.

Tau discriminator	η range	Generator matching category	Scale factor
<i>loose anti-electron</i>	$\eta < 1.460$	1 or 3	1.213 ± 0.07
loose anti-electron	$\eta > 1.558$	1 or 3	1.375 ± 0.05
<i>tight anti-muon</i>	$\eta < 0.4$	2 or 4	1.263 ± 0.07
tight anti-muon	$0.4 \leq \eta < 0.8$	2 or 4	1.364 ± 0.28
tight anti-muon	$0.8 \leq \eta < 1.2$	2 or 4	0.854 ± 0.04
tight anti-muon	$1.2 \leq \eta < 1.7$	2 or 4	1.712 ± 0.5
tight anti-muon	$1.7 \leq \eta < 2.3$	2 or 4	2.324 ± 0.5

Table 4.6: Muon/electron to tau fake rate scale factors.

Drell-Yan $e/\mu \rightarrow \tau_h$ fake rate correction

For the Drell-Yan events the additional, decay mode-dependent⁷ $e/\mu \rightarrow \tau_h$ fake rate SFs are calculated. It is done in a three step process. In the first step, the MC events are divided into two groups: 1) the events where generated lepton (electron or muon) is reconstructed as hadronic tau (lepton faking tau) and 2) all the rest, i.e. where reconstructed hadronic tau is not faked from generator lepton. In the second step, the total yield of the latter group is subtracted from the available collision data yield. In the last step, the obtained collision data visible mass distribution is Gaussian fitted to the same distribution for the events from the first group above of MC events. This way the SFs can be found separately for electrons and muons faking taus and also for different tau decay modes. Table 4.7 shows the appropriate event SFs.

Correction type	Decay Mode	Generator matching category	Scale factor
$e \rightarrow \tau_h$ fake rate	1-Prong	1 or 3	$0.98 \pm 12\%$
$e \rightarrow \tau_h$ fake rate	1-Prong + π^0 s	1 or 3	$1.2 \pm 12\%$
$\mu \rightarrow \tau_h$ fake rate	1-Prong	2 or 4	$0.74 \pm 25\%$
$\mu \rightarrow \tau_h$ fake rate	1-Prong + π^0 s	2 or 4	$1.0 \pm 25\%$

Table 4.7: DY $e/\mu \rightarrow \tau_h$ fake rate corrections. The SFs are applied for all Drell-Yan events with lepton faking hadronic tau.

Muon tracking and trigger efficiency correction

Tag-and-probe is used to find the muon tracking efficiency and trigger efficiencies for collision and MC data. It allows to identify "probe" object as a muon in two muons decay thanks to the resonance mass constraint. For the muon efficiency measurements this technique relies on the dimuon resonance ($Z \rightarrow \mu\mu$, $J/\psi \rightarrow \mu\mu$) events, for which stringent quality requirements are imposed on the one muon (*tag*). For the second muon (*probe*) the two categories (set of cuts) are created, which differ by a single variable in relation to which the efficiency is computed. The efficiency is obtained as a number of events passing the more restrictive category divided by the same for a less restrictive selection. If both muons in the event pass both criteria (for

⁷As we will see in Section 4.7 one of the discrimination variables in 0jet category is decay mode.

tag and probe), they are examined independently. In the first case, it is checked if the standalone muon is a global muon; and in the second case, it is checked if the probe is matched to the HLT triggers from baseline selection in a cone $\Delta R < 0.5$. The μ/τ_h cross triggers efficiency is also found for $\mu\tau_h$ events, where muon was used as a tag. All efficiencies are calculated in probe (p_T, η) bins, and for cross triggers additionally for different tau decay modes. The ratio of efficiencies for data and MC are used as *weights in the analysis* (corrections).

Lepton identification correction

Lepton ID scale factors correct for lepton identification efficiencies and are measured with tag-and-probe method using $Z \rightarrow ee$ (for electron ID SF) and $Z \rightarrow \mu\mu$ (for muon ID SF) events. The event extraction is based on the set of cuts presented in [82]. The weight is defined as the efficiency ratio $\epsilon_{data}/\epsilon_{MC}$, where ϵ itself is a ratio between the number of probes passing the ID and failing it, and it is extracted from the invariant mass distribution for the window around Z boson mass peak. Please note that the event cuts are the same for the real data and for MC. The generator information is not used here. The event weight is generated as a function of the probe lepton transverse momentum and pseudorapidity.

LO MADGRAPH DY SF

The LO MADGRAPH Drell-Yan samples reweighing corrects the shape of the distributions with respect to the generator level Z boson transverse momentum and mass modeling inaccuracies due to usage of LO calculations⁸. The weight is produced using $Z \rightarrow \mu\mu$ dataset, in Z -(p_T, m) bins, and is normalized so it does not modify the yield. The Z boson is reconstructed from generator particles. The requirements on the events for $Z \rightarrow \mu\mu$ extraction are given in [82]. The *weights are presented in Fig. 4.2*.

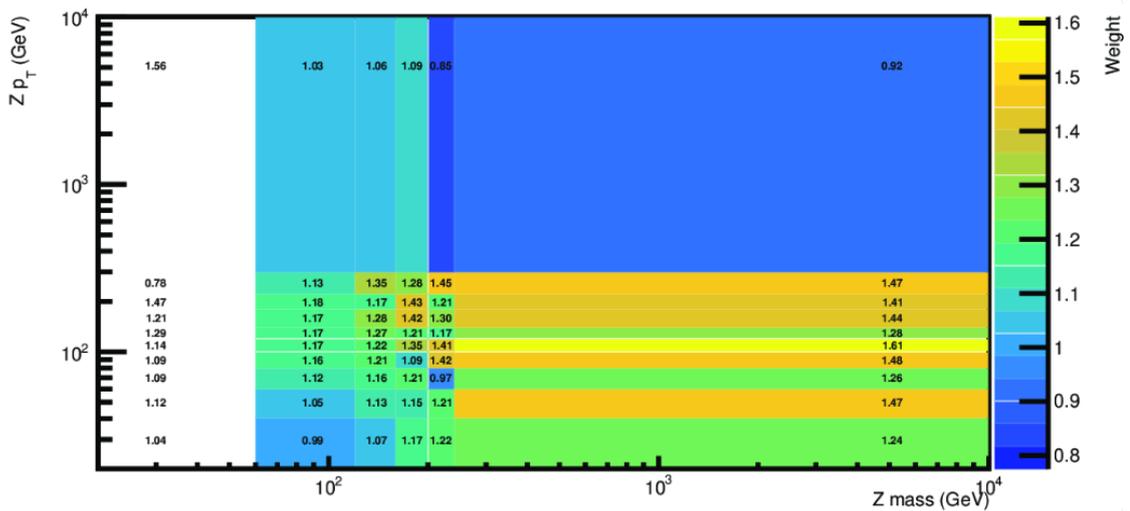


Figure 4.2: LO MADGRAPH Drell-Yan weights.

⁸DY is simulated at LO, as it allows to simulate samples with different jet multiplicity (up to 4 jets) for coverage of the full phase space.

4.5 Simulated samples normalization

Drell-Yan and W+jets

Simulated samples (MC) have to be normalized in order to represent the same yield as data collected in proton-proton collisions. The total weight for the event is the product of partial weights from the previous section and normalization.

The DY+jets MC samples are given in two versions: inclusive samples and samples with a number of additional final state partons (jets) forced at the generator level (between 1 and 4). This way it is easier to cover a large area of phase space taking into account constraints on the dataset volume. The different Z+jets samples have to be stitched into one with every event weighed in a proper manner. The weight is calculated based on cross section and number of events for a given multiplicity of final state partons. The procedure is the following:

- The total number of processed events T_i^{DY} for a given dataset ($i \in$ inclusive, DY+jet, DY+2jets, ..., high mass inclusive DY+Jets sample) is found.
- The normalized total number of processed events NT_i^{DY} is defined as a total number of processed events divided by appropriate cross section (given in Appendix A), i.e. $NT_i^{DY} = T_i^{DY} / \sigma$.
- The raw weight per event is calculated as

$$\frac{SF_{(NN)LO}^{DY}}{NT_{inclusive}^{DY} + NT_{partons}^{DY} + NT_{HM}^{DY}}. \quad (4.8)$$

Here $SF_{(NN)LO}^{DY}$ is NNLO-to-LO correction factor equal to 5765.4/4895 (cross section NNLO-to-LO ratio for inclusive Drell-Yan samples). Subscript *partons* is an integer corresponding to the number of jets in the dataset. $NT_{partons}^{DY}$ is zero for events with number of partons less than one or more than four. The NT_{HM}^{DY} is equal to NT_i^{DY} value for high mass inclusive DY+Jets sample for events with mass above 150 GeV (and zero otherwise). By "mass" we mean the invariant mass of the two *leptons* system in the event.

- Finally the weight is obtained from raw weight by multiplying it by the luminosity of analyzed collision data.

The W+jets MC samples stitching is done by the same procedure with two exceptions: 1) $SF_{(NN)LO}^{DY}$ is replaced with $SF_{(NN)LO}^W$ equal to 61526.7/50380 (cross section NNLO-to-LO ratio for inclusive W + jets samples) and 2) there is no dependency on event mass, so NT_{HM}^W is always zero.

Other MC

The normalization of the remaining MC samples is given by the formula:

$$\frac{\text{luminosity} * \text{cross section}}{\sum \text{generator weight}} \quad (4.9)$$

where summation in the denominator is performed over all processed events. The generator weight is a number assigned to every event in simulated samples and is almost always equal to unity. The only exception are diboson samples, which

are generated with MADGRAPH program at NLO cross sections (AMCATNLO) and where generator weights can take both positive or negative values. Typically the fraction of events with negative weight is 10-20% and since they are canceling 10-20% of the positive weighted events the statistics are effectively reduced (to 60-80% of original size) for those datasets. The reason for negative weights stem from the way the NLO event generator is matched to the (N)LL accuracy parton-shower and hadronization algorithm (see [92, 93, 94]).

4.6 Background estimation

Drell-Yan

DY + Jets background is modeled wholly by the MC and only rescaled and normalized on per with description given in Sections 4.4 and 4.5. Every event from *DY + jets* samples falls into one out of three categories:

- ZTT category contains events where hadronic tau in Higgs boson candidate is matched to hadronic tau particle at the generator level (category no. 5 in Table 4.2). Drell-Yan production with a tau pair in the final state is the main irreducible background in $H \rightarrow \tau_\mu \tau_h$ decay channel.
- ZL category contains events where hadronic tau in Higgs boson candidate is matched to muon or electron at the generator level (cat. no. 1-4). Here, the background comes e.g. from $Z \rightarrow \mu\mu$ events where one muon is faking a hadronic tau.
- ZJ category contains events where hadronic tau in Higgs boson candidate is matched to the jet at the generator level (cat. no. 6). Here, the background comes e.g. from $Z \rightarrow \mu\mu$ events where one muon is lost and additional final state jet is faking a hadronic tau.

W+jets

The W+jets background plays an important role in $H \rightarrow \tau_\mu \tau_h$ channel in case where W boson decays leptonically and jet is misidentified as a hadronic tau. The shape of this background is taken from the simulation, but to cope against the small yield in the signal region causing significant statistical fluctuations relaxed selection is used. In relaxed selection the tight WP tau MVA isolation ID is replaced with medium WP and the muon isolation is raised from 0.15 to 0.30. It is confirmed that this does not bias the shape of the distribution with respect to the nominal cuts.

Let us denote nominal (i.e. signal region), low transverse mass ($M_T^\mu < 50$ GeV) W+jets yield by W_n^\downarrow , nominal high transverse mass ($M_T^\mu > 80$ GeV) yield by W_n^\uparrow , and relaxed yields accordingly with lower index r . The signal region yield can be written as

$$W_n^\downarrow = W_r^\downarrow(MC) \cdot SF_{r \rightarrow n}, \quad (4.10)$$

where scale factor $SF_{r \rightarrow n} \equiv W_n^\downarrow / W_r^\downarrow$ describes the transition from relaxed to nominal conditions. In the analysis it is actually derived using data-driven method where collision data sample of high- M_T^μ (W+jets enriched) events is used:

$$SF_{r \rightarrow n} = \frac{W_n^\uparrow(Data)}{W_r^\uparrow(MC)}. \quad (4.11)$$

The $W_n^\uparrow(Data)$ is defined as data yield with QCD and all MC backgrounds (except W+jets) samples subtracted. Because the QCD contribution is also estimated from data and dependent on the W+jets MC, the normalization of the W+jets background is dependent on itself. However, this is a second order effect of roughly 4% and is accounted as a systematic uncertainty. The whole W+jets estimation process is repeated for every category separately. The 5% (0-jet and boosted category) or 25% (vbf category) uncertainty related to the (statistical) uncertainty of data, W+jets, QCD and $t\bar{t}$ in high- M_T^μ sideband are applied. A higher uncertainty value for vbf stems from a higher statistical uncertainty and a lower purity of W+jets background. Additional 5% (boosted cat.) or 10% (0-jet and vbf cat.) uncertainties are related to the high- M_T^μ to low- M_T^μ extrapolation factor.

QCD

The contamination from QCD multi-jet events is an important reducible source of background in the $H \rightarrow \tau_\mu \tau_h$ channel. It is estimated from collision data using same sign (SS) events⁹ and given by the formula:

$$QCD_{OS} = SF_{QCD}^{SS \rightarrow OS} \times \left[\text{Data} - \text{DY}, t\bar{t}, VV, \text{EWK}, \text{Other} - SF_W^{\text{Data} \rightarrow \text{MC}} \times W \right]_{SS},$$

where

- OS (SS) subscript indicates the yield coming from OS (SS) events;
- $SF_{QCD}^{SS \rightarrow OS}$ is a scale factor correcting the difference between the number of OS and SS events observed in QCD enriched region (obtained by inverting muon isolation criteria).
- $SF_W^{\text{Data} \rightarrow \text{MC}} = W_n^\uparrow(Data) / W_n^\uparrow(MC)$ is a scale factor describing the ratio of Data to MC events for nominal high- M_T^μ yields.

Ditop

The $t\bar{t}$ background is estimated from MC sample and similarly to the DY background is split into sub-samples depending on the tau generator matching. The two categories are "TTT" if tau is matched to hadronic tau and "TTJ" otherwise.

Others

Other backgrounds, like diboson and single-top-quark event samples, are small and their contributions are estimated with MC simulation.

4.7 Signal variables

In the previous Run I $H \rightarrow \tau\tau$ analysis the Higgs boson signal was extracted from SVfit ditau mass distributions. In order to maximize signal to background (S/B) ratio a multiple (up to 7) different categories were considered for every tau pair decay into a final state. For the Run II another approach was developed, where SVfit mass was replaced with 2D distributions. The first variable in 2D distributions represents

⁹"Same sign" concerns the charge of particles in the pair, i.e. the event is SS if muon and tau have the same (negative or positive) charge. The signal events are of opposite sign (OS) type.

the tau pair system mass ($m_{\tau\tau}$) and the second is selected specifically for every category. This way it is possible to maintain good analysis sensitivity with only three categories – 0jet, boosted and vbf – defined using a limited number of variables and very similar for every final state. At the same time, the simplified collection of cuts did not affect the statistics – the background could be controlled with a number and range of bins keeping a high amount of events in the 2D distribution.

In the $H \rightarrow \tau_\mu \tau_h$ final state the following variables have been used:

- The high resolution of SVfit mass is not crucial in 0jet category since it has the lowest S/B ratio. Instead of delivering good Higgs mass reconstruction, the 0jet category comes useful in constraining the background contributions. The chosen lepton pair visible mass (m_{vis}) as a discriminating variable in the final fit helps to separate poorly-described $Z \rightarrow ll$ background. Another constraint on $Z \rightarrow \mu\mu$ background comes from the hadronic tau reconstruction, as muon to tau fakes are not present in the 3-prong decay mode. Decay mode is a second variable in the 2D distributions, with three bins representing 1-prong, 1-prong with neutral pion(s) and 3-prong decays.
- In the boosted category Higgs boson candidate invariant (SVFit) mass ($m_{\tau\tau}$) is the first discriminating variable. The second one is transverse momentum of Higgs boson candidate p_T^H (see Section 4.3) as it takes high values for Higgs gluon fusion events with Lorentz-boosted boson recoiling against jets. High- p_T^H events are very unlikely to many backgrounds, especially W + jets or QCD ones.
- Higgs boson candidate invariant mass ($m_{\tau\tau}$) and the jet pair invariant mass (m_{jj}) are chosen for variables in the vbf category. High- m_{jj} are suppressed in mainly QCD multi-jet events.

4.8 Analysis cross check

The form of 2D discriminating variables distributions is neither very clear for human reading nor convenient to process in statistical analysis tools. Therefore, before further processing, all 2D plots are projected into a single axis. More specifically all bins of visible mass (for 0jet category) or SVFit mass (for boosted and vbf categories) are placed side-by-side consecutively for every bin of the second variable (decay mode, Higgs boson candidate transverse momentum or the jet pair invariant mass). These so called "unrolled" distributions, created for all channels ($\tau_h \tau_h, \mu \tau_h, \dots$ etc) and replicated multiple times with rescaled/re-weighted variables according to the systematic uncertainties implementation, are an input into the global maximum likelihood fit performed for the official CMS publication. Moreover, they are used to create stacked data-to-background distributions (so called "control plots"). The official analysis control plots are given in Appendix C and the final results from official $H \rightarrow \tau\tau$ analysis are presented in the next section. For reference, the official CMS control plot for $H \rightarrow \tau_\mu \tau_h$ vbf category is shown in Fig. 4.3.

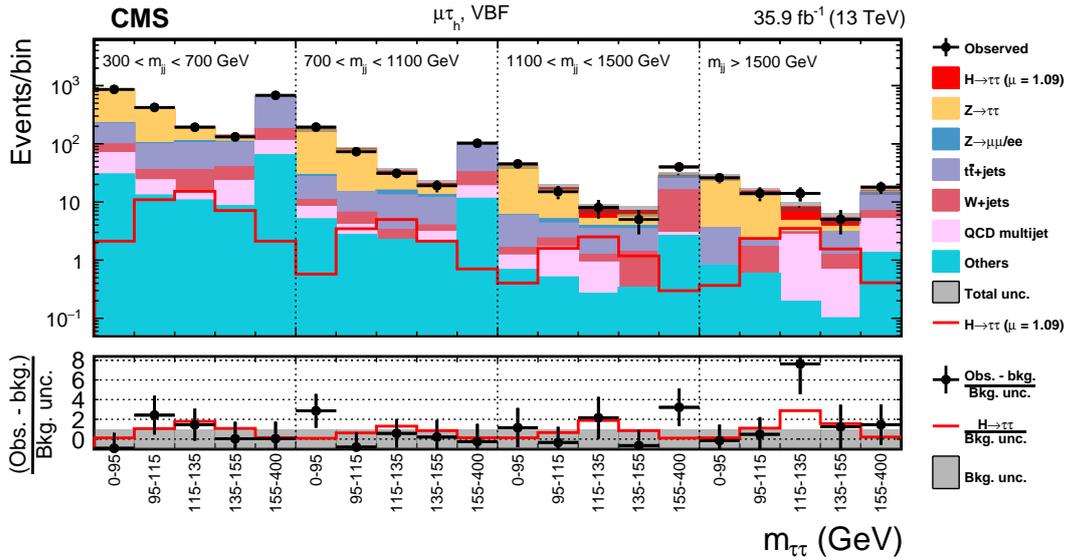


Figure 4.3: Observed and predicted 2D distributions in the vbf category of the $\tau_\mu\tau_h$ decay channel. The normalization of the predicted background distributions corresponds to the result of the global fit. The signal distribution is normalized to its best fit signal strength. The background histograms are stacked. The "Others" background contribution includes events from diboson and single top quark production, as well as Higgs boson decays to a pair of W bosons. The background uncertainty band accounts for all sources of background uncertainty, systematic as well as statistical, after the global fit. The signal is shown both as a stacked filled histogram and an open overlaid histogram [82].

The nominal discriminating variable unrolled distributions are also used to check the agreement ("synchronization") of different $H \rightarrow \tau\tau$ analysis versions between all groups involved. The selection of plots (for collision data, signal and background) from this synchronization analysis made for Official CMS, Cross-check CMS and Author implementations are presented in Figures 4.4 and 4.5. Official CMS distributions are the ones which were used to obtain the results in the official publication. Cross-check CMS distributions are delivered by an independent CMS group and were used internally by the CMS Collaboration before the publication. Author distributions are generated as a result of $H \rightarrow \tau_\mu\tau_h$ analysis presented in this chapter. The very complicated $H \rightarrow \tau_\mu\tau_h$ analysis made it practically impossible to obtain exactly the same distributions. The differences stem from e.g. the different CMSSW versions used, the different order of data processing steps, the different cuts and weights implementation (which can be further histogram- or value-based) or even different computing system architectures and programming languages which end up in different number of significant figures of (floating) numbers (and rounding results). There are two main conclusions coming out from the synchronization analysis. First, that checked distributions are in agreement in a sense, that almost all points on the histograms are within Official CMS distributions error bars. Second, that Author implementation presents better agreement with Official CMS than Cross-check CMS. The only exception is for 0jet category ZTT background, where total yield of the Author distribution is about 1.5% below the Official CMS and Cross-check CMS ones.

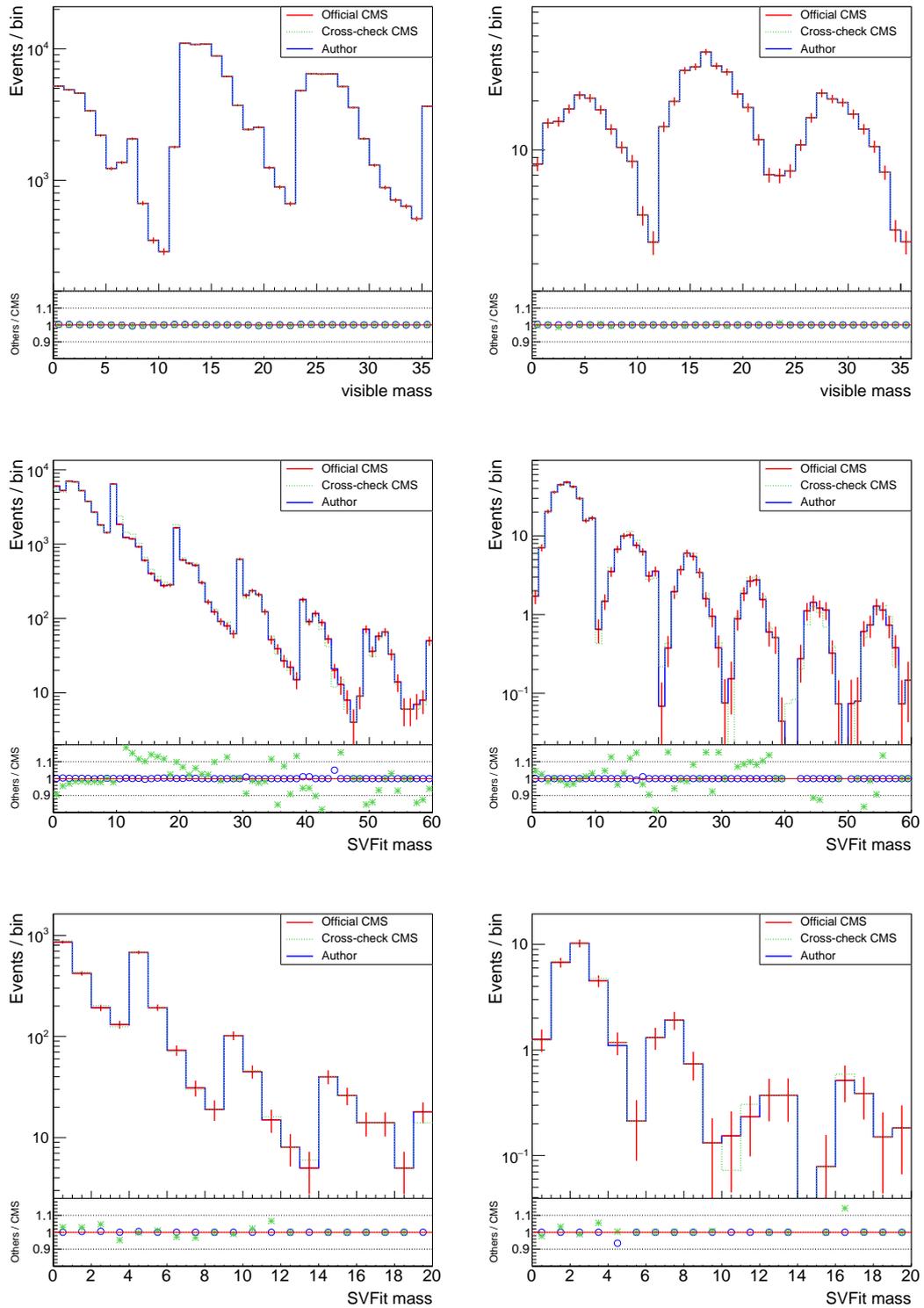


Figure 4.4: The events yield for collision data (left column) and signal gluon fusion Higgs production MC (right column) samples for Official CMS (red), Cross-check CMS (green) and Author (blue) analysis implementations. The 0jet (top row), boosted (middle row) and vbf (bottom row) categories are presented in different rows. The definitions of bins are taken from control plots (see Appendix C).

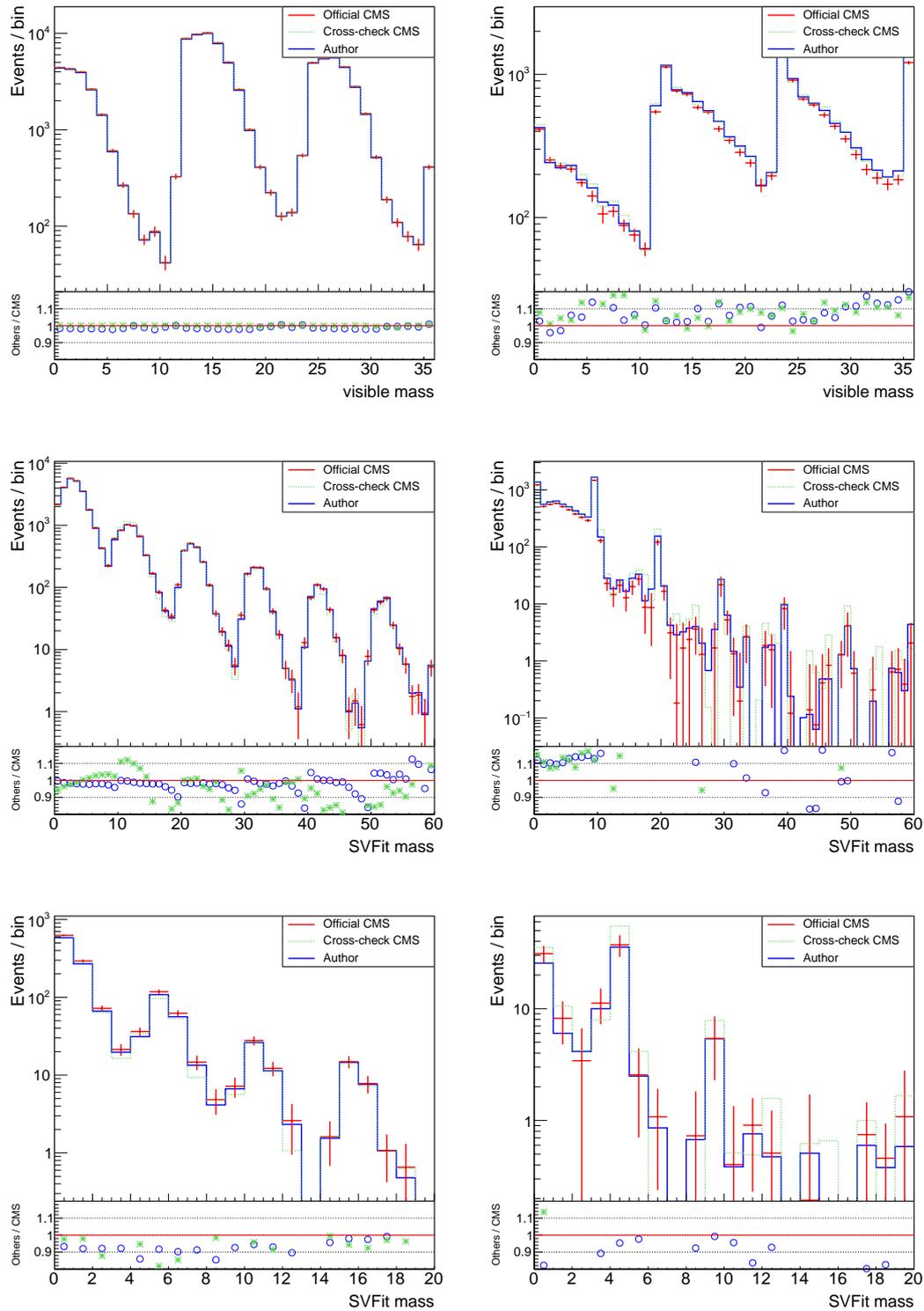


Figure 4.5: The events yield for ZTT (left column) and QCD (right column) backgrounds for Official CMS (red), Cross-check CMS (green) and Author (blue) analysis implementations. The 0jet (top row), boosted (middle row) and vbf (bottom row) are presented in different rows. The definitions of bins are taken from control plots (see Appendix C).

4.9 Official CMS analysis results

The global maximum likelihood fit is used to find the significance for the SM Higgs boson and $H \rightarrow \tau\tau$ signal strength (μ). All distributions mentioned at the end of previous section are combined to create the binned likelihood involving the expected and the observed number of events. The predicted number of events for the $H \rightarrow \tau\tau$ process, scaled by a signal strength modifier (a free parameter of a fit), is taken as expected number of signal events. The nuisance parameters in the fit representing the systematic uncertainties are varying according to their probability density functions (log-normal or Gaussian).

The excess of observed events in all decay modes and with respect to the SM background expectation can be visualized on a histogram of decimal logarithm of signal to signal-plus-background ratio (Fig. 4.6 left). The cleanest Higgs signal comes from a channel with two hadronic taus in the final state.

A profile likelihood ratio test statistic [95] is used to generate the local p -value (Fig. 4.6 right). The observed significance for the 125.09 GeV Higgs signal is 4.9 standard deviations compared to the expected significance of 4.7 standard deviations for 35.9 fb^{-1} at 13 TeV. The corresponding $H \rightarrow \tau\tau$ signal strength is $\mu = 1.09^{+0.27}_{-0.26}$.

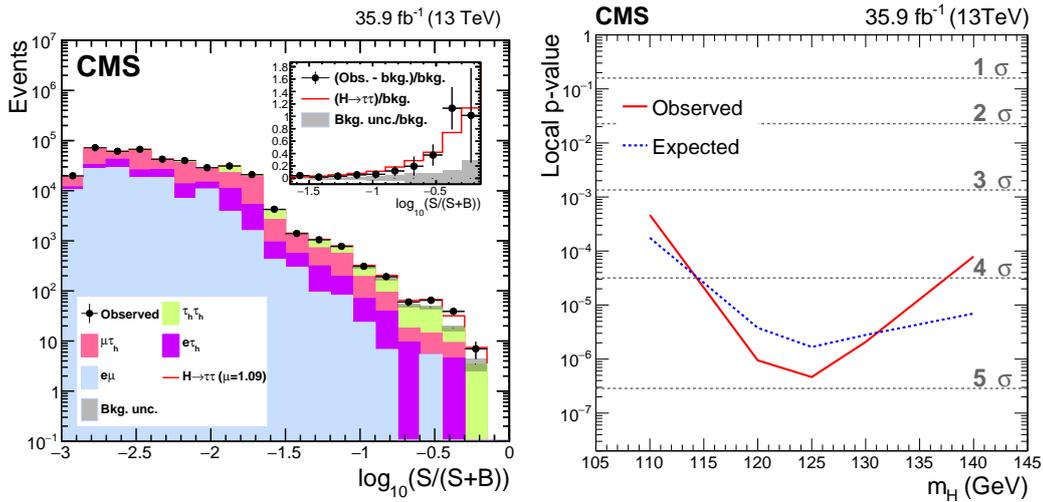


Figure 4.6: *Left*: Distribution of the decimal logarithm of the ratio between the expected signal and the sum of expected signal and expected background in each bin of the mass distributions used to extract the results, in all signal regions. *Right*: Local p -value and significance as a function of the SM Higgs boson mass hypothesis [82].

Systematic uncertainties

The 2.6% uncertainty is related to the integrated luminosity. Additional uncertainty of 7 to 20% is applied on the yield for Drell-Yan MC samples, depending on the final state and category. Similarly, the diboson and single top cross sections are estimated with 6% uncertainties. Uncertainties coming from theory and related to the cross sections, signal acceptances, PDFs, α_s , renormalization and branching fractions add up to 20% (see Table 4.8).

As the transverse momentum of leading jets influences the dijet system mass, the variations in jets energy may change the total number of events in bins of discriminating variable histogram in vbf category (see Section 4.3). There are almost 30 sources of systematic uncertainties related to the jet energy scale coming from e.g. absolute scale (mainly combined photon and muon reference scale and initial / final state radiation correction) uncertainty, single particle and quark / gluon response in ECAL and HCAL uncertainty or observed JEC time dependence due to calorimeter radiation damage. All JES uncertainties are considered as fully uncorrelated to avoid constraining jets unaffected by uncertainties related only to the specific regions of the detector (e.g. endcap-only ones).

In the $H \rightarrow \tau_\mu \tau_h$ final state up to 2.5% systematic uncertainty related to the MET energy scale is applied on the MC yields [82]. This effect is largest in 0jet category and comes from the relationship between MET and (muon) transverse mass (see Eq. 4.6 and Section 4.3).

The 1.2% uncertainty is related to the TES for taus. It is uncorrelated between decay modes and correlated between final states.

The 2% uncertainty is related to the trigger efficiency.

The uncertainty for processes where a jet is faking hadronic tau depends on the jet transverse momentum $p_{T,j}$ and is calculated as $20\% \times 0.01 \times p_{T,j}$. For jets with transverse momentum $p_{T,j} > 200$ GeV the flat 40% uncertainty is applied because of a lack of statistics.

In the 0-jet category in $H \rightarrow \mu \tau_h$ final state the events are separated using the reconstructed tau decay mode. The 3% uncertainty is related to differences in reconstruction/identification of hadronic tau in various decay modes. This uncertainty affects the shape of the distributions and not the overall normalization.

The 5% uncertainty related to data/MC MVA tau isolation ID correction is applied to the simulated events where tau is matched to generator hadronic tau (belongs to generator category no. 5).

The 25% uncertainty related to the $\mu \rightarrow \tau_h$ fake rate and applied to the $Z \rightarrow \mu \mu$ events is present in the $H \rightarrow \tau_\mu \tau_h$ final state.

The 0.3-0.5% uncertainty on the $Z \rightarrow \mu \mu$ peak position in the 0jet category is applied in $H \rightarrow \tau_\mu \tau_h$ events.

The 2% uncertainty is related to the lepton isolation and ID and applied on the yield for all MC-derived backgrounds¹⁰.

The shape uncertainty of the Drell-Yan background is computed by applying 0.9 or 1.1 times the LO MADGRAPH correction (see p. 56).

The 20% uncertainty is put on the $SF_{\text{QCD}}^{\text{SS} \rightarrow \text{OS}}$ scale factor (Section 4.6). An additional 20% uncertainty is applied on the QCD yield related to the QCD extrapolation between anti-isolated control region and signal region.

The Table 4.8 summarizes the systematic uncertainties relevant to the analysis.

¹⁰QCD (obtained from same-sign data events) and W+jets (derived from high- M_T^μ data region) are not affected in $H \rightarrow \tau_\mu \tau_h$ (see Section 4.6).

Source of uncertainty	Prefit	Postfit (%)
τ_h energy scale	1.2% in energy scale	0.2–0.3
e energy scale	1–2.5% in energy scale	0.2–0.5
e misidentified as τ_h energy scale	3% in energy scale	0.6–0.8
μ misidentified as τ_h energy scale	1.5% in energy scale	0.3–1.0
Jet energy scale	Dependent upon p_T and η	—
\vec{p}_T^{miss} energy scale	Dependent upon p_T and η	—
τ_h ID & isolation	5% per τ_h	3.5
τ_h trigger	5% per τ_h	3
τ_h reconstruction per decay mode	3% migration between decay modes	2
e ID & isolation & trigger	2%	—
μ ID & isolation & trigger	2%	—
e misidentified as τ_h rate	12%	5
μ misidentified as τ_h rate	25%	3–8
Jet misidentified as τ_h rate	20% per 100 GeV τ_h p_T	15
$Z \rightarrow \tau\tau/\ell\ell$ estimation	Normalization: 7–15% Uncertainty in $m_{\ell\ell/\tau\tau}$, $p_T(\ell\ell/\tau\tau)$, and m_{jj} corrections	3–15 —
W + jets estimation	Normalization ($e\mu$, $\tau_h\tau_h$): 4–20% Unc. from CR ($e\tau_h$, $\mu\tau_h$): \simeq 5–15 Extrap. from high- m_T CR ($e\tau_h$, $\mu\tau_h$): 5–10%	— — —
QCD multijet estimation	Normalization ($e\mu$): 10–20% Unc. from CR ($e\tau_h$, $\tau_h\tau_h$, $\mu\tau_h$): \simeq 5–15% Extrap. from anti-iso. CR ($e\tau_h$, $\mu\tau_h$): 20% Extrap. from anti-iso. CR ($\tau_h\tau_h$): 3–15%	5–20% — 7–10 3–10
Diboson normalization	5%	—
Single top quark normalization	5%	—
$t\bar{t}$ estimation	Normalization from CR: \simeq 5% Uncertainty on top quark p_T reweighting	— —
Integrated luminosity	2.5%	—
b-tagged jet rejection ($e\mu$)	3.5–5.0%	—
Limited number of events	Statistical uncertainty in individual bins	—
Signal theoretical uncertainty	Up to 20%	—

Table 4.8: Sources of systematic uncertainty. The acronyms CR and ID stand for control region and identification, respectively. The global fit to the signal and control regions may constrain the uncertainty – in such case the updated value is given in the third column [82].

Summary

In this chapter the following goals have been achieved:

- The $H \rightarrow \tau_\mu\tau_h$ analysis has been described along with its implementation made by the Author.
- It has been demonstrated that the presented implementation and the one used to obtain the public CMS result are in agreement. This allows to proceed with investigations oriented on machine learning based methods in discussed Higgs boson decay channel (Chapters 5 and 6).
- The official CMS results in $H \rightarrow \tau\tau$ have been reported. The obtained observed signal significance is 4.9 and the signal strength is $1.09^{+0.27}_{-0.26}$.

Chapter 5

Event identification using Machine Learning techniques

Introduction

Machine learning (ML) term was first introduced by A. L. Samuel (1901–1990) in his 1959 paper "Some Studies in Machine Learning Using the Game of Checkers" [96]. His simple algorithm was able to beat a local human expert in a game of checkers and already at this time justified the statement that it may be eventually economically feasible to use learning schemes to real-life problems. Today, ML becomes incorporated at increasing number of applications and is even called the leading force of the fourth (or third) industrial revolution [97, 98].

There is no single definition of ML, but one of the most popular states that it is the scientific study of algorithms and statistical models that computer systems use to progressively improve their performance on a specific task [99]. It is also often pointed out that ML is an application of artificial intelligence (AI) and allows computers to act without being explicitly programmed [100, 101]. Finally M. Stamp writes in his book that ML "is a form of statistical discrimination", where "the computer learns important information (...) trying to extract useful information from seemingly inscrutable data" [102]. This last quotation is particularly useful in justifying usage of ML for particle physics detector data, as all that people do in the analyses can be summarized as statistic-based extraction of information about the origin of visible particles.

Although learning process for the machine in many aspects resembles learning of a human being, it always has to be based on an algorithm which falls into one of four categories [101]. In the category called *supervised learning* the machine is ordered to find the function:

$$f : X \rightarrow Y \quad (5.1)$$

called the *classifier*¹, which is able to cast an observed object of interest (called *example* or *instance* or *observation* or *event* in our case) characterized by the set X of (input) variables $x \in X$ (*features*) onto a class $y \in Y$ (*label*). The category name comes from requirement of existence of supervised *training dataset*, i.e. labeled sample of data from which the classifier can be deduced (trained). The model can be evaluated using a *test dataset*, disjoined from the training one to ensure unbiased results. Additionally, we introduce *insample* as dataset comprised of training and test datasets used during models training and benchmarking, and *outsample* as test dataset for the final model

¹Strictly speaking the supervised learning can be divided into *Classification Learning*, *Preference Learning* and *Function Learning* [103]. Also, the classifier which is used to predict a categorical (discrete) outcome may be replaced by *regressor* used for the continuous response.

validation. In the insample the events can float between training and test datasets for consecutive trainings, which helps for discarding the overtrained models.

The event identification problem is an exemplary supervised learning problem: we are equipped with labeled training datasets (MC samples) and a set of features for every event to use.

The other three learning categories are:

- unsupervised machine learning algorithms. Here the training dataset is unlabeled. The model explores hidden structures/patterns from data instead of predicting the right output. Unsupervised learning often uses principal component and cluster analysis and allows modeling probability densities of given inputs [104];
- semi-supervised machine learning algorithms. The training is performed on two datasets: small labeled and large unlabeled. Using two datasets can improve the accuracy of the obtained model (e.g. in classification problem);
- reinforcement machine learning algorithms. Refers to goal-oriented algorithms acting in an environment and learned how to attain an objective using the reward-penalty system for the algorithm decisions. This approach is useful to e.g. creating gaming models. Reinforcement learning was used by the famous Google AlphaGo program to beat the European Go champion in 2015 [105].

For the last three categories, the learning method is not necessarily related to the concept of the classifier, therefore the broader term of *estimator* was coined. An estimator is a learning (or any) algorithm that fits (or trains) on data [106]. The estimator term can also refer to:

- high-level API (Application Programming Interface) encapsulating training, evaluation, prediction and exportation of the model (e.g. in TensorFlow or sklearn) [107];
- the method of finding the parameters for classifier or probability density for the model, e.g. Maximum Likelihood Estimator (MLE), Bayesian Estimator or estimator as optimization of loss criterion [108].

We give the above two meanings for a reference as these often come up in the literature. In this chapter, we will be using the estimator word in the first meaning exclusively.

In the previous chapter the $H \rightarrow \tau\tau$ analysis has been presented. Its main goal is to discriminate the Higgs boson signal from the background. Now we will generalize the problem to discriminating an arbitrary process (H or W+jets or $t\bar{t}$ and so on), called signal, from the remaining processes (backgrounds). In the following sections we solve this problem using a set of well established machine learning techniques (decision trees, random forests, boosted decision trees and neural networks). Created models will be evaluated using receiver operating characteristic curves. The performance of the best classifier found will be compared to the standard (cut based) analysis in the next chapter. The collision data samples will not be used for analysis in following chapters, because they do not provide the generator information (used for classification)².

²Please note, that trained model can be used for collision data as we will not use generator information as model input variables. However, the interpretation of results would not be trivial and goes beyond the scope of this thesis.

The chapter is organized as follows. In Section 5.1 the ML analysis structure is described, together with output classes definitions, used data samples and features standardization. Sections 5.2 – 5.5 are devoted to present various ML techniques, starting from the simplest model (Decision Tree) to the most advanced (Neural Network). Decision Trees are used to introduce the ML model performance metrics. This chapter ends with a short summary of the tested classifiers.

5.1 Initial analysis setup and data preprocessing

The output classes (categories) defined for the analysis are given in Table 5.1. The samples used are the same as in Chapter 4, with exception of additional QCD, ggH and VBF datasets used in order to improve statistics and enumerated in Appendix A. The data-driven methods for QCD background extraction are not used³.

Category Process	ggH125 $gg \rightarrow H \rightarrow \tau_h \tau_h$	qqH125 $qq \rightarrow H \rightarrow \tau_h \tau_h$		WH125 $W+H \rightarrow \tau_h \tau_h$	ZH125 $Z+H \rightarrow \tau_h \tau_h$
Category Process	ZTT $Z \rightarrow \tau_h \tau_h$	ZL $Z+\text{jets} \rightarrow ll,$ $l \rightarrow \tau_h$	ZJ $Z + \text{jets} \rightarrow ll,$ $\text{jet} \rightarrow \tau_h$	W W + jets	EWKZ Electroweak Z boson plus two jets
Category Process	QCD QCD	TT $t\bar{t}$	ST Single top quark	VV Diboson	Other Processes which do not fall into any other category

Table 5.1: Output classes for ML classification.

All events used have to contain only one lepton pair and are required to pass baseline selection (see p. 50) and event vetos (see p. 51). We however remove the constrain on the very loose anti-electron discriminator for tau lepton. This step is undertaken in order to further increase the number of available events for training, works efficiently for the QCD events and does not significantly change overall input variables distributions. The loss of anti-electron discriminator variable is partly compensated by adding the raw (float) discriminator value to the features list⁴. The events number gain versus standard analysis are presented in Table 5.2.

Because the format of binary (ROOT) files [109] used in the CMS analysis is not compatible with machine learning tools, the data files are converted to the HDF5 data format [110]. The new files contain a two dimensional array (matrix), with features values in columns and consecutive events in rows. A full list of features (available for every event) is presented in Appendix D.

In the standard analysis all available events in a given category undergo the same procedure (i.e. application of cuts and weights), therefore their order is not relevant and the manner of the category dataset creation (by default concatenation of different data samples, e.g. Drell-Yan with zero, one, ... number of jets) does not effect the result. Before proceeding with ML computation we have to split the samples into insample and outsample, so it is crucial to ensure the same properties of those datasets inside a single category (e.g. similar distribution of number of jets). This

³Please note, that if data-driven methods were used, the lepton charge would have been removed from the features list. This is because the QCD events from the data are same sign ones.

⁴The usage of word "partly" comes from the fact that discriminator working points are not simple cuts on the raw discriminator value, but are dependent on the tau-specific category and tau transverse momentum.

No.	category name	standard analysis	machine learning analysis	events gain [%]
1	ggH125	29571	53242	80
2	qqH125	47529	134254	182
3	WH125	37745	43475	15
4	ZH125	16730	19149	14
5	ZTT	279635	316711	13
6	ZL	22824	28880	27
7	ZJ	599171	693878	16
8	W	3687419	4235549	15
9	EWKZ	18501	21192	15
10	QCD	-	63492	-
11	TT	3197220	3732232	17
12	ST	1720897	2003396	16
13	VV	2903838	3417318	18
14	Other	304687	357906	17

Table 5.2: Comparison of number of events per category after baseline selection.

is done by shuffling the events using pseudorandom number generator with given seed to retain reversibility and reproducibility.

The common procedure in ML during a preprocessing step is features standardization [111]. In this process the values of the (feature) distribution are rescaled to some presumed range. The reasons for doing this will be given later in this chapter, and for now it will just be shown how it is done here. From the whole bunch of different standardization methods, two have been found useful in this analysis: *standard score* (z-score) and *normalized empirical cumulative distribution function* (normalized ECDF).

The standard score is arguably the most popular standardization technique in ML. For random variable⁵ (r.v.) X the standardized r.v. Z is given as:

$$Z = \frac{X - \mu}{\sigma}, \quad (5.2)$$

where μ is mean (expectation value) and σ is standard deviation of X . The standard score is the standard statistical tool when comparing observations with different normal distributions [113], because it does not change the shape of the distributions. The standard score normalized distribution of the SVFit mass (see Section 3.4) is shown in Fig. 5.1. From the plot one can deduce two conclusions: 1) the standard score changes only the x -axis normalization (not the shape) and 2) the SVFit mass distribution is not the normal distribution.

Since many ML techniques assume normally distributed input variables, we will deal with the fat-tailed (i.e. with large skewness or kurtosis) distributions taking advantage of the probability integral transform [114]. Every continuously distributed

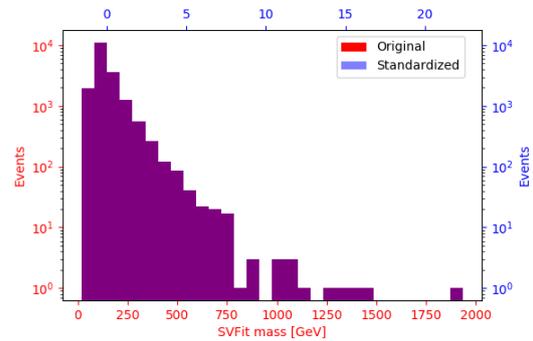


Figure 5.1: The SVFit mass distributions for ZH125 output class before and after standardization. In the figure the nominal distribution (in red) overlays the distribution standardized using standard score (in blue) giving violet. The standard score is computed with $\mu = 146.15$ GeV and $\sigma = 82.04$ GeV. The bottom and the left axis labels refer to the nominal distribution and the top and right labels to the standardized distributions.

⁵Random variable is a function defined on a sample space \mathcal{S} and taking values in the real line $\mathcal{R} = (-\infty, \infty)$ [112].

feature X can be standardized to the r.v. Y with a uniform distribution:

$$Y = F_X(X), \quad (5.3)$$

where F_X is a cumulative distribution function (CDF) of a feature. The CDFs of features are not known, therefore CDF version based on the observed data (ECDF) is used.

The ECDF is defined as:

$$F_n(t) = \frac{1}{n} \sum_{j=1}^n \mathbf{1}_{\{Z_j \leq t\}}, \quad (5.4)$$

where $\mathbf{1}_\epsilon$ is the indicator of the (countable, finite) set ϵ , Z_j is a value j of features set Z and Z is sorted [115]. In this analysis the ECDFs were generated using a sub-sample of 1000 events per sample. Please note, that the ECDF is r.v. itself and exhibits small fluctuations. Also note that because the uniformity of Y is preserved only for continuously distributed r.v., the ECDF-based standardization will not be applied for discrete-valued features (e.g. number of jets) [116]. The uniform distribution Y can be converted into a normal distribution using inverse transform sampling [117]. The result for the ZH125 sample SVFit variable is shown in Fig. 5.2.

The insample is created by concatenation of shuffled (standardized or not) samples, by taking first half of events of each sample. The remaining events are used to create outsample⁶.

After preprocessing, it is time to proceed with the ML computations. Since the standard analysis is based on the complex chain of selections, it is natural to start with estimator that is the closest to this workflow: a decision tree. The description of decision trees given in the next section will be followed with an introduction to the methods of benchmarking the performance of the ML models.

5.2 Decision Tree and ML performance measurements

Decision Tree (DT) is an estimator which recursively partitions the events space [118]. The single partition is realized by the *if-else* cut-like statement on feature(s) and splits the space into sub-spaces [119]. In case of this analysis (which is also by far the most frequent case), the cut is performed on a single feature and the space is always divided into two sub-spaces⁷. At the end, all partitions are assigned to the classes having the largest representation of events in a given sub-space.

⁶We took no more than 400k events per sample, as we noticed it accelerates the computations without sacrificing the quality (performance) of results (model).

⁷The overrun of the single-feature-per-split limitation would require at least $\mathcal{O}(n^m)$ complexity algorithm, where n is a number of events and m is a number of features considered in the single condition. This in turn would seriously restrict the maximum number of events available for training.

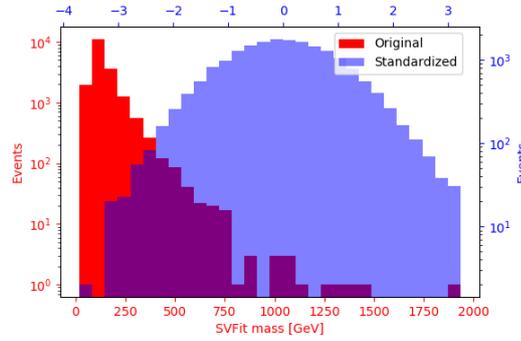


Figure 5.2: The SVFit mass distributions for ZH125 output class before and after standardization. In the figure the nominal distribution (in red) is plotted together with distribution standardized using empirical cumulative distribution function (in blue). The axis labels are as in Fig 5.1.

The above procedure can be identified with – and takes its name from – the concept of a *tree*, which in graph theory is an undirected graph with vertices connected by exactly one path⁸ [121]. The DT starts from the *root*, which is the sole node without incoming edges. The remaining nodes have one incoming edge and either multiple (usually two) or no outgoing edges. The former are called *internal (test) nodes*, the latter *leaves (or decision nodes)* (see Fig. 5.3).

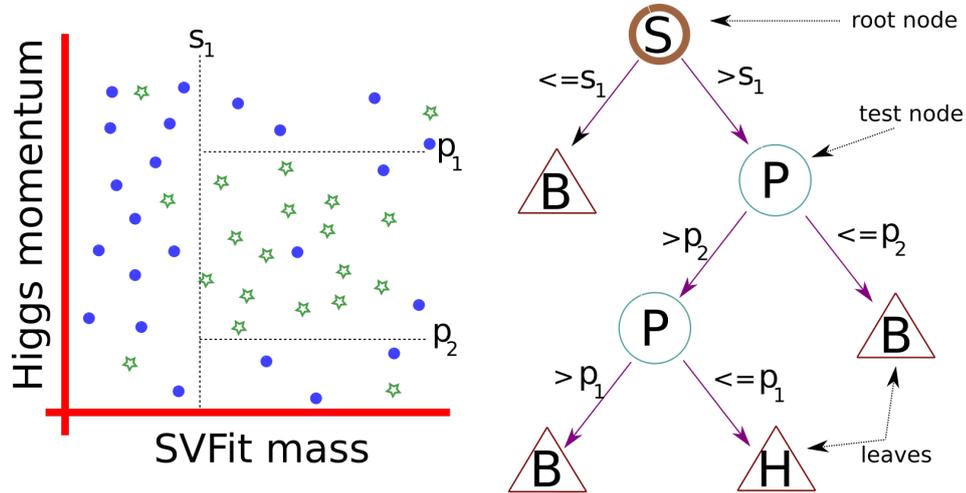


Figure 5.3: The illustration of the Decision Tree estimator. *Left*: 2D events space. The two features – SVFit mass (s) and Higgs momentum (p) – are used to split the space to background (blue dots) and Higgs (green stars) events. *Right*: matching graph. In the root and test nodes, the letters S and P indicate the feature for which the split is performed. In the leaves the letters B (bkg.) and H (signal) indicate the output class of the classifier. The events distribution in this example is illustrative and not a real one.

Finding the optimal DT algorithm is at least NP-hard⁹ [122, 123], therefore the heuristic methods (usually top-down¹⁰) are required and implemented. The common interface for these methods consists of two steps: *tree growing* followed by *tree pruning*.

At the tree growing step, the tree is expanded by iterative execution of the code at each data node. Two things may happen: 1) if the *stopping criterion* is satisfied, the node is converted into a leaf and marked with an appropriate class; 2) two (for the binary tree; denoted as *left* and *right*) or more new nodes are created and connected to the original node using *split criterion*. The examples of a stopping criterion are as follows:

- The maximum depth of the tree (i.e. the length of the longest path from a root to a leaf) has been reached;
- The leaf is pure, i.e. all events in the node belong to a single class;
- The node contains less than the given minimum number of events required to split;

⁸A tree is a connected non-empty graph with no circuits [120].

⁹NP stands for Nondeterministic Polynomial. In complexity theory of decision problems the problem is NP-hard if it can be translated into one for solving any nondeterministic polynomial time (i.e. solvable in polynomial time by a nondeterministic Turing machine) problem.

¹⁰The method is called *top-down*, because it is designed to look at the entire data in each step - decision tree is initiated by choosing the most efficient single variable to split the whole dataset and then grown from root to leaves in a recursive manner. In contrary, in bottom-up approach the DT is build from leaves to the root.

- The split does not leave the minimal number of events in each of the branches;
- The split does not leave a minimum weighted fraction of the sum total of weights (of all the input samples) in each of the branches;
- The split does not provide the minimum impurity decrease;
- The impurity is below a given threshold.

The aforementioned *impurity* is a function $\phi : P \rightarrow R$, where $P = (p_1, p_2, \dots, p_k)$ is a distribution of k discrete values of the r.v. and $p_i \in [0, 1]$ (p_i is related to the proportion of the r.v. belonging to i -class), which satisfy [119]:

- $\phi(P) \geq 0$;
- $\phi(P)$ is minimum if $\exists i$ such that component $p_i = 1$;
- $\phi(P)$ is maximum if $\forall i, 1 \leq i \leq k, p_i = 1/k$;
- $\phi(P)$ is symmetric with respect to components of P ;
- $\phi(P)$ possesses derivatives of all orders in its domain i.e. is C^∞ function.

Intuitively, the impurity can be understood as a measure of homogeneity (or heterogeneity) of the dataset (or subset) with respect to the output class. If a given dataset contains events belonging to several classes, then we say the dataset is impure or heterogeneous. The impurity is largest when all classes are equally represented in the dataset. If the dataset contains events belonging to a single class it is pure or homogeneous. If we define the proportion of the samples that belong to the class $k \in Y$ at node t as:

$$p(k|t) = \frac{1}{N} \sum_{z_i \in R_t} I(y_i = k),$$

where the sum is over events z_i in region (partition) R_t , then the popular impurity functions are given by the formulas:

$$\phi = \sum_k p(k|t)(1 - p(k|t)) \quad \text{Gini impurity,} \quad (5.5)$$

and

$$\phi = - \sum_k p(k|t) \log(p(k|t)) \quad \text{Entropy impurity.} \quad (5.6)$$

The Gini impurity returns the probability of incorrect classification of a randomly picked event out of the dataset. Entropy (*information entropy* or *Shannon entropy*) is a measure of unpredictability of the state and was introduced in 1948 by C. Shannon [124]. The base for the logarithm is equal to two, although in principle could be any number. Entropy has desired property of being additive when the probabilities are multiplied¹¹ [125], but also has a drawback of being computationally heavy.

Impurity can be used as a base for defining the split criterion. Let us denote *split* as $\theta = (x, t_p)$, where x is a feature and t_p is the cut value. Then for a dataset Q_p (where $Q_p = Q_- \cup Q_+$ and $Q_- = (x, y) | x_i \leq t_p$ for a binary tree) we can define the

¹¹Please note, that for very small probabilities $p(k|t)$, the multiplication (Gini index) can lead to rounding error.

information gain G as:

$$G(Q_p, \theta) = \phi(Q_p) - \sum_{j=1}^m \frac{N_j}{N} \phi(Q_j) \stackrel{\text{binary tree}}{=} \phi(Q_p) - \frac{N_-}{N} \phi(Q_-(\theta)) - \frac{N_+}{N} \phi(Q_+(\theta)), \quad (5.7)$$

where N and $N_{j,\pm}$ are the number of observations at parent node p and child nodes j ("±" for left and right), respectively. Information gain measures how much information a split gives about the class as we go down the tree. The split criterion used in many DT algorithms (e.g. CART) obtains the split by maximizing the $G(Q_p, \theta)$.

DTs tend to overtrain themselves for loose stopping criteria (for a finite training set one can always have one-element regions, unless features vector is the same for multiple classes). On the other hand, the tight stopping criteria often end up in underfitted models. Finding the right balance for the stopping criteria is a difficult task, therefore an alternative method was developed by L. Breiman *et al.* [126] called pruning. In this methodology, the overfit tree built with loose criteria is simplified back by cutting out branches that do not add up to the generalization accuracy. The *reduced error* is an example of a pruning technique, where the tree is traversed bottom-up checking if replacing an internal node (non-leaf subtree) with the most frequent class (best possible leaf) reduces tree accuracy (increases misclassification rate). If not, the node is removed.

The DTs have multiple advantages: they are easy to visualize and interpret¹², do not require data standardization, the prediction algorithm has a low computational complexity ($\mathcal{O}(\log(n))$), can handle multi-class output variables and are easily statistically validated. The DT makes also no assumptions about underlying data distributions i.e. is a nonparametric learning method and every situation is observable in the model (it is a white box model). However, they are susceptible to over(under)fitting, sensitive to the changes in the training samples and finding the optimal DT is practically impossible for more complicated scenarios. The cure for this last drawback of DT – heuristic methods – is based on greedy algorithms [127], which are suboptimal.

In this analysis the sklearn [128] package is used, where DT implementation exploits the *Classification and regression trees (CART)* technique introduced in [126]. CART produces binary trees finding the best split with information gain (Eq. 5.7), with either Gini (default) or Entropy impurity, and with pruning of the final tree.

In order to initially check the performance of the DT and the validity of the procedures used in this chapter, the following category recognition exercise has been created:

- For all analysis categories (0jet, boosted and vbf) every defining condition (see Table 4.5) is cast onto Boolean variable;
- For all events, the set of features is created comprised of numbers created above;
- The output classes for DT model are defined as category names, plus "trash" class for events which do not fall into any of the categories;
- The model is trained to recognize the proper category (or trash) for the event. In other words, the DT has to figure out the categories definition from the data.

¹²Although the size of a classification tree is strongly correlated to the size of the training set.

The visualization in form of a graph of the resulting model is presented in Fig. 5.4. DT managed to find the ideal solution for the defined task and classify all events in validation sample properly.

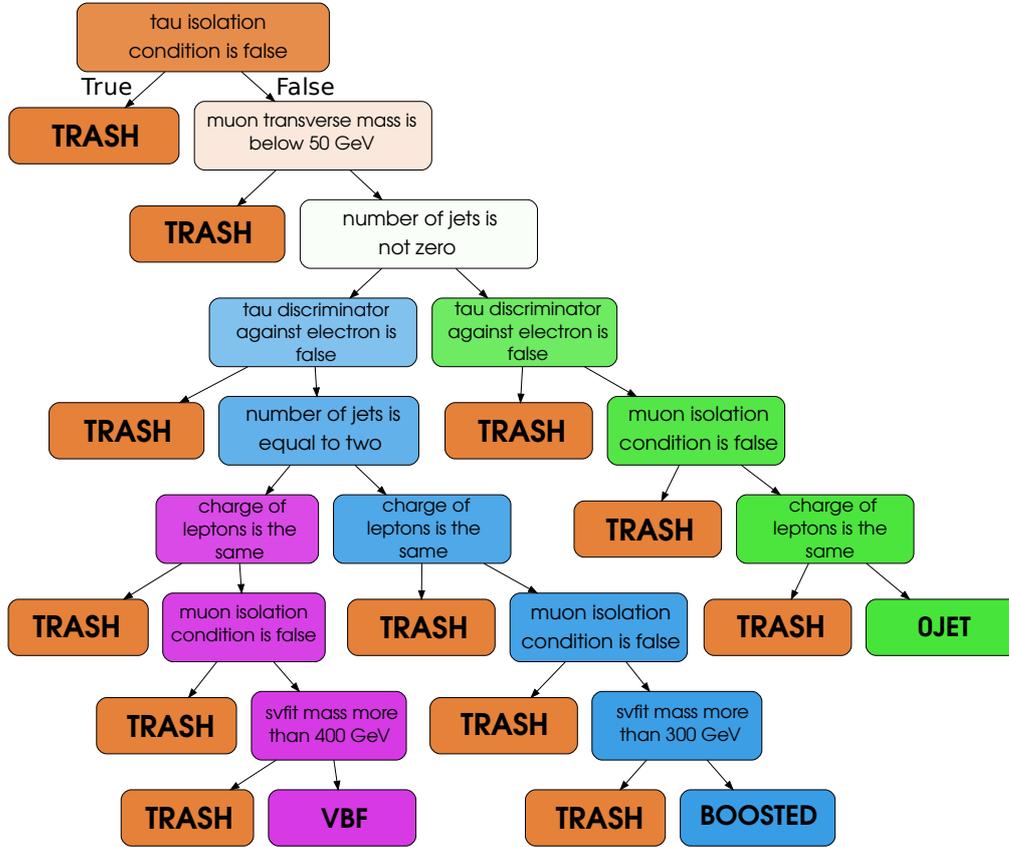


Figure 5.4: Decision tree model visualization for category recognition problem as a result of a ML training process. In boxes are given either condition (in case of nodes, see Table 4.5) or output class (in case of leaves). The Boolean variables given at nodes positions are in human-readable form. Different colors correspond to different output classes.

There are several well established measures for evaluation of the model i.e. quantification of the effectiveness of a classification strategy. In this thesis the *confusion matrix*, *receiver operating characteristic curve (ROC curve or simply ROC)* and *ROC Area Under Curve (ROC AUC)* are used.

5.2.1 Confusion matrix

Confusion matrix (or *error matrix*) is used to visualize the performance of the algorithm. The rows of the matrix represent the actual class instances, whereas the columns the predicted class ones [119]. The cells of the matrix contain the measure of the amount of correctly or incorrectly classified events. Depending on the approach the numbers represent:

- the number of classified events - for unweighted, non-normalized case;
- the yield of classified events - for cross-section weighted, non-normalized case;
- the percentage of number of events/yield - for normalized case.

The four definitions are usually introduced to help with the model description:

- True positive (TP) - the number/yield/percentage of correctly assigned events for the class under consideration. TP lies on the intersection of actual row and predicted column for a given class in the confusion matrix.
- False positive (FP) - the number/yield/percentage of incorrectly assigned events as belonging to the class under consideration. FP is the sum of the elements on the predicted column minus TP in the confusion matrix;
- False negative (FN) - the number/yield/percentage of incorrectly assigned events as not belonging to the class under consideration. FN is the sum of the elements on the actual row minus TP in the confusion matrix;
- True negative (TN) - the number/yield/percentage of events correctly assigned as not belonging to the class under consideration. TN is the sum of all elements in the confusion matrix that do not enter into TP, FP or FN.

The variables above work particularly well in the two output classes scenario, where they are given by the appropriate numerical value in a single confusion matrix cell (for more than two output classes they are arrays). For the multi-class output, the example of the confusion matrix is presented in Fig. 5.5. Please note that the sum of the elements in every row is one.

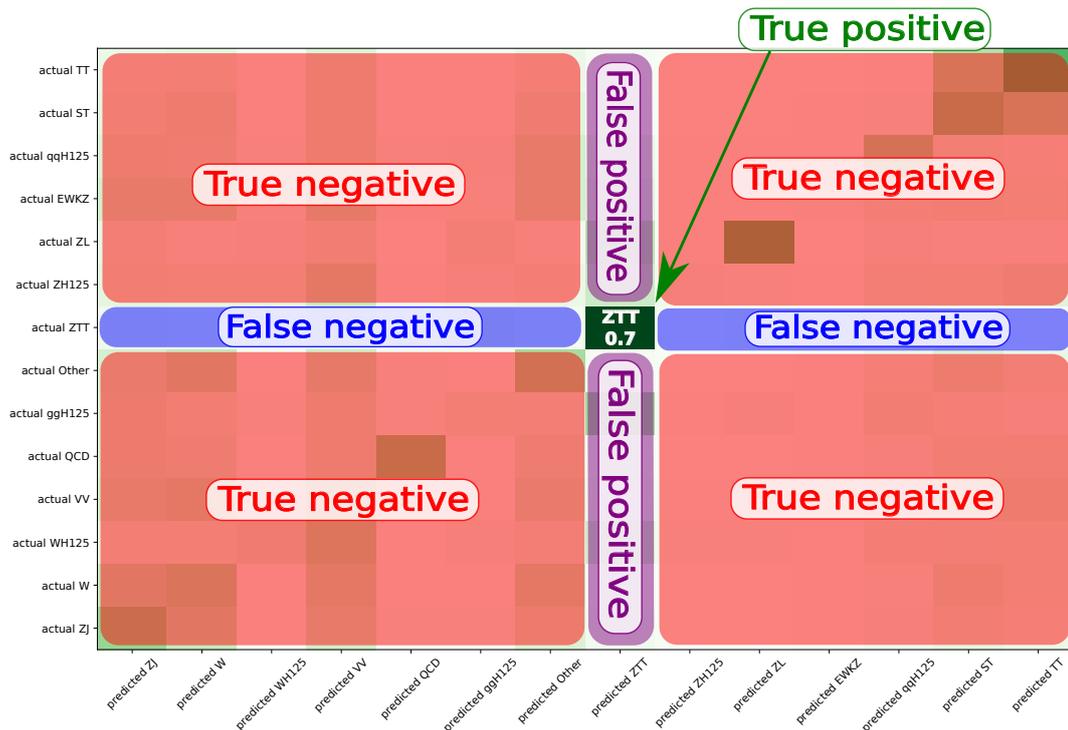


Figure 5.5: The illustration of relation between confusion matrix elements and true/false positive/negative for ZTT output class. The sum of the numbers under areas colored with green, blue, violet and red constituent TP, FN, FP and TN respectively.

To compute the confusion matrix C the vector of predicted (scores) and actual (target) values are required. Let us denote these two objects with A and B . Then we have:

$$C = A^T \cdot B$$

In this analysis both A and B are of (*number of classes* \times *number of events*) shape, therefore C has *dimension* = (14, 14). In order to normalize the confusion matrix, all elements need to be divided by sum of the events (in rows). Furthermore, to obtain cross-section weighted result, every event is multiplied by an appropriate weight.

The normalized confusion matrix for classification recognition model is very simple: as the model returns perfect results the confusion matrix is 4×4 identity matrix. The problem with confusion matrix is that it does not allow for simple model comparison in more complex cases.

5.2.2 ROC and ROC AUC

The quantities introduced in the previous section can be used to define appropriate rates. The *True Positive Rate* (TPR):

$$TPR = \frac{TP}{TP + FN}$$

is called the *sensitivity*, and gives the fraction of events that are correctly classified out of all events of a given class. The sensitivity is also often called *efficiency* in HEP, e.g. on the trigger related field. The *True Negative Rate* (TNR)

$$TNR = \frac{TN}{TN + FP}$$

is called *specificity*, and gives the fraction of events that are correctly not classified into the class under consideration. Please note that models which classify all events as signal (background) have sensitivity equal to one (zero) and specificity equal to zero (one). Therefore, the models with both high sensitivity and specificity are desirable. The (*balanced*) *accuracy* of the model is often introduced as the half of the sum of these two quantities. Although the domain of such accuracy is $[0, 1]$, it is assumed that for the models with accuracy below 0.5 it is always possible to reverse the classification criteria and obtain a result above 0.5.

The accuracy would be the perfect measure of goodness of a model, but there is an issue with it. The majority of the estimators do not generate the binary results (i.e. simple output with the name of the output class), but return the *score* i.e. the measure of the probability for the event to belong to the given output class¹³. For these models the rates – and hence the accuracy – can be modulated by the introduction of the thresholds on the score(s). The threshold independent model evaluation is possible with the ROC curve.

The ROC is a graphical plot of $\langle \text{sensitivity} \rangle$ against $\langle 1 - \text{specificity} \rangle$ for varied threshold values.¹⁴ There are few features of ROC that are readily deduced:

- ROC points are always equal to/above the $[(0,0), (1,1)]$ line, and ROC lying on this line corresponds to a random-walk model. The reasoning is the same as for the accuracy – a very bad model can be transformed into a better one by reversal of classification criteria.
- For binary output estimators (like DT), ROC is a single point. This stems from the fact, that the threshold for the binary score cannot be set;

¹³In fact, decision trees are the only binary models that will be considered here.

¹⁴The $\langle 1 - \text{specificity} \rangle$ is also called *fall-out* or *false positive rate* and is equal to $\frac{FP}{FP+TN}$. It may sometimes be confused with *false discovery rate* or *fake rate* – often used in physics and defined as $\frac{FP}{FP+TP}$.

- For the continuous score domain models, ROC is created by connecting the points for all threshold cuts. The line always contains (0,0) point (for sufficiently high score threshold no event is classified into any of the output classes) and (1,1) (for sufficiently low score threshold all events are classified into any of the output classes).
- ROC is defined for a binary classifier system. As mentioned earlier during confusion matrix description, for a model with more than two output classes, the TP(R) and TN(R) are given as arrays. Hence, for such models the number of ROC curves equal to the number of output classes need to be plotted.
- The perfect point on the ROC would be (0,1), for which all signal events are classified correctly and no background events are misclassified as signal. This case is realized in classification recognition exercise.

For convenience and to be able to compute ROC AUC even for the single-point ROC the line to (0,0) and (1,1) points are always drawn. The illustration of the ROC curve idea is presented in Fig. 5.6. The ROC curves in following sections will be generated using sklearn package implementation [129].

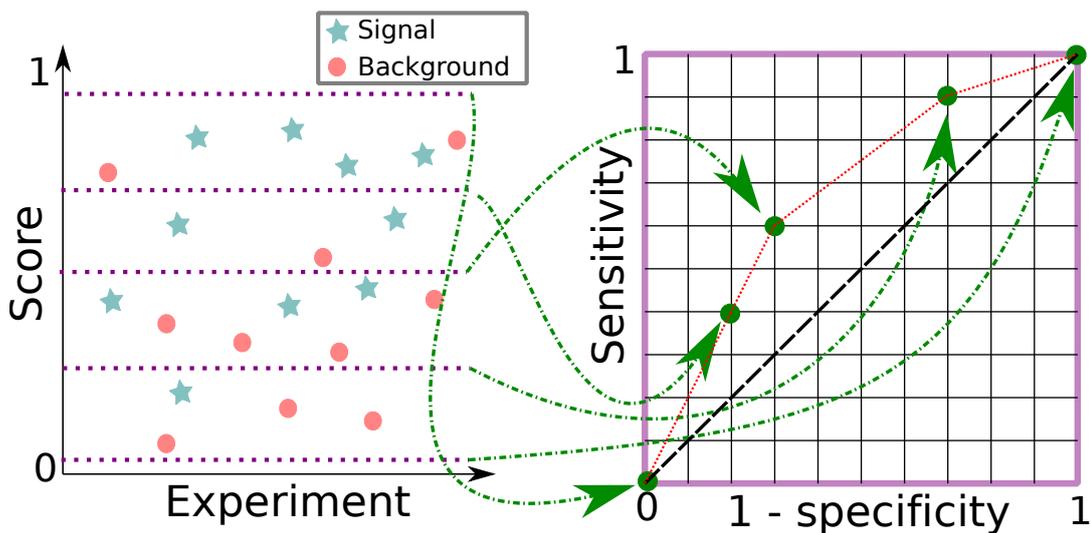


Figure 5.6: An illustration of the receiver operating characteristic curve. On the left is the output of the exemplary ML model for signal (green stars) and background (red dots) events. The violet dashed lines indicate the different threshold cuts. For every cut we can compute the sensitivity and specificity (defined in text) and mark (green dots) on the right hand side plot. The ROC curve is created by connecting the points.

ROC can serve as a rough model performance estimation, but its practicality for different models comparison is confined only to models in which ROC curves do not intersect. Since this is a rare case in the real world, it is useful to introduce yet another measure for that purpose – ROC AUC.

ROC AUC corresponds to the probability for the classifier to assign to the randomly selected signal event a score higher than to the randomly selected background event. This is also the most common statistic used to compare models in the machine learning community [130]. For classification recognition exercise all four ROC curves have AUC equal to unity.

In this chapter we are training classifiers with multiple output classes, therefore there are multiple ROC curves per model. The dominance relationship between

classifiers will be established with the mean of all ROC AUCs calculated for every model. The common allegation here is that since the Higgs samples are underrepresented due to smaller datasets and cross sections¹⁵, the raw ROC AUCs is no good for model evaluation and it should be somehow reweighed. Moreover, the standard cut based analysis is oriented on maximizing the signal-to-background ratio, which should be translated into maximizing the ROC values for a specific choice of working point (threshold) and not average ROC AUC. It is hard to disagree with the statements above, but we have decided to stick with mean ROC AUC approach because of two reasons. Firstly, the ROC AUC (as opposed to e.g. accuracy) is not sensitive to the imbalance problem [131], i.e. it will not be high if the minority class is badly predicted. The model with bigger (mean of) ROC AUC(s) will be better [119], regardless of samples representation¹⁶. Secondly, we wanted to avoid narrowing down the analysis by doing only "Higgs boson signal vs backgrounds" discrimination. The information about the origin of the event (not only if it is Higgs-like) is valuable for defining/constraining the contamination of various backgrounds of physical processes for control and signal regions, and can be used to improve the standard analysis results.

5.2.3 DT model results

Before creating a final model the impact of the size of the training dataset on the overall model efficiency has been checked. This is done by selecting a subset of insample events, training the model using this subset as training dataset, and computing the ratio of properly classified events to all events for the outsample. The last step is straightforward for decision trees as they return only the name of the output class for a given event. The results are given in Fig. 5.7. We found that the efficiency grows monotonically with the number of training events, which confirms the validity of the implemented procedure.

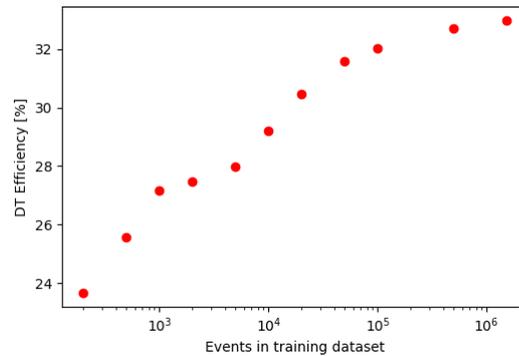


Figure 5.7: The efficiency (percent of correctly recognized events) versus the size of the training dataset. Errors are below the size of the points, e.g. for 5000 events set DT efficiency is equal to 27.98 ± 0.14 .

The final decision tree classifier [132] has been trained on full insample. The confusion matrix of the model is presented in Fig. 5.8. We can see that obtained estimator has the biggest troubles in discriminating between single top and $t\bar{t}$ and also between ZJ and W+jets processes. ZTT and ZL classes are on the other hand well distinguished. The ROCs with areas under curves values are presented in Fig. 5.9. The decision tree model works relatively well with ROC AUCs mean of 0.611. The best discriminated output class is ZTT with ROC AUC of 0.85 and the worst discriminated is EWKZ for which the model behaves the same as would random guessing (ROC AUC is equal to 0.5). The Higgs boson signal processes (ggH, qqH

¹⁵This is usually called the *imbalance problem*.

¹⁶There is one additional effect: models usually trains separately for different output classes (true especially for deep learning estimators) and only a small subset of trained models actually provides any discriminating power, so any reasonable ROC curves reweighing would not change the final model choice in a vast majority of scenarios.

and Z/W+H) are poorly classified with average ROC AUC of 0.56, from which the best performance is obtained for vector boson fusion Higgs boson production (ROC AUC = 0.62).



Figure 5.8: Confusion matrix for DT model

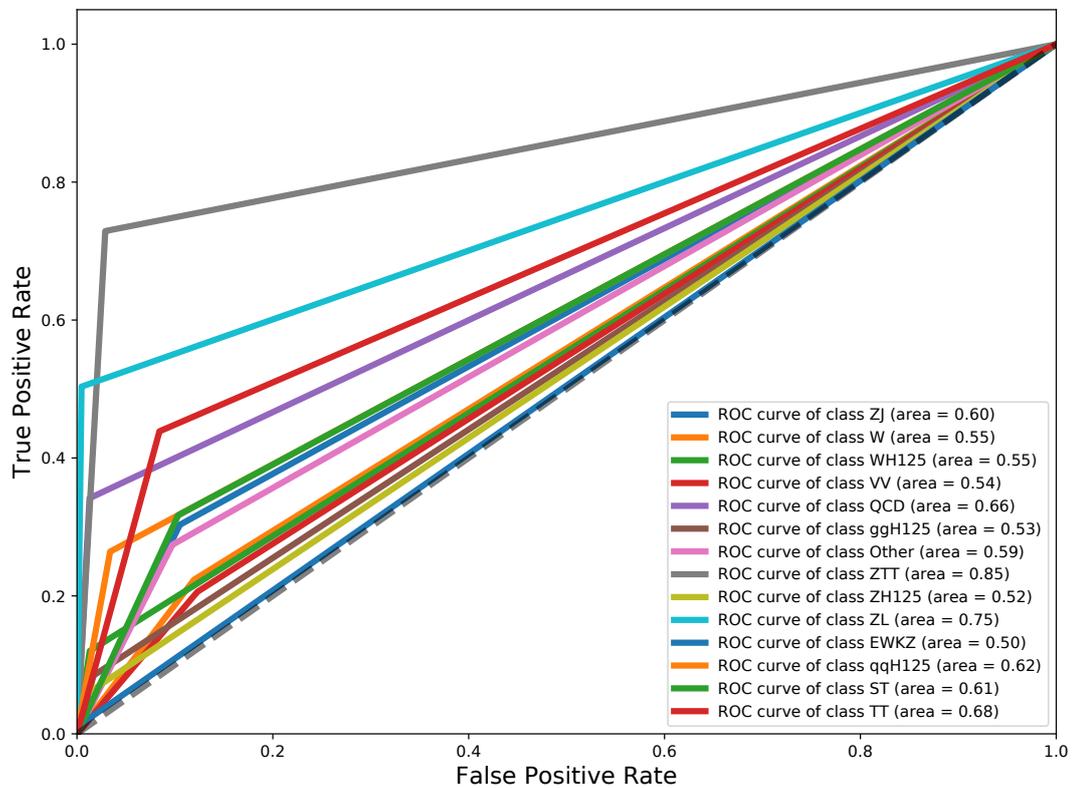


Figure 5.9: ROC curves for the DT model

5.3 Random Forest

The problem with the decision tree getting inevitable overfit to training data when increasing its complexity is solved in CART technique by pruning. This approach of building reasonably general tree by lowering the complexity of fairly overfit tree was questioned by T.K. Ho [133], who proposed an alternative method where the accuracy of the estimator on the train *and test* sample increases simultaneously, and with tree complexity¹⁷. In his method called Random Forest (RF) the estimator is built out of a *multitude of unpruned decision trees* (hence *forest*) for which only a subset of features is used at each node during tree growing. For every tree the features are selected *at random* independently for every node. This is done to reduce the correlation between decision trees in an ensemble [136]. The maximal number of input features considered in the node is one of the main estimator parameters that can be set before the training. In this analysis we will try the following values:

```
max_features = [0.1, auto, 0.9],
```

where numbers represent the fraction of features used¹⁸ and *auto* option sets the parameter to be equal to square root of the number of available features. The lower the value of this parameter the lower is the variance of the base estimators (DTs) and higher bias of the combined estimator (RF).

The idea of Ho was extended by L. Breiman [137], who introduced *bootstrap aggregating* ("*bagging*"). He proposed to train every decision tree in the forest not on the full training sample, but using only a randomly chosen subset of events¹⁹. Bagging helps to improve the accuracy and stability and constrains overfitting of the RF. It also reduces the variance of the base estimators.

The RF algorithm is able to return the prediction for the probability for a given output class. It is given as the mean of the class probabilities predicted by all trees in the forest. The class probability for a single tree is in turn computed as a fraction of events of the class under consideration present in the DT output leaf event space [138].

Most of the advantages of the DT are also valid for the RF: the data standardization is not required, it runs efficiently even for large datasets and can handle a multitude of input features, the model can be visualized, and gives an insight into variables importance in the problem (although it is not strictly a white-box estimator). Moreover, the RF does not overfit and even though a single decision tree in the forest is likely more biased and less accurate than this given with CART technique, the RF is on average much more accurate.

In order to find the best RF estimator, the training is repeated for different configurations of model parameters (hyperparameters). Except for the maximal number of features in the node, the following quantities have been modified:

- number of decision trees in the forest:
`n_estimators = [15, 30, 45, 80],`

¹⁷The resistance to overtraining for the forest method can be explained in a frame of stochastic discrimination theory [134, 135]

¹⁸In other words if n is a total number of features, then we at random pick $0.1 * n$ or $0.9 * n$ of features to be used to maximize the information gain. The features that have not been selected can be taken into account at the other node.

¹⁹In bagging method the events are selected with replacement, and the size of the training subset is kept the same as the full sample.

- impurity function for base estimators:
`criterion = ["gini", "entropy"],`
 where names are self-explanatory and refer to Eqs. 5.5 and 5.6.
- minimal number of events required to split the node:
`min_samples_split = [2, 8, 0.05],`
 where the default value is 2. The floating value indicates the fraction of events required out of the whole training dataset.
- the seed used by the random number generator.
`random_state = [1, 2, 3].`
 This is used to check, if the events order/choice influence the results.

The models are compared with ROC AUC (the higher the better). To improve the stability of the comparison the cross-validation is implemented. First, the insample is split into four. Then the model is trained four times using a sum of three (each time different) subsets and validated on the remaining one. Finally, the ROC AUC of the model is given as an average ROC AUC for these four trainings. The results for 216 available combinations of parameters are summarized in Fig. 5.10. The worst obtained ROC AUC is 0.767 ± 0.001 , and the best is 0.849 (with standard deviation less than $5 \cdot 10^{-4}$). The amount of trees in the forest and the number of events required to split the node turned out to have the biggest influence on the model accuracy. On the other hand, the random generator seed value does not change the performance of the model, which indicates validity of the preprocessing step of events randomization process. The bottom line is that the hyperparameter optimization can provide significant model improvement and should be done if possible.

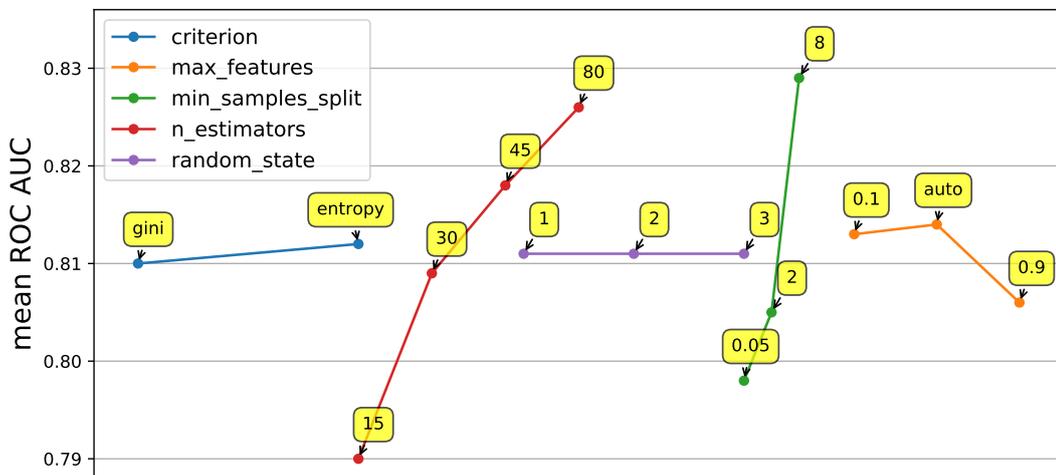


Figure 5.10: The impact of the hyperparameter optimization on the ROC AUC. For every hyperparameter, the points have been calculated by averaging the model ROC AUC values over all remaining hyperparameters.

RF with best selected hyperparameters is trained once again using full insample, and validated in the outsample. The confusion matrix for RF is presented in Fig. 5.11 and the ROCs with areas under curves values are presented in Fig. 5.12. The random forest model average ROC AUCs for all output classes is 0.849, with ZTT being the best discriminated background. The average ROC AUCs for four Higgs boson output classes is 0.853. The RF presents considerably better performance than use of sole decision tree.



Figure 5.11: Confusion matrix for RF model

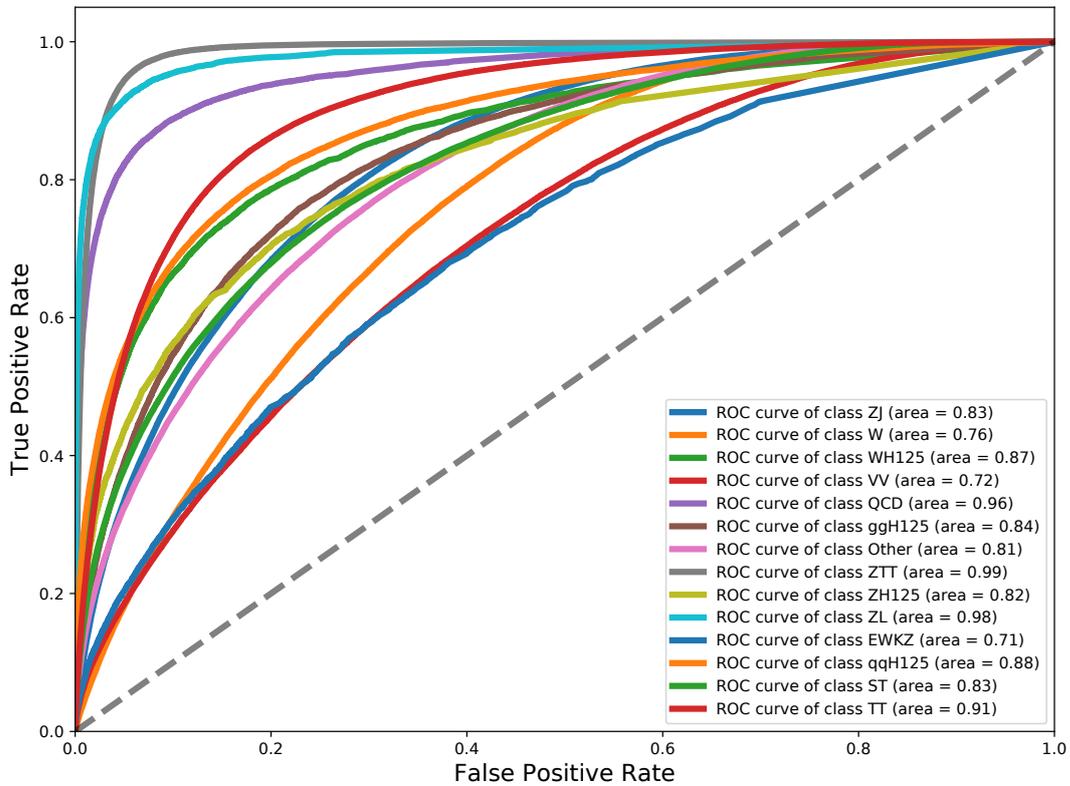


Figure 5.12: ROC curves for the RF model

5.4 Gradient boosted decision trees

Small decision tree classifiers are often considered to be "weak" predictor models, i.e. they are computationally simple and perform relatively poorly (only slightly better than random guessing) in complicated scenarios. The question of existence of arbitrary "strong" (i.e. accurate) learner built out of a set of weak learners²⁰ was first posed by M. Kearns [139] and L. Valiant [140] as *hypothesis boosting problem*²¹, and answered affirmatively in 1990 by R. Schapire [141]. Although these theoretical considerations can be in practice applied for an arbitrary (weak) estimator, it is decision trees that gained the most attraction in the field of boosting. The first famous boosting algorithm, developed by Y. Freund and again R. Schapire in 1996 [142], *AdaBoost* is traditionally implemented using *stumps*, i.e. one-level decision trees. AdaBoost was originally developed for binary problems, and its first extensions to multi-class problems required quite strong base learners and did not gain much attention [143]. The boosting algorithm we have used in the analysis is a generalization of AdaBoost called *gradient boosting algorithm*, which is widely used and well suited for multi-class classifications [144].

The gradient boosting algorithm is based on the idea of finding a function (classifier) f of input variables (features) $x \in X$ which minimizes a specified loss function $L(y, f(x))$ for a given set of output classes $y \in Y$. The employed loss function depends on the problem, and in our case of multi-class classification it is the negative multinomial log-likelihood. It can be shown that minimizing the negative log-likelihood loss function is equivalent to minimizing the categorical cross-entropy between the target (in training dataset) and prediction given by the classifier.

Gradient boosted decision trees (BDT) algorithm is composed of three parts. In the first part the classifier $f_k(x)$ is initialized with constant value (=zero) for all classes $k \in K$. The second part comprises of iterative refinement of the classifier. On every iteration step $m \in \{1, \dots, M\}$, the new decision tree is trained on targets given by the gradient of loss function evaluated at $f = f_{m-1}$. The worse the current prediction is for the event, the higher will the (absolute) value of target on the next iteration be. Please note that even though we search for the classifier, the trees fitted in this method are regression trees and not classification trees. A fitted decision tree is added to existing classifier at the end of every loop. In the third part of algorithm the final classifier is set as the one from the last iteration: $f_k = f_{kM}$. For more detailed description of the BDT please see e.g. [145].

The output of the classifier can be used to compute the class probabilities for an event with the following formula (called *softmax activation function*):

$$p_k(x) = \frac{\exp\{f_k(x)\}}{\sum_{l=1}^K \exp\{f_l(x)\}}. \quad (5.8)$$

The form of Eq. 5.8 ensures that $0 \leq p_k(x) \leq 1$ and that the sum of the probabilities over all classes equals one.

The boosting algorithms are susceptible to overfitting. Regularization techniques constraining the fitting procedure are introduced to reduce this undesired effect. The first obvious regularization method is to constrain the number of boosting stages M (`n_estimators`). The second method, which is sensitive to the mutual interactions of the input features, sets the maximal depth (`max_depth`) of the individual tree i.e. it

²⁰Gradient boosting is similar to the RF in the sense that the combined estimator is built in the form of an ensemble of base estimators.

²¹"Boosting" term refers to boosting the performance (accuracy) of the base estimator.

modifies the number of nodes. A third popular way of constraining overfitting is to modify the so called *learning rate* (`learning_rate`). Learning rate is a weight given to every new tree added while updating the classifier in the second part of the boosting algorithm. Finally, there are regularization methods which are inspired by RF algorithm and consist of limiting: 1) the minimal number of events required to split the node (`min_samples_split`); 2) the fraction of features available to be used in the split at the node (`max_features`); 3) the fraction of events to be used for training the trees (`subsample`). All above regularization techniques are introduced to the model at the stage of estimator initialization (before the training) via hyperparameters. We performed hyperparameter optimization to find the best model settings by training 324 different models. Results are summarized in Fig. 5.13.

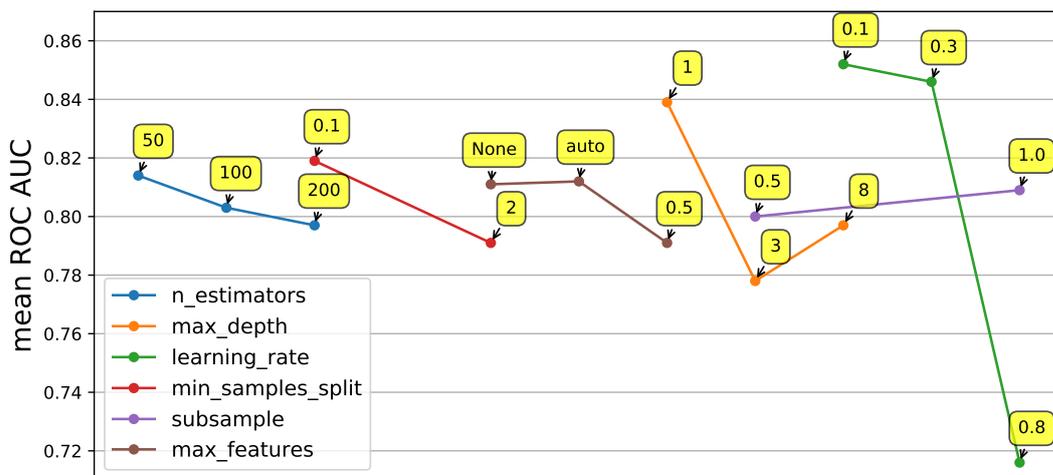


Figure 5.13: The impact of the hyperparameter optimization on the ROC AUC. For every hyperparameter the points have been calculated by averaging the model ROC AUC values over all remaining hyperparameters. The meaning of the hyperparameter names are given in the text. The *None* value for *max_features* means no limit for this variable (all features are used).

Overfitting is not the only downside of BDT. Other potential and significant negative aspects of this algorithm are 1) a high computational complexity of the algorithm and 2) that it still is a shallow estimator. The former means that the training takes a lot of processing time and sometimes enforces shrinking of the training dataset. The latter means that BDT has limited power to capture the abstract relations between features and requires building training variables deliberately (using a posteriori knowledge) in order to make it work with best performance. Regardless of those issues, BDT is one of the best out-of-the-box estimators in supervised learning scenarios, producing a combined classifier giving small errors on the training dataset, an overall low variance compared to base estimator and is fairly robust to overfitting owing to regularization.

The confusion matrix for BDT is presented in Fig. 5.14 and the ROCs and ROC AUCs values are in Fig. 5.15. The boosted decision tree average ROC AUC for all output classes is 0.860, with ZL being the best discriminated background. The average ROC AUC for Higgs boson output classes is 0.87. The BDT presents better performance than RF.



Figure 5.14: Confusion matrix for BDT model

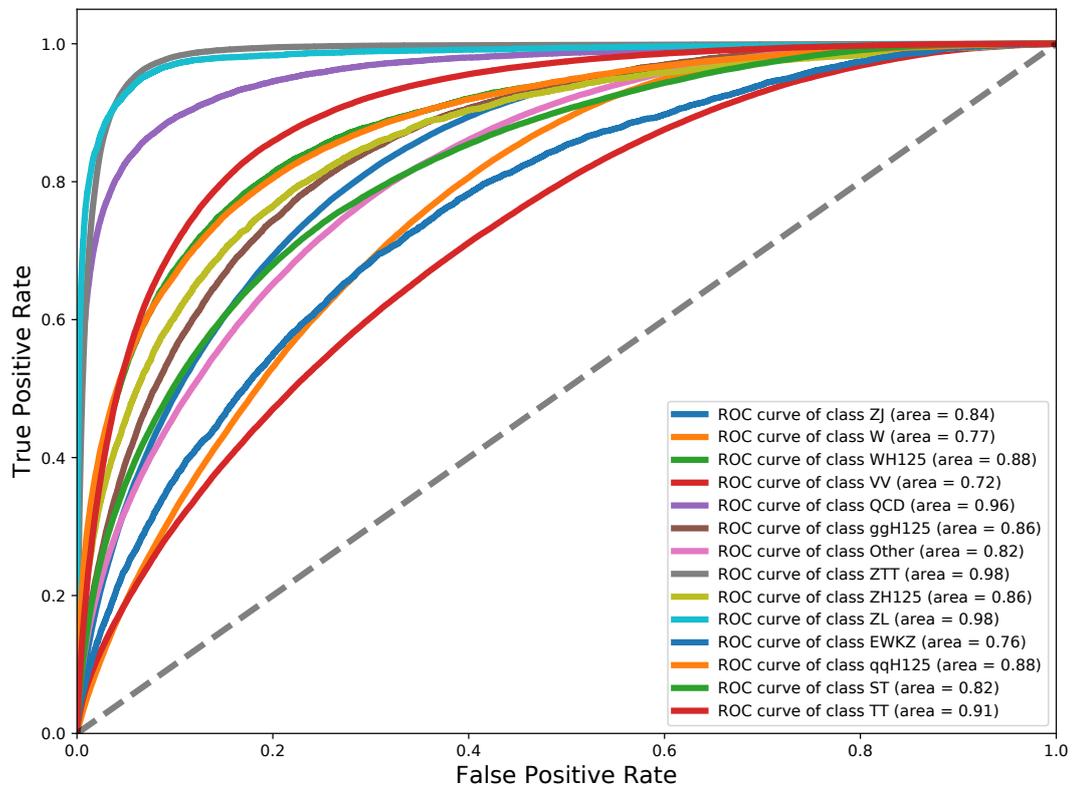


Figure 5.15: ROC curves for the BDT model

5.5 Deep learning

Deep learning (DL) is a name for machine learning techniques based on artificial neural networks²², i.e. computing systems inspired by biological neural networks. The exhaustive description of DL is a book long task. We will therefore concentrate only on methods implemented in this analysis, which are based on the PYTORCH library [146] and constrained by the hardware.²³

The basic building block of DL algorithms is a *neuron* (Fig. 5.16) – the generalization of invented already in 1957 *perceptron*. Similarly to the biological neuron, the artificial neuron (also called *node*) has dendrites, cell body and an axon. The dendrites represent incoming connections x_i , i.e. features values (in case of first neurons layer) or output from (connected) neurons of the previous layer. The signals moving along the dendrite are rescaled by the *weights* w_i , which are free parameters of the model adjusted during learning process. In cell body the signals from dendrites are summed up. Additionally, there can be a bias b (another free parameter) added to the computed sum, which ensures activation of the neuron²⁴ for low strength signals. The function

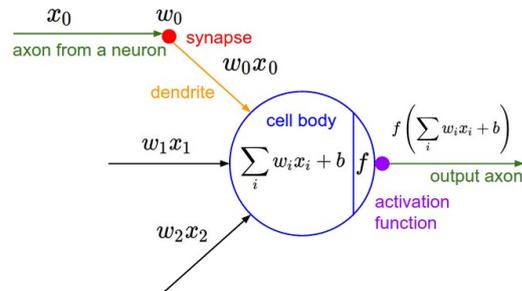


Figure 5.16: Neuron model [147]. The x_i denote the input variables, w_i are weights and b is the bias.

$$f = \sum_i w_i x_i + b,$$

where i goes from 1 to *number of dendrites* is called a *transfer function*. Finally, the value of the transfer function is passed over to the axon, where it is wrapped by the *activation function* and transmitted further to either the neuron(s) of the next layer or the output of the net. The activation function is used in order to introduce non-linearity to the neuron response – without it a single layer model would not be able to perform more than (multidimensional) linear regression. In practice the activation function performs filtering and rescaling of output from the neuron. Popular activation functions are Heaviside step function (in perceptron), sigmoid function (used historically before the third wave²⁵ of DL popularity) and Rectified Linear Units (ReLU) which is recently favored for various real-life problems. Typically a single layer of neurons share the same activation function type, which is determined by specific layer purpose.

The neural network is composed of multitude of neurons. One step above the neuron in the net abstraction hierarchy is a *layer*, which is simply a bunch of completely independent (not connected) neurons. The layers can be further stacked to

²²We do not want to be too strict here. For the purpose of this thesis let us assume that Deep Learning (DL) is a subset of ML based on Deep Neural Networks (DNNs) and DNN is simply Neural Network (NN) with at least three layers (including the input and output one).

²³We used single low-end gaming GPU with a 3 GB memory. This influenced the maximum number of neurons/layers and data batch size.

²⁴The neuron is *activated* when outputs nonzero value, otherwise is *dead*.

²⁵The AI history can be divided into three periods between its establishment in the 1950's and now, splitted by two *AI winters* in 1970's and 1985-90s.

create *multilayer neural network*. The way in which the neurons in layers are connected depends on the utilized network architecture, e.g. all neurons between layers are fully connected (in *feed-forward neural networks*), the neurons in a layer are connected only to a subset of previous layer neurons (in *convolutional neural networks* used primarily for image recognition) or neurons are connected creating directed graphs i.e. the neuron in next layer can feed back the neuron from previous layer (in *recurrent neural networks* i.e. networks with memory). In this analysis we implemented the multilayer feed-forward net, where we can distinguish three types of layers: input, hidden and output. Single *input layer* is fed by the features values and typically has the number of neurons equal to the number of features. The potentially many *hidden layers* are crucial to solve non-linear problems. The number of hidden layers and the number of neurons in every hidden layer are arbitrary and are the hyperparameters of the model. The last hidden layer is connected to the *output layer* for which the number of outputs is the same as the number of output classes (for classification). The output layer returns the measure of probability for the event to belong to a given output class and uses usually a sigmoid or softmax (see Eq. 5.8) activation function.

When the neural network architecture, number of (hidden) layers and the number of neurons and activation functions for every layer are set up, it is time to proceed with net training. The goal of this process is to find the values of weights and biases for which neural network gives the best estimates for output classes. Two questions need to be answered: "*How to measure goodness of the model estimates?*" and "*How to translate this measure to changes of weights and biases?*".

The goodness of the model estimates is measured with loss function. Similarly to the situation with BDT algorithm, the best form of the loss function depends on the neural network application. The loss function takes weights and biases as parameters and computes error of net prediction w.r.t. target. Finding the optimal parameters for arbitrary loss function would be NP-hard combinatorial optimization problem. To go around this problem the common requirement for loss functions is to be convex and smooth, which allows to compute its gradients estimates w.r.t. parameters. The task of finding the gradient of multivariate functions (with multiple local minima) is a science by itself and core fundamental concept of almost all DL methods [148]. The most popular solution here is to use *backpropagation* algorithm [149]. This algorithm invented in the 1960's implements the gradient descent of loss function in weight/bias space exploiting the chain rule. The free parameters of the model are updated iteratively and recursively going backwards from output layer to input layer. The backpropagation algorithm is fast and allows to distribute prediction error across all parameters of the net. Moreover, it is possible to ensure global convergence of backpropagation under a few assumptions on the loss function [150].

The class of algorithms called sometimes *optimizers* is designed to perform the translation of gradient of the loss function into parameters changes. Regardless of optimizer implementation, the parameters changes are in the direction of decreasing prediction error. In this analysis we have settled on the ADaptive Moment estimation (*Adam*) optimizer [151]. Let us denote the parameters and gradient of the loss function by θ and g , respectively. Then the Adam algorithm correction is given by the formula (arrow indicates the direction of updates):

$$\theta_t \leftarrow \theta_{t-1} - \left[\alpha \cdot \left(\sqrt{\frac{v_t}{1 - \beta_2^t}} + \epsilon \right)^{-1} \cdot \frac{m_t}{1 - \beta_1^t} \right], \quad (5.9)$$

where m_t is exponential moving average of the gradient (estimate of gradient first moment/mean)

$$m_t \leftarrow \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t,$$

and v_t is exponential moving average of the squared gradient (estimate of gradient second moment/uncentered variance)

$$v_t \leftarrow \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t^2.$$

The algorithm updates the values of gradient moments and models parameters iteratively until the stopping criterion is satisfied, i.e. θ_t will converge. The index t (acronym from *time*) is the iterator in the loop and if denoted as subscript indicates the number of iterations and if placed in superscript means the power. For $t = 0$ we set $m_t = \vec{0}$ and $v_t = \vec{0}$. In the formula above, β_1 and β_2 are the exponential decay rates for the moment estimates. They are also the hyperparameters of the model, set in range between zero and one. In Eq. 5.9 there are three parts in the main bracket:

- The learning rate or stepsize α . The role of this hyperparameter is similar to the one in BDT algorithm, which is to modulate the learning speed. This regularizes the model by helping avoid overtraining on noisy data. It is a challenging task to choose the right value for this important variable, as it may vary depending on model, dataset or even layer and epoch number.
- The middle part of the bracket introduces the adaptive learning rate. The term in the denominator under the square root is used to counteract the bias toward zero value of (initialized by the zero vector) second moment. The small constant ϵ prevents the divisions by zero in the computation.
- The last part is simply the bias-corrected gradient first momentum estimate.

The above algorithm hides yet another important hyperparameter. For the loss function $L(f(x; \theta), y)$, where f is the model, x is the features vector and y is the target, the gradient estimate computation (processed by backpropagation algorithm) can be written as:

$$g \leftarrow \frac{1}{m} \nabla_{\theta} \left(\sum_{i=1}^m L(f(x^i; \theta), y^i) \right),$$

where the summation is over the subset of events in the training sample. The size of this subset (i.e. value of m) is called *batch size*. With batch size hyperparameter it is possible to improve the performance and decrease the hardware memory usage of the algorithm without losing the generality of results. This in turn allows to design bigger nets using limited computing resources.

The training can be repeated multiple times using *training epoch number* hyperparameter. Each training epoch is a full Adam algorithm passed over the whole training dataset.

The DL models are sensitive to features initial normalization. There are at least three reasons for that. First, the normalization affects the activation of the neurons. If some input values are too small the neurons can become dead for those features. Conversely, the large input variable range can end up in too large neuron activations. It is valuable to match the features normalization to the range of the activation function, e.g. 0 to 1 for sigmoid, ReLU and softmax or -1 to 1 for hyperbolic tangent. Second, some optimizers can assume the normalized features to make initial weights/biases and/or internal algorithm parameters values sensible. Third, the

gradient descent used in backpropagation algorithm may not be able to find the best gradient direction and converges much slower if the features are not normalized.

The DL methods became very popular recently; arguably they are the most popular ML techniques nowadays. They are fast to learn and execute (also due to GPU-optimization), learn well on huge datasets and have a long history of usage and development. They are also "...the second best way to solve any problem²⁶...". There are however some drawbacks of DL, e.g.: 1) they are black-box models, hence do not give an easy way to obtain human-readable insight into the problem; 2) they are non-deterministic and sensitive to the initial parameters. Therefore, in our case the grid-search based hyperparameter optimization has no real value; 3) they are not probabilistic, so it is difficult to estimate the prediction (error) for the model (classification).

In the analysis we trained the model using four hidden layers with the first three layers made of 600 and the fourth layer of 300 neurons, and with ReLU activation functions. We used a constant learning rate of $\alpha = 0.01$, adaptive batch size (varying from $m = 2000$ to $m = 50000$ events depending on epoch) and 1000 training epochs. The input and output layers have sigmoid activation functions and the score has been translated into probability measures using softmax function. The above architecture was chosen by author empirically as the number of parameters is too large for DL to make use of grid search practical, especially when parallel processing is not possible due to only one GPU available.

The confusion matrix for DL is presented in Fig. 5.17 and the ROCs and ROC AUCs values are shown in Fig. 5.18. The deep learning model average ROC AUC for all output classes is 0.875, with ZTT being the best discriminated background. The average ROC AUC for Higgs boson output classes is 0.885. The DL is overall slightly better than BDT.

Summary

In this chapter four machine learning models have been developed: decision tree (ROC AUC of 0.611), random forest (ROC AUC of 0.849), BDT (ROC AUC of 0.860) and neural network (ROC AUC of 0.875). The output classes discrimination turned out to be qualitatively similar in all four cases. The ZTT, ZL, QCD and TT classes have always been the best distinguished. On the other hand, the events of VV and EWKZ processes have been poorly recognized. The models behave similarly also with respect to the relative discrimination power between different output classes. Single top and ditop are always difficult to differentiate, and the same goes for ZJ and W classes.

One of the goals in this chapter has been to present the overview of available machine learning methods and their utilization in the HEP field. As such we tried to devote a similar amount of time for developing every method. There are two conclusions. First one is stated above: the models, although intrinsically different, return quantitatively similar results. Second conclusion is that currently neural network (DL) presents the best performance.

DL can be relatively quickly implemented, trains fast with help of GPU and returns best results. RF and especially DT perform poorly compared to DL. BDT is - depending on settings - either very bad in classification or very slow in training. Moreover, DL has the largest margin for improvement, e.g. with usage of stronger hardware. However, the DL method requires a quite deep understanding in order

²⁶"The best way is to actually understand the problem" [152].

to give good results - a simple hyperparameter optimization using grid search has not worked well in our case.



Figure 5.17: Confusion matrix for DL model

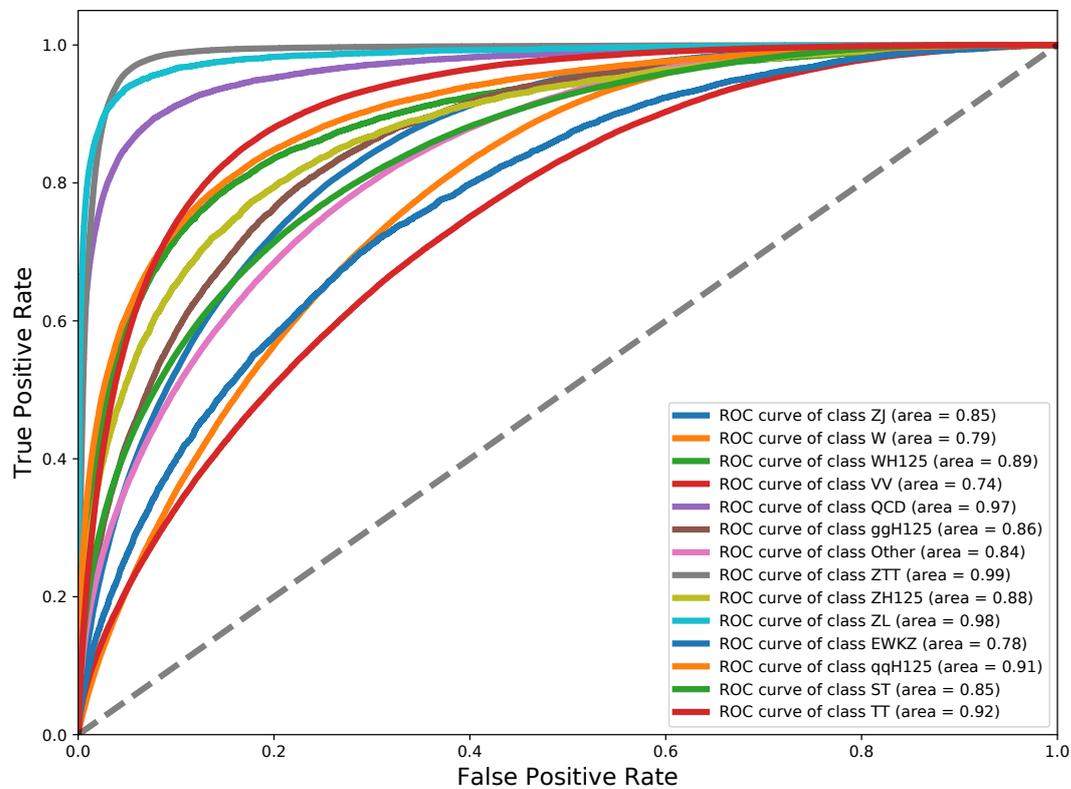


Figure 5.18: ROC curves for the DL model

Chapter 6

Machine Learning application into the $H \rightarrow \tau\tau$ analysis

In the CMS $H \rightarrow \tau\tau$ publication [153] (described in details in Chap. 4), the excess of signal 125 GeV Higgs boson events is given quantitatively in the form of local background-only p -values, i.e. the probabilities of background fluctuations to deliver an excess of events at least as large as observed [154]. In this process the 2D distributions of the discriminating variables are used as an input for the global maximum likelihood fit. The nuisance parameters with appropriate probability density function (Gaussian or log-normal) are introduced into the fit in order to account for the systematic uncertainties. Finally, the obtained p -values are translated into corresponding observed significance S .

The described procedure requires certain amount of additional data preprocessing as it is an standardized procedure implemented within the Higgs PAG combine-tool *Combined Limit* [155]. The data samples preparation is a time-demanding task on its own, needed in order to obtain limits, significances and likelihood scans for final data-driven analysis. However, since we are only interested in significance ratio between the methods a simplified approach for the significance calculations is sufficient. It is described in Appendix E. In particular all operations will be done on the nominal distributions with fixed signal and background yield.

There are multiple possible ways to bring machine learning into physics analysis. Here we propose to use the output of the ML algorithm as an standalone discriminating variable¹. The procedure is applied on top of the categorization used in the CMS analysis (i.e. only for events passing the 0jet, boosted or vbf category cuts) to ensure a fair comparison.

Our goal in this chapter is to calculate the significance for cut based analysis using the nominal distributions of discriminating variables (original plots are presented in Appendix C), where the Higgs signal events and background will be simulated with MC samples. Furthermore, we will compare it to the significance computed for the proposed machine learning based approach.

6.1 Setup

To proceed with the computations, the steps from Chapter 4 have been repeated, with several modifications, separately for insample and outsample. The following changes have been made with respect to the original analysis:

¹We have also tested the approach with ML score playing the role of additional variable to be used together with Higgs mass estimate variable, i.e be swapped with the second variable on the 2D signal variable distributions. The approach presented in this chapter turned out to be, however, both simpler and more efficient.

- The background has not been split into subcategories (eg. "Electroweak", "QCD", etc). As we are interested only with signal/background yield ratio this change does not affect the results;
- Real data is not used, thus only expected sensitivity is evaluated;
- The W + jets yield is taken from MC, i.e. the procedure of distribution "smoothing" has not been applied;
- QCD has been taken from MC and not from the data.

The third and fourth points are necessary to hold the definition of outsample consistent throughout the thesis, and have a negligible influence on the final conclusions. The signal yield is taken from the MC as a sum of the events from samples $ggH125$, $qqH125$, $ZH125$ and $WH125$ that have passed considered category selections.

For all the events left after categorization, the DL from Section 5.5 has been used to generate the event-by-event scores. In the original analysis we are interested only in the Higgs signal and not in determining a specific event origin. Therefore, for every event the fourteen-element score array needs to be cast into the measure of the probability for the event to belong only to the Higgs category. Let us create two classes: Higgs (\mathcal{S}) and background (\mathcal{B}). Further, the score for the class Higgs is defined – for every event – as:

$$\text{score}(\mathcal{S}) = \sum_i \text{score}(i), \quad (6.1)$$

where i goes over $ggH125$, $qqH125$, $ZH125$ and $WH125$ categories. Obviously, the relation $\text{score}(\mathcal{B})=1-\text{score}(\mathcal{S})$ always holds.

Model sensitivity evaluation

Sensitivity of a given (ML- or cut-based) approach is evaluated by measuring the significance S (see Appendix E):

$$S = \sqrt{2 \log Q} = \sqrt{2 \left\{ \sum_{\text{bins}} -s_i + (s_i + b_i) \log \left[1 + \frac{s_i}{b_i} \right] \right\}}, \quad (6.2)$$

where the sum goes over all bins in 2D distributions (number of bins is equal to number of bins of the first variable times the number of bins of the second variable) for cut based analysis and over all score bins for ML based analysis. The results are computed separately for all analysis categories.

6.2 An insight into machine learning approach

The basic idea behind the Higgs boson searches is related to the *discriminating variables*, i.e. variables with some power to distinguish between the signal and background events. Standard $H \rightarrow \tau_\mu \tau_h$ analysis relies on ditau system mass and transverse momentum, dijet system mass and to some extent on hadronic tau decay mode. All of them are somehow sensitive to the physical process taking place in the event (see Chapter 4). Since DL model uses input features containing all aforementioned variables, we are using only its output (score in Eq. 6.1) as the sole discriminating variable.

The ideal (fake) discriminating variable (e.g. top plot in Fig. 6.1) would allow to always determine the origin of the event and therefore be used to discard (extract) all background (signal) events. Real DL score is not ideal - its distribution for signal and background for MC samples is shown on the bottom plot in Fig. 6.1.

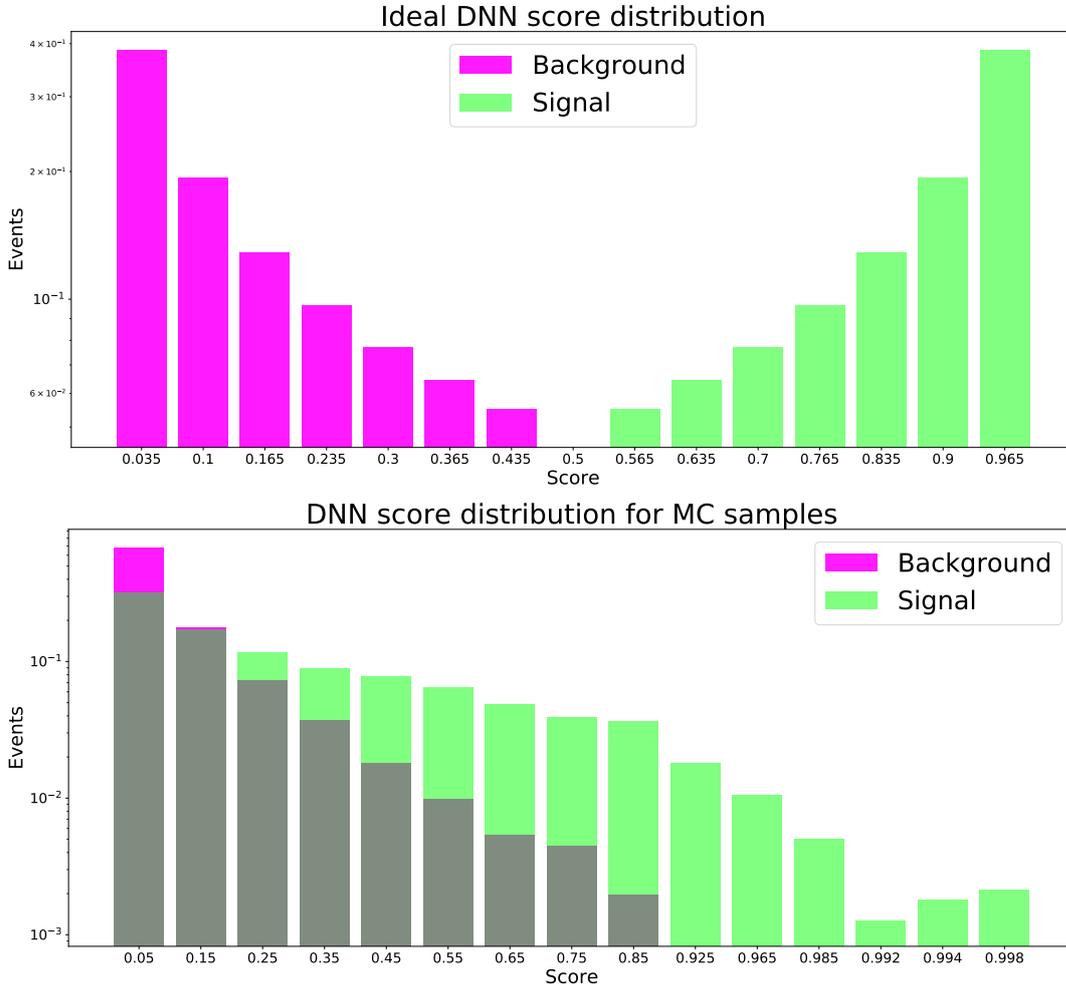


Figure 6.1: Scores distributions, normalized to unity separately for signal and background. *Top* plot presents the "ideal" (fake) distribution where signal and background can be perfectly separated using only two score bins. *Bottom* plot presents the real distribution for MC samples used. Please note the uneven bin ranges.

For every distribution of discriminating variable we are able to compute the number of events of signal and background in every bin and further obtain the significance using the formula in Eq. 6.2. Two contradictory effects occurs. On the one hand, in order to get a large significance we want to have as many as possible (to build up the sum part) terms with the highest possible signal-to-background ratio in Eq. 6.2. In other words, more bins in the histogram usually mean a larger significance. On the other hand, by slicing finely the distribution we soon reach the point where only a small number of events will be present in every bin. This in turn will cause significant fluctuations in events number (per bin) and a large systematic uncertainty of the final result. The trade-off between the significance value and its uncertainty may be an issue in an analysis and in our case has been worked out by looking at the cut based case solution. More specifically, we modeled binning of our

score distributions to have the same number of bins and with the number of events not significantly smaller (see next page) than the worst case in cut based analysis.

From the bottom plot in Fig. 6.1 one can see that the bin splitting with constant bin widths is not optimal. There are at least two reasons: 1) splitting bins with low-score values does not bring much significance improvement due to large background and 2) the high-score bins may contain zero or negligible background yield boosting final significance to very large values but also with unacceptably large uncertainty. We decided to derive the proper bin splitting using a two-step process:

- Primary bin splitting by optimization. In this step we have decided to take advantage of scalar function minimization methods and optimize the (negative of) significance as a function of bins splitting. The following assumptions have been adopted:
 1. The optimization has been performed on the insample to obtain unbiased results;
 2. The number of score bins is the same as the number of bins of second variable (i.e. decay mode, p_T^H or m_{jj}) in the cut based analysis distributions (three for 0jet, six for boosted and four for vbf categories). We have not used the total number of bins in 2D distributions as the optimization does not work well for large number of degrees of freedom and the results are worse than when using additional quantile-based splitting².
 3. The edges of bins are given with precision of three significant digits;
 4. The minimal, expected, normalized background event yield and the number of used generator MC events number in a single bin are required to exceed those values for the cut based method for a given category³. To ensure the result will obey this requirement also on outsample we raised the actual requirement values for boosted and vbf categories (to 1 for yield and 5 for number of events);
 5. The optimization has been performed with the Nelder-Mead simplex algorithm [156] – the most widely used algorithm for unconstrained optimization problem [157].

To increase the likelihood for a found local minimum to be a global minimum, the stochastic Basin-Hopping algorithm [158] has been used on top of Nelder-Mead, with five hundred algorithm iterations and with the initial step size for use in the random displacement set up as an inverse of the number of bins. The results are presented in Table 6.1.

Category	Bins edges
0jet	[0, 0.326, 0.53, 1]
boosted	[0, 0.05, 0.242, 0.41, 0.676, 0.966, 1]
vbf	[0, 0.215, 0.548, 0.949, 1]

Table 6.1: The best splitting per category found with insample significance optimization.

²This is true with the assumption of using reasonable number of iterations in Basin-Hopping algorithm (see below). We set up cutoff on optimization algorithm execution time for 2 hours.

³In we denote the yield by y and number of events by n then we have: $y_{\min, \text{bkg}}^{\text{0jet, cut}} \simeq 17$, $n_{\min, \text{bkg}}^{\text{0jet, cut}} = 22$, $y_{\min, \text{bkg}}^{\text{boosted, cut}} \simeq 0.0009$, $n_{\min, \text{bkg}}^{\text{boosted, cut}} = 1$ and $y_{\min, \text{bkg}}^{\text{vbf, cut}} \simeq 0.007$, $n_{\min, \text{bkg}}^{\text{vbf, cut}} = 1$.

- Secondary bin splitting using background events quantile. In this step we are further subdividing the bins from the primary bin splitting to get the total bin number equal to the number of bins in cut based analysis (separately for every category). The main objective here is to avoid introduction of large systematic uncertainties by creating bins with a very low background yield/events number. Since - by construction - the number of background events decreases with the score value, we do not split the last bin. The other bins are divided by hand, taking care of keeping assumption no. 4 to always hold, and using splitting by background quantile⁴. The results are presented in Table 6.2.

Category	Number of sub-bins	Total no. of bins
0jet	[21, 14, 1]	36
boosted	[25, 20, 2, 2, 10, 1]	60
vbf	[10, 5, 4, 1]	20

Table 6.2: The number of sub-bins in quantile splitting by background events and the total number of bins per category.

The number of events (signal and background) per bin of discriminating variable distribution - score for ML based method and decay mode/visible mass (0jet cat.), p_T^H /SVFit mass (boosted cat.) or m_{jj} /SVFit mass (vbf cat.) for cut based method - are plotted in Fig. 6.2.

6.3 Systematic uncertainties

In order to measure the sensitivity of the results to the systematic uncertainties we are following the method from the original publication [82] and rescale the values of the selected variables (up and down). The variables and the scale factors are inspired by Table 4.8, from which we took four most important systematic uncertainties in the $H \rightarrow \tau_\mu \tau_h$ decay channel. The following sources of systematic uncertainties are taken into account:

- Hadronic tau energy scale. We rescale the transverse momentum of the taus matched to generator hadronic tau (cat. no. 5 in Table 4.2) by $\pm 1.2\%$. After this change the SVFit mass, Higgs boson transverse mass and momentum, hadronic tau transverse mass, lepton pair transverse momentum and visible mass have to be recalculated.
- Muon misidentified as hadronic tau energy scale. We rescale the transverse momentum of the taus matched to generator lepton (cat. no. 1-4 in Table 4.2) by $\pm 1.5\%$. The other variables have been recalculated like in the previous point.
- MET transverse momentum energy scale. We rescale MET coordinates (\vec{E}_x , \vec{E}_y) all together by $\pm 3\%$. Other dependent and recalculated variables are: SVfit mass, lepton transverse masses and Higgs boson transverse mass and momentum.
- Jet energy scale. All jets in the event get rescaled transverse momenta depending upon their original p_T and η . We have used the standard formula:

⁴We have used event-wise quantile splitting, i.e. number of events in sub-bins are equal. We could use yield-wise quantile splitting just as well.

$p_T^{\text{new}} = p_T^{\text{old}} \times (1 \pm \phi(p_T, \eta))$, where $\phi(p_T, \eta)$ is recommended correction function (values) used in reference analysis. JES does not enforce SVFit mass recalculation.

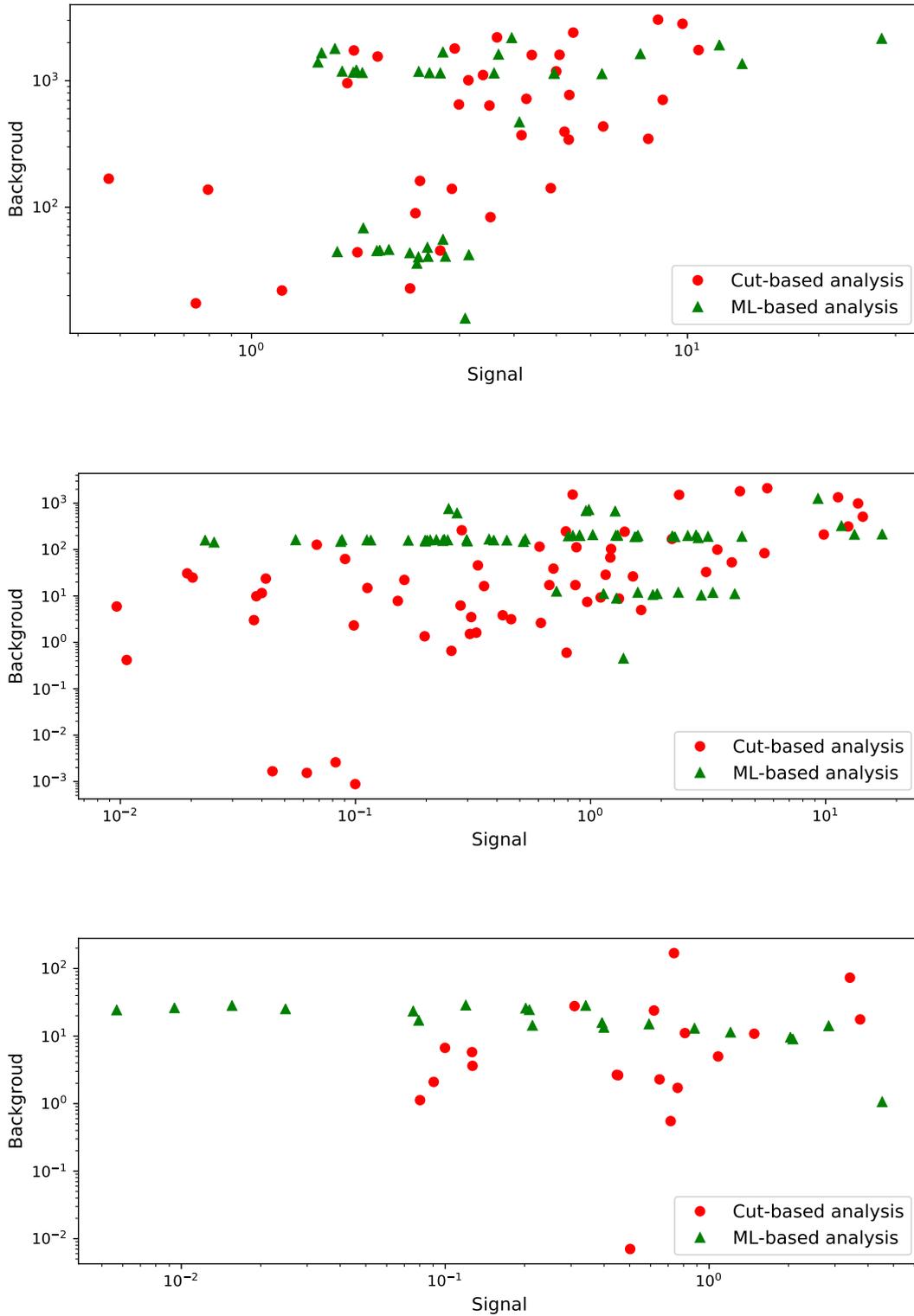


Figure 6.2: Plots of signal versus background yield. Every point represents a single bin of discriminating variable distribution. The values for cut-based (red circles) and ML-based (green triangles) analysis are given together. Three separate plot cover three categories: 0jet (top), boosted (middle) and vbf (bottom).

Different rescales may change discriminating variables distributions in different ways. Hadronic tau energy scale influences the overall yield by moving events outside or inside signal region due to baseline selecton $p_{T,\tau_h} > 30$ cut (see p. 50) and changes the shape of SVFit mass (is one of its input parameters). MET transverse momentum energy scale does the same via $M_T^\mu > 50$ cut in categorization (see Section 4.3). JES may change which category the events are assigned to (changes the number of jets (njets) by making jet not pass the $p_T > 30$ cut - see p. 46) and to which bin in vbf category (changes dijet system mass (m_{jj}) value). Rescales may also influence the classification results given by the neural network (see Section 5.5).

6.4 Results

The values of points from Fig. 6.2 (and similar distributions rescaled according to the previous section) have been used to compute the significance from Eq. 6.2. The results are:

- in 0jet category the significance for cut based analysis is $1.389^{+0.1\%}_{-3.7\%}$ and for ML based analysis is $1.751^{+3.1\%}_{-3.3\%}$. The neural network allows to obtain about a 26% better result with a comparable systematic uncertainty.
- in boosted category the significance for cut based analysis is $2.640^{+14\%}_{-9.9\%}$ and for ML based analysis is $3.171^{+0.3\%}_{-5.9\%}$. The neural network allows to obtain about a 20% better result with even a smaller systematic uncertainty.
- in vbf category the significance for cut based analysis is $2.461^{+41\%}_{-2.2\%}$ and for ML based analysis is $3.340^{+2.4\%}_{-38\%}$. The neural network allows to obtain nominally about a 35% better result, however, the systematic uncertainties are large here, for both cut- and ML-based analyses. The best case for cut based analysis is slightly better than nominal value for ML method and the worst case for ML analysis is slightly worse than nominal cut based case.

The uncertainties in above figures are the result of rescaling the variables presented in previous section, i.e. come from systematic effects. The numerical results for the significance analysis are summarized in Table 6.3. The final significance number is the superposition of two things: 1) the discriminating power of the considered discriminating variable with respect to the signal events and 2) the stability of its value when varying inputs conditions. The former allows to construct distributions with high signal-to-background bins, and the latter to practically use them without worrying about exploding systematic uncertainties. Presented results show that developed machine learning methods have comparable stability to the cut based approach and present even a better signal discriminating potential. Hence, they can be thought as a simple extension (or replacement) to the standard cut-based analysis categorization.

	0jet cut based	boosted cut based	vbf cut based
NOMINAL VALUE	1.389	2.640	2.461
hadronic tau energy scale	+0.0 -0.05	+0.241 -0.159	+0.079 -0.0
muon misidentified as hadronic tau energy scale	+0.006 -0.013	+0.016 -0.064	+0.003 -0.003
MET transverse momentum energy scale	+0.0 -0.0	+0.202 -0.053	+1.026 -0.0
jet energy scale	+0.027 -0.0	+0.192 -0.189	+0.054 -0.054
RESULT (cut based analysis)	1.389 ^{+0.1%} -3.7%	2.640 ^{+14%} -9.9%	2.461 ^{+41%} -2.2%
	0jet ML based	boosted ML based	vbf ML based
NOMINAL VALUE	1.751	3.171	3.340
hadronic tau energy scale	+0.015 -0.049	+0.0 -0.116	+0.08 -0.421
muon misidentified as hadronic tau energy scale	+0.024 -0.0	+0.0 -0.018	+0.003 -0.003
MET transverse momentum energy scale	+0.029 -0.029	+0.01 -0.0	+0.0 -0.796
jet energy scale	+0.043 -0.0	+0.0 -0.145	+0.0 -0.887
RESULT (ML based analysis)	1.751 ^{+3.1%} -3.3%	3.171 ^{+0.3%} -5.9%	3.340 ^{+2.4%} -38%

Table 6.3: Significance for cut based and ML based analysis together with the impact of various rescales of tau/MET/jet energy scales on the result. Values for every category (*0jet*, *boosted*, *vbf*) are given in separate columns.

Conclusions

The aim of this thesis has been to apply the latest developments in scientific study of algorithms and statistical models, commonly called machine learning, on the working re-implementation of the 2016 CMS $H \rightarrow \tau\tau \rightarrow \mu\nu_\mu\nu_\tau\tau_h\nu_\tau$ analysis. All relevant details of the analysis have been presented. It has been shown that it is possible to obtain agreement at the level of a few percent with official nominal CMS distributions of discriminating variables. Furthermore, we have explained why and how the machine learning methods could be applied to the event identification and how it could be translated onto a ground of Higgs decays analysis. While doing that we have prepared four machine learning estimators (decision tree, random forest, boosted decision tree and neural network) and have shown that neural network works best in considered scenario. The obtained performance given in terms of ROC AUC is equal to 0.875.

The machine learning classifier can work as an extension for categorization procedure employed in official CMS $H \rightarrow \tau\tau$ analysis, which is based on a carefully chosen set of kinematical cuts. We propose a new method in which the two dimensional distributions of discriminating variables are replaced with only output of the deep neural network. We then compare these two approaches by computing the significance. Machine learning has turned out to be considerably better improving the result by 20% to 35% (depending on the category) without increasing systematic uncertainties with respect to the original method.

The techniques proposed in this thesis can be used for example to improve the sensitivity to the signal events or to better determine the background contributions in control regions used in the analysis. They could further be used to obtain better Higgs properties measurements. One of the most important physical results that could be enhanced is related to the Higgs boson signal strength modifiers μ_i , which are defined as the ratio of the particular process branching ratio BR^f to the SM prediction $(\text{BR}^f)_{\text{SM}}$:

$$\mu^f = \frac{\text{BR}^f}{(\text{BR}^f)_{\text{SM}}},$$

where $f = ZZ, WW, \gamma\gamma, \tau\tau, bb$. The combined CMS result of the signal strengths for 35.9 fb^{-1} at 13 TeV are depicted on the left-hand side in Fig. 6.3. From the plot one can see that the uncertainties in the ditau Higgs decay mode channel are bigger than all of the other channels, with the exception of the Higgs decay into bb . The ML techniques could improve the result by constraining the uncertainties and effectively rule out some of the beyond SM models.

The other measurement that can benefit from utilization of ML techniques is related to the coupling modifier κ_j :

$$\kappa_j^2 = \frac{\sigma_j}{\sigma_j^{\text{SM}}} \quad \text{or} \quad \kappa_j^2 = \frac{\Gamma^j}{\Gamma_{\text{SM}}^j}$$

σ_j being cross-sections and Γ^j branching ratios (super-/subscript indicates SM values). The κ_j can be conveniently used to measure the deviation of the observed number of Higgs events from the SM prediction. The deviation is calculated in terms of the parameters (M, ε) , which relate to the coupling modifiers by the $\kappa_F = vm_f^\varepsilon / M^{1+\varepsilon}$ (for fermions) and $\kappa_V = vm_V^{2\varepsilon} / M^{1+2\varepsilon}$ (for vector boson) formulas. The result of such phenomenological fit is presented on right-hand side of Fig. 6.3. The better coupling modifier determination obtained with ML techniques would translate into smaller M and ε errors.

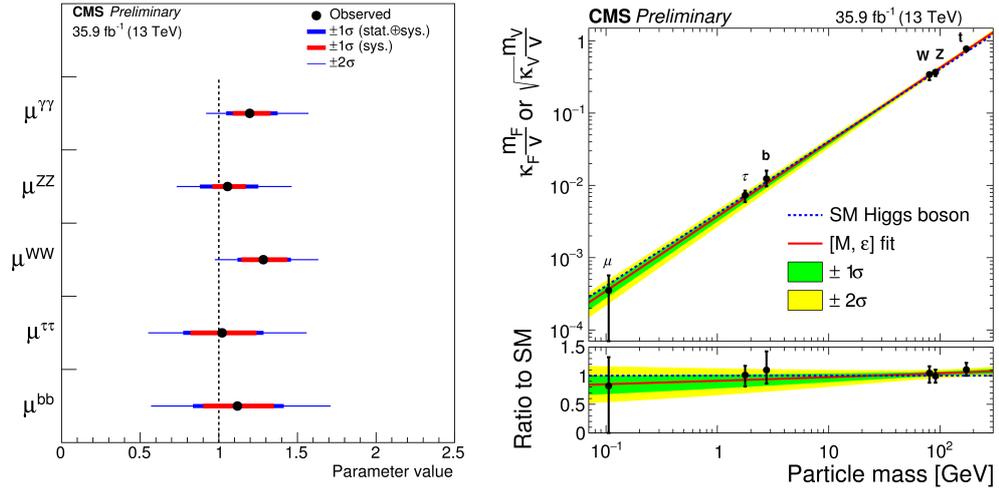


Figure 6.3: Left: The Higgs decay signal strength modifiers μ . The thick and thin blue horizontal bars indicate the $\pm 1\sigma$ and $\pm 2\sigma$ uncertainties, respectively. The red line shows $\pm 1\sigma$ systematic components of the uncertainties. Right: The fit of the six parameter κ model plotted versus the particle mass [159].

The ML-based analysis of $H \rightarrow \tau\tau$ decay channel is already presented by the CMS Collaboration in a public analysis note (PAS-only technical report) [160]. In the publication the neural network score is used as a discriminating variable and the methods which are used are to some extent similar to the ones presented in this thesis for DL. For example the DNN used also has two hidden layers and softmax function at the output layer. On the other hand, this analysis contains simplified input data sets (by requiring isolated hadronic taus and leptons, which effectively reduces background contamination), smaller number of input features (18 vs 68 for $\tau_\mu \tau_h$ final state) and output classes (8 vs 14), and a larger number of categories in classification scheme (12 vs 3). The analyzed 2016 and 2017 collision data corresponding to 77.4 fb^{-1} of integrated luminosity allowed to obtain the total inclusive cross section for $H \rightarrow \tau\tau$ process equal to $\sigma = 2.56 \pm 0.48(\text{stat}) \pm 0.34(\text{syst})$. The future official (not preliminary) CMS Collaboration $H \rightarrow \tau\tau$ public analysis based on the neural networks is currently in preparation.

Bibliography

- [1] Mark Srednicki. *Quantum Field Theory*. Cambridge University Press, 2007. ISBN: 9781139462761. URL: <https://books.google.co.in/books?id=50epxIG42B4C>.
- [2] E. Noether. “Invariante Variationsprobleme”. ger. In: *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse* 1918 (1918), pp. 235–257. URL: <http://eudml.org/doc/59024>.
- [3] Sheldon L. Glashow. “The renormalizability of vector meson interactions”. In: *Nuclear Physics* 10 (1959), pp. 107–117. ISSN: 0029-5582. DOI: [https://doi.org/10.1016/0029-5582\(59\)90196-8](https://doi.org/10.1016/0029-5582(59)90196-8). URL: <http://www.sciencedirect.com/science/article/pii/0029558259901968>.
- [4] Abdus Salam and J. C. Ward. “Weak and electromagnetic interactions”. In: *Il Nuovo Cimento (1955-1965)* 11.4 (1959), pp. 568–577. ISSN: 1827-6121. DOI: [10.1007/BF02726525](https://doi.org/10.1007/BF02726525). URL: <https://doi.org/10.1007/BF02726525>.
- [5] S. Weinberg. “Model of leptons.” In: *Phys. Rev. Lett.*, 19: 1264–6 (Jan. 1967).
- [6] T. P. Cheng and L. F. Li. *Gauge theory of elementary particle physics*. 1984. ISBN: 9780198519614.
- [7] Ziro Maki, Masami Nakagawa, and Shoichi Sakata. “Remarks on the unified model of elementary particles”. In: *Prog. Theor. Phys.* 28 (1962), pp. 870–880. DOI: [10.1143/PTP.28.870](https://doi.org/10.1143/PTP.28.870).
- [8] T2K Collaboration. “Search for CP Violation in Neutrino and Antineutrino Oscillations by the T2K Experiment.” In: *121.17* (2018), p. 171802. DOI: [10.1103/PhysRevLett.121.171802](https://doi.org/10.1103/PhysRevLett.121.171802). arXiv: [1807.07891](https://arxiv.org/abs/1807.07891).
- [9] P. W. Anderson. “Plasmons, Gauge Invariance, and Mass”. In: *Phys. Rev.* 130 (1 1963), pp. 439–442. DOI: [10.1103/PhysRev.130.439](https://doi.org/10.1103/PhysRev.130.439). URL: <https://link.aps.org/doi/10.1103/PhysRev.130.439>.
- [10] F. Englert and R. Brout. “Broken Symmetry and the Mass of Gauge Vector Mesons”. In: *Phys. Rev. Lett.* 13 (9 1964), pp. 321–323. DOI: [10.1103/PhysRevLett.13.321](https://doi.org/10.1103/PhysRevLett.13.321). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.321>.
- [11] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. “Global Conservation Laws and Massless Particles”. In: *Phys. Rev. Lett.* 13 (20 1964), pp. 585–587. DOI: [10.1103/PhysRevLett.13.585](https://doi.org/10.1103/PhysRevLett.13.585). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.585>.
- [12] Peter W. Higgs. “Broken Symmetries and the Masses of Gauge Bosons”. In: *Phys. Rev. Lett.* 13 (16 1964), pp. 508–509. DOI: [10.1103/PhysRevLett.13.508](https://doi.org/10.1103/PhysRevLett.13.508). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.508>.
- [13] J. Goldstone. “Field Theories with Superconductor Solutions”. In: *Nuovo Cim.* 19 (1961), pp. 154–164. DOI: [10.1007/BF02812722](https://doi.org/10.1007/BF02812722).

- [14] CERN. *LHC Higgs Cross Section Working Group*. 2019. URL: <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHXSWG>.
- [15] Nicolas Deutschmann et al. “Gluon-fusion Higgs production in the Standard Model Effective Field Theory”. In: *Journal of High Energy Physics* 2017.12, 63 (2017), p. 63. DOI: [10.1007/JHEP12\(2017\)063](https://doi.org/10.1007/JHEP12(2017)063). arXiv: [1708.00460](https://arxiv.org/abs/1708.00460) [hep-ph].
- [16] Charalampos Anastasiou, Radja Boughezal, and Frank Petriello. “Mixed QCD-electroweak corrections to Higgs boson production in gluon fusion”. In: *Journal of High Energy Physics* 2009.4, 003 (2009), p. 003. DOI: [10.1088/1126-6708/2009/04/003](https://doi.org/10.1088/1126-6708/2009/04/003). arXiv: [0811.3458](https://arxiv.org/abs/0811.3458) [hep-ph].
- [17] Jonathan M. Butterworth et al. “Jet Substructure as a New Higgs-Search Channel at the Large Hadron Collider”. In: 100.24, 242001 (2008), p. 242001. DOI: [PhysRevLett.100.242001](https://doi.org/10.1103/PhysRevLett.100.242001). arXiv: [0802.2470](https://arxiv.org/abs/0802.2470) [hep-ph].
- [18] Robert V. Harlander and William B. Kilgore. “Next-to-Next-to-Leading Order Higgs Production at Hadron Colliders”. In: 88.20, 201801 (2002), p. 201801. DOI: [10.1103/PhysRevLett.88.201801](https://doi.org/10.1103/PhysRevLett.88.201801). arXiv: [hep-ph/0201206](https://arxiv.org/abs/hep-ph/0201206) [hep-ph].
- [19] Oliver Brein, Robert V. Harlander, and Tom J. E. Zirke. “vh@nlo-Higgs Strahlung at hadron colliders”. In: *Computer Physics Communications* 184.3 (2013), pp. 998–1003. DOI: [10.1016/j.cpc.2012.11.002](https://doi.org/10.1016/j.cpc.2012.11.002). arXiv: [1210.5347](https://arxiv.org/abs/1210.5347) [hep-ph].
- [20] M. L. Ciccolini, S. Dittmaier, and M. Krämer. “Electroweak radiative corrections to associated WH and ZH production at hadron colliders”. In: 68.7, 073003 (2003), p. 073003. DOI: [10.1103/PhysRevD.68.073003](https://doi.org/10.1103/PhysRevD.68.073003). arXiv: [hep-ph/0306234](https://arxiv.org/abs/hep-ph/0306234) [hep-ph].
- [21] ATLAS Collaboration. “Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector”. In: *arXiv e-prints*, arXiv:1806.00425 (2018), arXiv:1806.00425. arXiv: [1806.00425](https://arxiv.org/abs/1806.00425) [hep-ex].
- [22] Z. Kunszt. “Associated production of heavy Higgs boson with top quarks”. In: *Nuclear Physics B* 247.2 (1984), pp. 339–359. ISSN: 0550-3213. DOI: [https://doi.org/10.1016/0550-3213\(84\)90553-4](https://doi.org/10.1016/0550-3213(84)90553-4). URL: <http://www.sciencedirect.com/science/article/pii/0550321384905534>.
- [23] L. Reina and S. Dawson. “Next-to-Leading Order Results for $t\bar{t}b\bar{h}$ Production at the Tevatron”. In: 87.20, 201804 (2001), p. 201804. DOI: [10.1103/PhysRevLett.87.201804](https://doi.org/10.1103/PhysRevLett.87.201804). arXiv: [hep-ph/0107101](https://arxiv.org/abs/hep-ph/0107101) [hep-ph].
- [24] Zhang Yu et al. “QCD NLO and EW NLO corrections to $t\bar{t}H$ production with top quark decays at hadron collider”. In: *arXiv e-prints*, arXiv:1407.1110 (2014), arXiv:1407.1110. arXiv: [1407.1110](https://arxiv.org/abs/1407.1110) [hep-ph].
- [25] D. de Florian et al. “Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector”. In: *arXiv e-prints*, arXiv:1610.07922 (2016), arXiv:1610.07922. arXiv: [1610.07922](https://arxiv.org/abs/1610.07922) [hep-ph].
- [26] M. Tanabashi et al. “Review of Particle Physics”. In: *Phys. Rev. D* 98 (3 2018), p. 030001. DOI: [10.1103/PhysRevD.98.030001](https://doi.org/10.1103/PhysRevD.98.030001). URL: <https://link.aps.org/doi/10.1103/PhysRevD.98.030001>.
- [27] CMS Collaboration. “Measurements of Higgs boson properties in the diphoton decay channel in proton-proton collisions at $\sqrt{s} = 13$ TeV”. In: *arXiv e-prints*, arXiv:1804.02716 (2018), arXiv:1804.02716. arXiv: [1804.02716](https://arxiv.org/abs/1804.02716) [hep-ex].

- [28] M. Aaboud et al. “Measurements of Higgs boson properties in the diphoton decay channel with 36 fb^{-1} of pp collision data at $\sqrt{s} = 13 \text{ TeV}$ with the ATLAS detector”. In: *Phys. Rev. D* 98 (5 2018), p. 052005. DOI: [10.1103/PhysRevD.98.052005](https://doi.org/10.1103/PhysRevD.98.052005). URL: <https://link.aps.org/doi/10.1103/PhysRevD.98.052005>.
- [29] CMS Collaboration. “Measurements of properties of the Higgs boson decaying into the four-lepton final state in pp collisions at $\sqrt{s} = 13 \text{ TeV}$ ”. In: *arXiv e-prints*, arXiv:1706.09936 (2017), arXiv:1706.09936. arXiv: [1706.09936](https://arxiv.org/abs/1706.09936) [[hep-ex](#)].
- [30] Morad Aaboud et al. “Measurement of the Higgs boson mass in the $H \rightarrow ZZ^* \rightarrow 4\ell$ and $H \rightarrow \gamma\gamma$ channels with $\sqrt{s} = 13 \text{ TeV}$ pp collisions using the ATLAS detector”. In: *Phys. Lett. B* 784 (2018), pp. 345–366. DOI: [10.1016/j.physletb.2018.07.050](https://doi.org/10.1016/j.physletb.2018.07.050). arXiv: [1806.00242](https://arxiv.org/abs/1806.00242) [[hep-ex](#)].
- [31] CMS Collaboration. “Observation of Higgs Boson Decay to Bottom Quarks”. In: *Phys. Lett. B* 784 (2018), p. 121801. DOI: [10.1103/PhysRevLett.121.121801](https://doi.org/10.1016/j.physletb.2018.07.050). arXiv: [1808.08242](https://arxiv.org/abs/1808.08242) [[hep-ex](#)].
- [32] M. Aaboud et al. “Observation of $H \rightarrow b\bar{b}$ decays and VH production with the ATLAS detector”. In: *Physics Letters B* 786 (2018), pp. 59–86. ISSN: 0370-2693. DOI: <https://doi.org/10.1016/j.physletb.2018.09.013>. URL: <http://www.sciencedirect.com/science/article/pii/S0370269318307056>.
- [33] CMS Collaboration. “Measurements of properties of the Higgs boson decaying to a W boson pair in pp collisions at $\sqrt{s} = 13 \text{ TeV}$ ”. In: *arXiv e-prints*, arXiv:1806.05246 (2018), arXiv:1806.05246. arXiv: [1806.05246](https://arxiv.org/abs/1806.05246) [[hep-ex](#)].
- [34] G. Aad et al. “Observation and measurement of Higgs boson decays to WW^* with the ATLAS detector”. In: *Phys. Rev. D* 92 (1 2015), p. 012006. DOI: [10.1103/PhysRevD.92.012006](https://doi.org/10.1103/PhysRevD.92.012006). URL: <https://link.aps.org/doi/10.1103/PhysRevD.92.012006>.
- [35] CMS Collaboration. “Observation of the Higgs boson decay to a pair of τ leptons with the CMS detector”. In: *Physics Letters B* 779 (2018), pp. 283–316. DOI: [10.1016/j.physletb.2018.02.004](https://doi.org/10.1016/j.physletb.2018.02.004). arXiv: [1708.00373](https://arxiv.org/abs/1708.00373) [[hep-ex](#)].
- [36] M. Aaboud et al. “Cross-section measurements of the Higgs boson decaying into a pair of τ -leptons in proton-proton collisions at $\sqrt{s} = 13 \text{ TeV}$ with the ATLAS detector”. In: *Phys. Rev. D* 99 (7 2019), p. 072001. DOI: [10.1103/PhysRevD.99.072001](https://doi.org/10.1103/PhysRevD.99.072001). URL: <https://link.aps.org/doi/10.1103/PhysRevD.99.072001>.
- [37] Patricia R. Burchat. (E) *DECAYS OF THE TAU LEPTON**. 1986. URL: <https://www.slac.stanford.edu/pubs/slacreports/reports14/slac-r-292.pdf>.
- [38] Achim Stahl. *Physics with Tau Leptons*. Springer-Verlag Berlin Heidelberg, 2000. ISBN: 978-3-540-48458-5. DOI: [10.1007/BFb0109618](https://doi.org/10.1007/BFb0109618).
- [39] M. L. Perl et al. “Evidence for Anomalous Lepton Production in $e^+ - e^-$ Annihilation”. In: *Phys. Rev. Lett.* 35 (22 1975), pp. 1489–1492. DOI: [10.1103/PhysRevLett.35.1489](https://doi.org/10.1103/PhysRevLett.35.1489). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.35.1489>.
- [40] Heavy Flavor Averaging Group et al. “Averages of b -hadron, c -hadron, and τ -lepton properties as of summer 2016”. In: *arXiv e-prints*, arXiv:1612.07233 (2016), arXiv:1612.07233. arXiv: [1612.07233](https://arxiv.org/abs/1612.07233) [[hep-ex](#)].

- [41] CMS Collaboration. “Reconstruction and identification of τ lepton decays to hadrons and ν_τ at CMS. Reconstruction and identification of tau lepton decays to hadrons and tau neutrino at CMS”. In: *JINST* 11.CMS-TAU-14-001. CMS-TAU-14-001. CERN-PH-EP-2015-261 (2015), P01019. 72 p. DOI: 10.1088/1748-0221/11/01/P01019. URL: <https://cds.cern.ch/record/2062435>.
- [42] CERN. *Origins of CERN*. <http://opendata.cern.ch/docs/cms-guide-for-condition-database>.
- [43] CERN. *History of CERN*. 2018. URL: <https://timeline.web.cern.ch/timelines/the-history-of-cern/overlay#1949-12-09%2000:45:00>.
- [44] F. Bordry. “Status of the large hadron collider (lhc)”. In: *Proceedings of EPAC08* (2008). URL: <http://accelconf.web.cern.ch/AccelConf/e08/papers/moxagm01.pdf>.
- [45] L. Evans and P. Bryant. “LHC Machine”. In: *JINST* 8 (2008). URL: <https://doi.org/10.1088/1748-0221/3/08/S08001>.
- [46] CERN. *Restarting the LHC: Why 13 TeV?* 2018. URL: <https://home.cern/about/engineering/restarting-lhc-why-13-tev>.
- [47] Nikolaos Karastathis et al. “Crossing Angle Anti-Leveling at the LHC in 2017”. In: *J. Phys. : Conf. Ser.* 1067.CERN-ACC-2018-125. 2 (2018), MOPMF040. 5 p. DOI: 10.18429/JACoW-IPAC2018-MOPMF040. URL: <https://cds.cern.ch/record/2648700>.
- [48] CERN. *Beta & Emittance. Taking a closer look at LHC*. 2018. URL: https://www.lhc-closer.es/taking_a_closer_look_at_lhc/0.beta__emittance.
- [49] CERN. *LHC report: full house for the LHC*. 2017. URL: <https://home.cern/news/news/accelerators/lhc-report-full-house-lhc>.
- [50] CMS Collaboration. *Public CMS Luminosity Information*. 2019. URL: <https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults>.
- [51] David Barney and Tai Sakuma. “Sketchup images highlighting the sub-detectors”. In: (2017). URL: <https://cds.cern.ch/record/2628519>.
- [52] CMS Collaboration. “The CMS experiment at the CERN LHC”. In: *Journal of Instrumentation* 3.08 (2008), S08004. URL: <http://stacks.iop.org/1748-0221/3/i=08/a=S08004>.
- [53] CMS Collaboration. *CMS Physics. Technical Design Report*. Tech. rep. CERN/LHCC 2006-001. 2006. URL: <http://cmsdoc.cern.ch/cms/cpt/tdr/>.
- [54] Manfred Krammer. *Silicon Detectors*. 2010. URL: http://www.hephy.at/fileadmin/user_upload/Lehre/Unterlagen/Praktikum/Halbleiterdetektoren.pdf.
- [55] Viktor Veszpremi. “Operation and performance of the CMS tracker”. In: arXiv:1402.0675v1 (2014). URL: <https://arxiv.org/pdf/1402.0675.pdf>.
- [56] Katja Klein. “The Phase-1 upgrade of the CMS pixel detector”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 845 (2017). Proceedings of the Vienna Conference on Instrumentation 2016, pp. 101–105. ISSN: 0168-9002. DOI: <https://doi.org/10.1016/j.nima.2016.06.039>. URL: <http://www.sciencedirect.com/science/article/pii/S0168900216305824>.

- [57] The CMS Collaboration. "Description and performance of track and primary-vertex reconstruction with the CMS tracker". In: *Journal of Instrumentation* 9.10, P10009 (2014), P10009. DOI: [10.1088/1748-0221/9/10/P10009](https://doi.org/10.1088/1748-0221/9/10/P10009). arXiv: [1405.6569](https://arxiv.org/abs/1405.6569).
- [58] CMS Collaboration. "Energy calibration and resolution of the CMS electromagnetic calorimeter in pp collisions at $\sqrt{s} = 7$ TeV". In: *Journal of Instrumentation* 8.9, P09009 (2013), P09009. DOI: [10.1088/1748-0221/8/09/P09009](https://doi.org/10.1088/1748-0221/8/09/P09009). arXiv: [1306.2016](https://arxiv.org/abs/1306.2016) [[hep-ex](https://arxiv.org/abs/1306.2016)].
- [59] CMS Collaboration. "Performance and operation of the CMS electromagnetic calorimeter". In: *Journal of Instrumentation* 5.03 (2010), T03010. URL: <http://stacks.iop.org/1748-0221/5/i=03/a=T03010>.
- [60] CMS Collaboration. "Performance of the CMS hadron calorimeter with cosmic ray muons and LHC beam data". In: *Journal of Instrumentation* 5.3 (2010), T03012. DOI: [10.1088/1748-0221/5/03/T03012](https://doi.org/10.1088/1748-0221/5/03/T03012). arXiv: [0911.4991](https://arxiv.org/abs/0911.4991) [[physics.ins-det](https://arxiv.org/abs/0911.4991)].
- [61] P. Fonte. "Applications and new developments in Resistive Plate Chambers". In: *IEEE Nuclear, Science Symposium* (2001).
- [62] A Tapper and Darin Acosta. *CMS Technical Design Report for the Level-1 Trigger Upgrade*. Tech. rep. CERN-LHCC-2013-011. CMS-TDR-12. Additional contacts: Jeffrey Spalding, Fermilab, Jeffrey.Spalding@cern.ch Didier Contardo, Universite Claude Bernard-Lyon I, didier.claude.contardo@cern.ch. 2013. URL: <https://cds.cern.ch/record/1556311>.
- [63] A. Zabi et al. "The CMS Level-1 calorimeter trigger upgrade for the LHC Run II". In: *International Conference on Technology and Instrumentation in Particle Physics*. Vol. 12. 01. IOP Publishing, 2017, pp. C01065–C01065. DOI: [10.1088/1748-0221/12/01/c01065](https://doi.org/10.1088/1748-0221/12/01/c01065). URL: <https://doi.org/10.1088/1748-0221/12/01/c01065>.
- [64] M. Konecki. "The CMS Level-1 muon triggers for the LHC Run II". In: *ICHEP*. 2018.
- [65] Laurent Thomas. "CMS High Level Trigger performance at 13 TeV". In: *Proceedings of Science* (2018).
- [66] CMS Collaboration. "Particle-flow reconstruction and global event description with the CMS detector. Particle-flow reconstruction and global event description with the CMS detector". In: *JINST* 12.CMS-PRF-14-001. CMS-PRF-14-001-004. 10 (2017). Replaced with the published version. Added the journal reference and DOI. All the figures and tables can be found at <http://cms-results.web.cern.ch/cms-results/public-results/publications/PRF-14-001> (CMS Public Pages), P10003. 82 p. URL: <http://cds.cern.ch/record/2270046>.
- [67] Wolfgang Adam et al. *Track Reconstruction in the CMS tracker*. Tech. rep. CMS-NOTE-2006-041. Geneva: CERN, 2006. URL: <https://cds.cern.ch/record/934067>.
- [68] CMS Collaboration. "Electron Reconstruction within the Particle Flow Algorithm". In: (2010).
- [69] *Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET*. Tech. rep. CMS-PAS-PFT-09-001. Geneva: CERN, 2009. URL: <https://cds.cern.ch/record/1194487>.

- [70] Gerald C. Blazey et al. "Run II Jet Physics: Proceedings of the Run II QCD and Weak Boson Physics Workshop". In: *arXiv e-prints*, hep-ex/0005012 (2000), hep-ex/0005012. arXiv: [hep-ex/0005012](https://arxiv.org/abs/hep-ex/0005012) [[hep-ex](#)].
- [71] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "The anti- k_t jet clustering algorithm". In: *JHEP* 04 (2008), p. 063. DOI: [10.1088/1126-6708/2008/04/063](https://doi.org/10.1088/1126-6708/2008/04/063). arXiv: [0802.1189](https://arxiv.org/abs/0802.1189) [[hep-ph](#)].
- [72] Vardan Khachatryan et al. "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV". In: *JINST* 12.02 (2017), P02014. DOI: [10.1088/1748-0221/12/02/P02014](https://doi.org/10.1088/1748-0221/12/02/P02014). arXiv: [1607.03663](https://arxiv.org/abs/1607.03663) [[hep-ex](#)].
- [73] CMS Collaboration. "Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s}=8\text{TeV}$ ". In: 10 (2015), P06005. DOI: [10.1088/1748-0221/10/06/P06005](https://doi.org/10.1088/1748-0221/10/06/P06005).
- [74] CMS Collaboration. "Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7 \text{ TeV}$ ". In: *JINST* 7 (2012), P10002. DOI: [doi:10.1088/1748-0221/7/10/P10002](https://doi.org/10.1088/1748-0221/7/10/P10002).
- [75] CMS Collaboration. "Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and E_T^{miss} ". In: *CMS Physics Analysis Summary CMS-PAS-PFT-09-001* (2009).
- [76] CMS Collaboration. "Commissioning of the Particle-Flow Reconstruction in Minimum-Bias and Jet Events from pp Collisions at 7 TeV". In: *CMS-PAS-PFT-10-002* (2010).
- [77] CMS Collaboration. "Performance of τ -lepton reconstruction and identification in CMS". In: *JINST* 7 (2012), P01001. DOI: [10.1088/1748-0221/7/01/P01001](https://doi.org/10.1088/1748-0221/7/01/P01001).
- [78] The CMS Collaboration. "Reconstruction and identification of τ lepton decays to hadrons and ν_τ at CMS". In: *Journal of Instrumentation* 11.1, P01019 (2016), P01019. DOI: [10.1088/1748-0221/11/01/P01019](https://doi.org/10.1088/1748-0221/11/01/P01019). arXiv: [1510.07488](https://arxiv.org/abs/1510.07488).
- [79] CMS Collaboration. "Performance of reconstruction and identification of τ leptons decaying to hadrons and ν_τ in pp collisions at $\sqrt{s} = 13 \text{ TeV}$ ". In: *JINST* 13. arXiv:1809.02816. CMS-TAU-16-003-003 (2018). Replaced with the published version. Added the journal reference and the DOI. All the figures and tables can be found at <http://cms-results.web.cern.ch/cms-results/public-results/publications/TAU-16-003> (CMS Public Pages), P10005. 71 p. DOI: [10.1088/1748-0221/13/10/P10005](https://doi.org/10.1088/1748-0221/13/10/P10005). URL: <https://cds.cern.ch/record/2637646>.
- [80] C. Veelken L. Bianchini. "Reconstruction of the Higgs mass in events with Higgs bosons decaying into a pair of τ leptons using matrix element techniques". In: *Nuclear Instruments and Methods in Physics Research A* 862 (Aug. 2017), pp. 54–84. DOI: [10.1016/j.nima.2017.05.001](https://doi.org/10.1016/j.nima.2017.05.001). arXiv: [1603.05910](https://arxiv.org/abs/1603.05910) [[hep-ex](#)].
- [81] CMS Collaboration. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". In: *Physics Letters B* 716.1 (2012), pp. 30–61. DOI: [10.1016/j.physletb.2012.08.021](https://doi.org/10.1016/j.physletb.2012.08.021). arXiv: [1207.7235](https://arxiv.org/abs/1207.7235) [[hep-ex](#)].
- [82] Albert M Sirunyan et al. "Observation of the Higgs boson decay to a pair of τ leptons with the CMS detector". In: *Phys. Lett. B* 779 (2018), pp. 283–316. DOI: [10.1016/j.physletb.2018.02.004](https://doi.org/10.1016/j.physletb.2018.02.004). arXiv: [1708.00373](https://arxiv.org/abs/1708.00373) [[hep-ex](#)].

- [83] Simone Alioli et al. “A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX”. In: *Journal of High Energy Physics* 2010, 43 (2010), p. 43. DOI: [10.1007/JHEP06\(2010\)043](https://doi.org/10.1007/JHEP06(2010)043). arXiv: [1002.2581](https://arxiv.org/abs/1002.2581) [hep-ph].
- [84] A. Denner et al. “Standard model Higgs-boson branching ratios with uncertainties”. In: *European Physical Journal C* 71, 1753 (2011), p. 1753. DOI: [10.1140/epjc/s10052-011-1753-8](https://doi.org/10.1140/epjc/s10052-011-1753-8). arXiv: [1107.5909](https://arxiv.org/abs/1107.5909) [hep-ph].
- [85] J. Alwall et al. “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations”. In: *Journal of High Energy Physics* 2014, 79 (2014), p. 79. DOI: [10.1007/JHEP07\(2014\)079](https://doi.org/10.1007/JHEP07(2014)079). arXiv: [1405.0301](https://arxiv.org/abs/1405.0301) [hep-ph].
- [86] Torbjörn Sjöstrand et al. “An introduction to PYTHIA 8.2”. In: *Computer Physics Communications* 191 (2015), pp. 159–177. DOI: [10.1016/j.cpc.2015.01.024](https://doi.org/10.1016/j.cpc.2015.01.024). arXiv: [1410.3012](https://arxiv.org/abs/1410.3012) [hep-ph].
- [87] CMS Collaboration. “Event generator tunes obtained from underlying event and multiparton scattering measurements”. In: *arXiv e-prints*, arXiv:1512.00815 (2015), arXiv:1512.00815. arXiv: [1512.00815](https://arxiv.org/abs/1512.00815) [hep-ex].
- [88] S. Agostinelli et al. “Geant4—a simulation toolkit”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), pp. 250–303. ISSN: 0168-9002. DOI: [https://doi.org/10.1016/S0168-9002\(03\)01368-8](https://doi.org/10.1016/S0168-9002(03)01368-8). URL: <http://www.sciencedirect.com/science/article/pii/S0168900203013688>.
- [89] CERN. *CMS Guide to the CMS condition database*. <http://opendata.cern.ch/docs/cms-guide-for-condition-database>.
- [90] Matteo Cacciari and Gavin P. Salam. “Pileup subtraction using jet areas”. In: *Physics Letters B* 659.1 (2008), pp. 119–126. ISSN: 0370-2693. DOI: <https://doi.org/10.1016/j.physletb.2007.09.077>. URL: <http://www.sciencedirect.com/science/article/pii/S0370269307011094>.
- [91] *Search for the Standard-Model Higgs boson decaying to tau pairs in proton-proton collisions at $\sqrt{s} = 7$ and 8 TeV*. Tech. rep. CMS-PAS-HIG-13-004. Geneva: CERN, 2013. URL: <https://cds.cern.ch/record/1528271>.
- [92] Stefano Frixione and Bryan R. Webber. “Matching NLO QCD computations and parton shower simulations”. In: *Journal of High Energy Physics* 2002.6, 029 (2002), p. 029. DOI: [10.1088/1126-6708/2002/06/029](https://doi.org/10.1088/1126-6708/2002/06/029). arXiv: [hep-ph/0204244](https://arxiv.org/abs/hep-ph/0204244) [hep-ph].
- [93] S. Frixione and B. R. Webber. “The MC@NLO 3.3 Event Generator”. In: *arXiv e-prints*, hep-ph/0612272 (2006), hep-ph/0612272. arXiv: [hep-ph/0612272](https://arxiv.org/abs/hep-ph/0612272) [hep-ph].
- [94] J. Alwall et al. “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations”. In: *Journal of High Energy Physics* 2014, 79 (2014), p. 79. DOI: [10.1007/JHEP07\(2014\)079](https://doi.org/10.1007/JHEP07(2014)079). arXiv: [1405.0301](https://arxiv.org/abs/1405.0301) [hep-ph].
- [95] CMS Collaboration. “Combined results of searches for the standard model Higgs boson in pp collisions at $\sqrt{s} = 7$ TeV”. In: *Physics Letters B* 710.1 (2012), pp. 26–48. DOI: [10.1016/j.physletb.2012.02.064](https://doi.org/10.1016/j.physletb.2012.02.064). arXiv: [1202.1488](https://arxiv.org/abs/1202.1488) [hep-ex].

- [96] A. L. Samuel. "Some Studies in Machine Learning Using the Game of Checkers". In: *IBM J. Res. Dev.* 3.3 (July 1959), pp. 210–229. ISSN: 0018-8646. DOI: [10.1147/rd.33.0210](https://doi.org/10.1147/rd.33.0210). URL: <http://dx.doi.org/10.1147/rd.33.0210>.
- [97] David Kelnar. *The fourth industrial revolution: a primer on Artificial Intelligence (AI)*. <https://medium.com/mmc-writes/the-fourth-industrial-revolution-a-primer-on-artificial-intelligence-ai-ff5e7fffcae1>.
- [98] Goncalo de Vasconcelos. *The Third Industrial Revolution – Internet, Energy And A New Financial System*. www.forbes.com/sites/goncalodevasconcelos/2015/03/04/the-third-industrial-revolution-internet-energy-and-a-new-financial-system/.
- [99] Wikipedia. *Machine learning*. https://en.wikipedia.org/wiki/Machine_learning.
- [100] Coursera. *Machine learning*. <https://www.coursera.org/lecture/machine-learning/what-is-machine-learning-Ujm7v>.
- [101] Expert System. *Machine learning*. <https://www.expertsystem.com/machine-learning-definition/>.
- [102] Mark Stamp. *Introduction to Machine Learning with Applications in Information Security*. 1st ed. Vol. 1. Chapman and Hall/CRC, Sept. 2017. ISBN: 9781138626782.
- [103] Ralf Herbrich. *Learning Kernel Classifiers: Theory and Algorithms*. 1st ed. Vol. 1. The MIT Press, 2001. ISBN: 0-262-08306-X.
- [104] Olivier Bousquet, Ulrike von Luxburg, and Gunnar Rätsch, eds. *Advanced Lectures on Machine Learning, ML Summer Schools 2003, Canberra, Australia, February 2-14, 2003, Tübingen, Germany, August 4-16, 2003, Revised Lectures*. Vol. 3176. Lecture Notes in Computer Science. Springer, 2004. ISBN: 3-540-23122-6. DOI: [10.1007/b100712](https://doi.org/10.1007/b100712). URL: <https://doi.org/10.1007/b100712>.
- [105] Cade Metz. *In a Huge Breakthrough, Google's AI Beats a Top Player at the Game of Go*. <https://www.wired.com/2016/01/in-a-huge-breakthrough-googles-ai-beats-a-top-player-at-the-game-of-go/>.
- [106] Spark.ml. *Overview: estimators, transformers and pipelines*. <https://spark.apache.org/docs/1.6.0/ml-guide.html>.
- [107] Google. *Estimators*. <https://www.tensorflow.org/guide/estimators>.
- [108] Meadowlark Bradsher. *Classifier vs model vs estimator*. <https://stats.stackexchange.com/questions/103475/classifier-vs-model-vs-estimator>.
- [109] CERN. *Save Data*. 2018. URL: <https://root.cern/save-data>.
- [110] HDF Group. *The HDF Group*. <https://www.hdfgroup.org/>.
- [111] Inc. Wikimedia Foundation. *Feature scaling*. https://en.wikipedia.org/wiki/Feature_scaling.
- [112] George G. Roussas. *An Introduction to Probability and Statistical Inference*. 2nd ed. Vol. 1. Elsevier, 2015. ISBN: 978-0-12-800114-1.
- [113] Robert D. Grober. "PGA Tour Scores as a Gaussian Random Variable". In: (). arXiv: [0802.3890 \[cs\]](https://arxiv.org/abs/0802.3890).
- [114] Inc. Wikimedia Foundation. *Probability integral transform*. https://en.wikipedia.org/wiki/Probability_integral_transform.
- [115] Hosam M. Mahmoud. *Sorting: A Distribution Theory*. 1st ed. Vol. 1. John Wiley and Sons, Inc., 2011. ISBN: 9780471327103.

- [116] Stack Exchange Inc. *Why is the CDF of a sample uniformly distributed*. <https://stats.stackexchange.com/questions/161635/why-is-the-cdf-of-a-sample-uniformly-distributed>.
- [117] Inc. Wikimedia Foundation. *Inverse transform sampling*. https://en.wikipedia.org/wiki/Inverse_transform_sampling.
- [118] Shahar Cohen Haim Dahan. *Proactive Data Mining with Decision Trees*. 1st ed. SpringerBriefs in Electrical and Computer Engineering. Springer-Verlag New York, 2014. ISBN: 978-1-4939-0538-6,978-1-4939-0539-3. URL: <http://gen.lib.rus.ec/book/index.php?md5=d0bad5bce4bbcc91b82b43c04e1d8650>.
- [119] Oded Maimon Lior Rokach. *Data Mining With Decision Trees: Theory and Applications (2nd Edition)*. 2nd ed. Series in Machine Perception and Artificial Intelligence. World Scientific Publishing Company, 2015. ISBN: 981459007X,9789814590075. URL: <http://gen.lib.rus.ec/book/index.php?md5=6f9b56573e3f1d3b32a642f210281642>.
- [120] Katie Waddle. *Group actions on trees*. version 17 Aug 2007. REU work, VI-GRE 2007. 2007. URL: <http://gen.lib.rus.ec/book/index.php?md5=f5c6469122620f9382357b88d1f420be>.
- [121] Inc. Wikimedia Foundation. *Tree (graph theory)*. [https://en.wikipedia.org/wiki/Tree_\(graph_theory\)](https://en.wikipedia.org/wiki/Tree_(graph_theory)).
- [122] Thomas Hancock et al. "Lower Bounds on Learning Decision Lists and Trees". In: *Information and computation* 126.40 (1996), pp. 114–122. URL: <https://core.ac.uk/download/pdf/81927999.pdf>.
- [123] Laurent Hyafil and Ronald L. Rivest. "Constructing optimal binary decision trees is NP-complete". In: *Information Processing Letters* 5.1 (1976), pp. 15–17.
- [124] C. E. Shannon. "A Mathematical Theory of Communication". In: *SIGMOBILE Mob. Comput. Commun. Rev.* 5.1 (Jan. 2001), pp. 3–55. ISSN: 1559-1662. DOI: [10.1145/584091.584093](https://doi.org/10.1145/584091.584093). URL: <http://doi.acm.org/10.1145/584091.584093>.
- [125] Yisen Wang, Chaobing Song, and Shu-Tao Xia. "Unifying Decision Trees Split Criteria Using Tsallis Entropy". In: *arXiv e-prints*, arXiv:1511.08136 (Nov. 2015), arXiv:1511.08136. arXiv: [1511.08136](https://arxiv.org/abs/1511.08136) [stat.ML].
- [126] Leo Breiman et al. *Classification and regression trees*. The Wadsworth statistics / probability series. CRC, 1984. ISBN: 0-412-04841-8. URL: <http://gen.lib.rus.ec/book/index.php?md5=26908b6eda02ca4faf25adb57a12b26>.
- [127] Inc. Wikimedia Foundation. *Greedy algorithm*. https://en.wikipedia.org/wiki/Greedy_algorithm.
- [128] F. Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.
- [129] Inc. GitHub. *Scikit ROC*. <https://github.com/scikit-learn/scikit-learn/blob/7b136e9/sklearn/metrics/ranking.py>.
- [130] J.A. Hanley and Barbara Mcneil. "A Method of Comparing the Areas Under Receiver Operating Characteristic Curves Derived from the Same Cases". In: *Radiology* 148 (Oct. 1983), pp. 839–43. DOI: [10.1148/radiology.148.3.6878708](https://doi.org/10.1148/radiology.148.3.6878708).
- [131] Aleksander Kołcz et al. "Data duplication: an imbalance problem?" In: (Jan. 2003). URL: <https://www.site.uottawa.ca/~nat/Workshop2003/imbalance-kolcz.pdf>.

- [132] scikit learn. *Decision Tree Classifier*. <https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html>.
- [133] Tin Kam Ho. "Random Decision Forests". In: (1995). URL: <https://web.archive.org/web/20160417030218/http://ect.bell-labs.com/who/tkh/publications/papers/odt.pdf>.
- [134] E.M. Kleinberg. "Stochastic discrimination". In: (1990). DOI: 10.1007/BF01531079. URL: <https://ieeexplore.ieee.org/document/709601/metrics>.
- [135] E.M. Kleinberg. "On the Algorithmic Implementation of Stochastic Discrimination". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22 (2000), pp. 473–490.
- [136] Tin Kam Ho. "The Random Subspace Method for Constructing Decision Forests". In: (1998). DOI: 10.1109/34.709601. URL: <https://ieeexplore.ieee.org/document/709601/metrics>.
- [137] Leo Breiman. "Random Forests". In: (2001). DOI: 10.1023/A:1010933404324.
- [138] scikit learn. *Random Forest Classifier*. <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>.
- [139] M. Kearns. "Thoughts on Hypothesis Boosting". In: *Unpublished manuscript (Machine Learning class project, December 1988)* (1988). URL: <http://www.cis.upenn.edu/~mkearns/papers/boostnote.pdf>.
- [140] M. Kearns and L. G. Valiant. "Cryptographic Limitations on Learning Boolean Formulae and Finite Automata". In: *Proceedings of the Twenty-first Annual ACM Symposium on Theory of Computing*. STOC '89. Seattle, Washington, USA: ACM, 1989, pp. 433–444. ISBN: 0-89791-307-8. DOI: 10.1145/73007.73049. URL: <http://doi.acm.org/10.1145/73007.73049>.
- [141] R. E. Schapire. "The strength of weak learnability". In: *Machine Learning* 5.2 (1990), pp. 197–227. ISSN: 1573-0565. DOI: 10.1007/BF00116037. URL: <https://doi.org/10.1007/BF00116037>.
- [142] Y. Freund and R. E. Schapire. *A Decision-Theoretic Generalization of on-Line Learning and an Application to Boosting*. 1997.
- [143] Balázs Kégl. "The return of AdaBoost.MH: multi-class Hamming trees". In: *arXiv e-prints*, arXiv:1312.6086 (2013), arXiv:1312.6086. arXiv: 1312.6086 [cs.LG].
- [144] J. H. Friedman. *Greedy Function Approximation: A Gradient Boosting Machine*. 1999. URL: <https://statweb.stanford.edu/~jhf/ftp/trebst.pdf>.
- [145] T. Hastie, R. Tibshirani, and J. Friedman. *Elements of Statistical Learning Ed. 2*. 2009.
- [146] Adam Paszke et al. "Automatic differentiation in PyTorch". In: (2017).
- [147] Andrej Karpathy. *Convolutional Neural Networks*. 2019. URL: <https://cs231n.github.io/convolutional-networks>.
- [148] PyTorch. *Autograd: Automatic Differentiation*. 2019. URL: https://pytorch.org/tutorials/beginner/blitz/autograd_tutorial.html.
- [149] Josh Patterson. *Deep Learning: A Practitioner's Approach*. 2017.
- [150] G. D. Magoulas, M. N. Vrahatis, and G. S. Androulakis. "Improving the Convergence of the Backpropagation Algorithm Using Learning Rate Adaptation Methods". In: (1999). URL: <http://www.dcs.bbk.ac.uk/~gmagoulas/NEUCOMP.pdf>.

- [151] Diederik P. Kingma and Jimmy Ba. “Adam: A Method for Stochastic Optimization”. In: *arXiv e-prints*, arXiv:1412.6980 (2014), arXiv:1412.6980. arXiv: 1412.6980 [cs.LG].
- [152] Quora. *What are the pros and cons of neural networks from a practical perspective?* 2019. URL: <https://www.quora.com/What-are-the-pros-and-cons-of-neural-networks-from-a-practical-perspective-Personal-comments-from-heavy-users-welcome>.
- [153] CMS Collaboration. “Evidence for the 125 GeV Higgs boson decaying to a pair of tau leptons”. In: *Journal of High Energy Physics, JHEP 05 (2014) 104* 5.104 (2014). DOI: [doi:10.1007/JHEP05\(2014\)104](https://doi.org/10.1007/JHEP05(2014)104).
- [154] ATLAS and CMS Collaborations. “Procedure for the LHC Higgs boson search combination in Summer 2011”. In: (2011). URL: <https://cdsweb.cern.ch/record/1379837>.
- [155] GitHub. *Combine Limit*. 2019. URL: <https://github.com/cms-analysis/HiggsAnalysis-CombinedLimit/wiki>.
- [156] J. Nelder and R. Mead. “A simplex method for function minimization”. In: *The Computer Journal* (1965), pp. 308–313.
- [157] Fuchang Gao and Lixing Han. “Implementing the Nelder-Mead simplex algorithm with adaptive parameters”. In: (2010). DOI: [10.1007/s10589-010-9329-3](https://doi.org/10.1007/s10589-010-9329-3). URL: <https://www.webpages.uidaho.edu/~fuchang/res/ANMS.pdf>.
- [158] D. Wales and Doye J. “Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms”. In: *Journal of Physical Chemistry A* (1997).
- [159] *Combined measurements of the Higgs boson’s couplings at $\sqrt{s} = 13$ TeV*. Tech. rep. CMS-PAS-HIG-17-031. Geneva: CERN, 2018. URL: <https://cds.cern.ch/record/2308127>.
- [160] *Measurement of Higgs boson production and decay to the $\tau\tau$ final state*. Tech. rep. CMS-PAS-HIG-18-032. Geneva: CERN, 2019. URL: <https://cds.cern.ch/record/2668685>.
- [161] L Tuura et al. “CMS data quality monitoring: Systems and experiences”. In: *Journal of Physics: Conference Series* 219.7 (2010), p. 072020. DOI: [10.1088/1742-6596/219/7/072020](https://doi.org/10.1088/1742-6596/219/7/072020). URL: <https://doi.org/10.1088/1742-6596/219/7/072020>.
- [162] CERN. *CMS Simulated Dataset Names*. <http://opendata.cern.ch/docs/cms-simulated-dataset-names>.
- [163] *WorkBookMiniAOD*. <https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookMiniAOD>. Accessed: 2018-08-15.
- [164] L. Garre. *MONTE CARLO PARTICLE NUMBERING SCHEME*. <http://pdg.lbl.gov/2005/reviews/montecarlорpp.pdf>.
- [165] Valeria Bartsch and Gunter Quast. *Expected Signal Observability at Future Experiments*. Tech. rep. CMS-NOTE-2005-004. Geneva: CERN, 2005. URL: <https://cds.cern.ch/record/824351>.
- [166] Wikipedia. *Statistical hypothesis testing*. 2019. URL: https://en.wikipedia.org/wiki/Statistical_hypothesis_testing.
- [167] K. Cranmer. *Statistical Challenges for Searches for New Physics at the LHC*. Ed. by Louis Lyons and Müge Karagöz Ünel. 2006. DOI: [10.1142/9781860948985_0026](https://doi.org/10.1142/9781860948985_0026). arXiv: [physics/0511028](https://arxiv.org/abs/physics/0511028) [physics.data-an].

- [168] Glen Cowan et al. "Asymptotic formulae for likelihood-based tests of new physics". In: *Eur. Phys. J. C*71 (2011). [Erratum: *Eur. Phys. J. C*73,2501(2013)], p. 1554. DOI: [10.1140/epjc/s10052-011-1554-0](https://doi.org/10.1140/epjc/s10052-011-1554-0), [10.1140/epjc/s10052-013-2501-z](https://doi.org/10.1140/epjc/s10052-013-2501-z). arXiv: [1007.1727](https://arxiv.org/abs/1007.1727) [physics.data-an].

Appendix A

Datasets

DY + Jets, Recoil Corrections applied	
Sample	Cross section (pb)
/DYJetsToLL-M-50-TuneCUETP8M1-13TeV-madgraphMLM-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-ext1-v2/MINIAODSIM	5765.4
/DYJetsToLL-M-50-TuneCUETP8M1-13TeV-madgraphMLM-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-ext2-v1/MINIAODSIM	5765.4
/DY1JetsToLL-M-50-TuneCUETP8M1-13TeV-madgraphMLM-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-v1/MINIAODSIM	1012.5
/DY2JetsToLL-M-50-TuneCUETP8M1-13TeV-madgraphMLM-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-v1/MINIAODSIM	332.8
/DY3JetsToLL-M-50-TuneCUETP8M1-13TeV-madgraphMLM-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-v1/MINIAODSIM	101.8
/DY4JetsToLL-M-50-TuneCUETP8M1-13TeV-madgraphMLM-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-v1/MINIAODSIM	54.8
WZ, no Recoil Corrections	
Sample	Cross section (pb)
/WZTo1L3Nu-13TeV-amcatnloFXFX-madspin-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-v1/MINIAODSIM	3.05
/WZTo1L1Nu2Q-13TeV-amcatnloFXFX-madspin-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-v3/MINIAODSIM	10.71
/WZTo2L2Q-13TeV-amcatnloFXFX-madspin-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-v1/MINIAODSIM	5.595
/WZTo3LNU-TuneCUETP8M1-13TeV-amcatnloFXFX-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-v1/MINIAODSIM	4.708
/WZTo3LNU-TuneCUETP8M1-13TeV-powheg-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-v1/MINIAODSIM	4.708
W + Jets, Recoil Corrections applied	
Sample	Cross section (pb)
/WJetsToLNu-TuneCUETP8M1-13TeV-madgraphMLM-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-v1/MINIAODSIM	61526.7
/WJetsToLNu-TuneCUETP8M1-13TeV-madgraphMLM-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-ext2-v1/MINIAODSI	61526.7
/W1JetsToLNu-TuneCUETP8M1-13TeV-madgraphMLM-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-v1/MINIAODSIM	9644.5

/W2JetsToLNu-TuneCUETP8M1-13TeV-madgraphMLM-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-v1/MINIAODSIM	3144.5
/W2JetsToLNu-TuneCUETP8M1-13TeV-madgraphMLM-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-ext1-v1/MINIAODSIM	3144.5
/W3JetsToLNu-TuneCUETP8M1-13TeV-madgraphMLM-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-v1/MINIAODSIM	954.8
/W3JetsToLNu-TuneCUETP8M1-13TeV-madgraphMLM-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-ext1-v1/MINIAODSIM	954.8
/W4JetsToLNu-TuneCUETP8M1-13TeV-madgraphMLM-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-v1/MINIAODSIM	485.6
/W4JetsToLNu-TuneCUETP8M1-13TeV-madgraphMLM-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-ext1-v1/MINIAODSIM	485.6
/W4JetsToLNu-TuneCUETP8M1-13TeV-madgraphMLM-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-ext2-v1/MINIAODSIM	485.6
TTbar, no Recoil Corrections	
Sample	Cross section (pb)
/TT-TuneCUETP8M2T4-13TeV-powheg-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-v1/MINIAODSIM	831.76
Higgs, no Recoil Corrections	
Sample	Cross section (pb)
/GluGluHToTauTau-M125-13TeV-powheg-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-v1/MINIAODSIM	48.58*0.0627
/VBFHToTauTau-M125-13TeV-powheg-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-v1/MINIAODSIM	3.782*0.0627
/WplusHToTauTau-M125-13TeV-powheg-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-v1/MINIAODSIM	0.8400*0.0627
/WminusHToTauTau-M125-13TeV-powheg-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-v1/MINIAODSIM	0.5328*0.0627
Higgs small, no Recoil Corrections	
Sample	Cross section (pb)
/ZHToTauTau-M125-13TeV-powheg-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-v1/MINIAODSIM	0.884*0.0627
/VBFHToWWTo2L2Nu-M125-13TeV-powheg-JHUGen-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-v1/MINIAODSIM	0.0858
/GluGluHToWWTo2L2Nu-M125-13TeV-powheg-JHUGen-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-v1/MINIAODSIM	1.001
Single top, no Recoil Corrections	
Sample	Cross section (pb)
/ST-tW-antitop-5f-inclusiveDecays-13TeV-powheg-pythia8-TuneCUETP8M1/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-ext1-v1/MINIAODSIM	38.09
/ST-tW-top-5f-inclusiveDecays-13TeV-powheg-pythia8-TuneCUETP8M1/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-ext1-v1/MINIAODSIM	38.09

/ST-t-channel-antitop-4f-inclusiveDecays-13TeV-powhegV2-madspin-pythia8-TuneCUETP8M1/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-v1/MINIAODSIM	24.11
/ST-t-channel-top-4f-inclusiveDecays-13TeV-powhegV2-madspin-pythia8-TuneCUETP8M1/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-v1/MINIAODSIM	39.852
WW, no Recoil Corrections	
Sample	Cross section (pb)
/WWToLNUQQ-13TeV-powheg/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-v1/MINIAODSIM	43.53
/WWToLNUQQ-13TeV-powheg/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-ext1-v1/MINIAODSIM	43.53
/WWTo1L1Nu2Q-13TeV-amcatnloFFFX-madspin-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-v1/MINIAODSIM	43.53
/WWTo2L2Nu-13TeV-powheg/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-v1/MINIAODSIM	10.48
ZZ, no Recoil Corrections	
Sample	Cross section (pb)
/ZZTo2L2Q-13TeV-amcatnloFFFX-madspin-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-v1/MINIAODSIM	3.22
/ZZTo2Q2Nu-13TeV-amcatnloFFFX-madspin-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-v1/MINIAODSIM	4.03
/ZZTo2L2Nu-13TeV-powheg-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-v1/MINIAODSIM	3.22
/ZZTo4L-13TeV-amcatnloFFFX-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-ext1-v1/MINIAODSIM	1.212
Electroweak, no Recoil Corrections	
Sample	Cross section (pb)
/EWKWMinus2Jets-WToLNU-M-50-13TeV-madgraph-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-v1/MINIAODSIM	20.25
/EWKWMinus2Jets-WToLNU-M-50-13TeV-madgraph-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-ext1-v1/MINIAODSIM	20.25
/EWKWMinus2Jets-WToLNU-M-50-13TeV-madgraph-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-ext2-v1/MINIAODSIM	20.25
/EWKWPlus2Jets-WToLNU-M-50-13TeV-madgraph-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-v1/MINIAODSIM	25.62
/EWKWPlus2Jets-WToLNU-M-50-13TeV-madgraph-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-ext1-v1/MINIAODSIM	25.62
/EWKWPlus2Jets-WToLNU-M-50-13TeV-madgraph-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-ext2-v1/MINIAODSIM	25.62
/EWKZ2Jets-ZToLL-M-50-13TeV-madgraph-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-v1/MINIAODSIM	3.987
/EWKZ2Jets-ZToLL-M-50-13TeV-madgraph-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-ext1-v1/MINIAODSIM	3.987

/EWKZ2Jets-ZToLL-M-50-13TeV-madgraph-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-ext2-v1/MINIAODSIM	3.987
/EWKZ2Jets-ZToNuNu-13TeV-madgraph-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-v1/MINIAODSIM	10.01
/EWKZ2Jets-ZToNuNu-13TeV-madgraph-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-ext1-v1/MINIAODSIM	10.01
/EWKZ2Jets-ZToNuNu-13TeV-madgraph-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-ext2-v1/MINIAODSIM	10.01
WGamma, no Recoil Corrections	
Sample	Cross section (pb)
/WGToLNUG-TuneCUETP8M1-13TeV-amcatnloFXFX-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-ext1-v1/MINIAODSIM	489.0
/WGToLNUG-TuneCUETP8M1-13TeV-amcatnloFXFX-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-ext2-v1/MINIAODSIM	489.0
/WGToLNUG-TuneCUETP8M1-13TeV-amcatnloFXFX-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-ext3-v1/MINIAODSIM	489.0
/WGstarToLNUMuMu-012Jets-13TeV-madgraph/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-v1/MINIAODSIM	2.793
/WGstarToLNUEE-012Jets-13TeV-madgraph/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-v1/MINIAODSIM	3.526
VV, no Recoil Corrections	
Sample	Cross section (pb)
/VVToL2L2Nu-13TeV-amcatnloFXFX-madspin-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-v1/MINIAODSIM	11.95
/VVToL2L2Nu-13TeV-amcatnloFXFX-madspin-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-ext1-v1/MINIAODSIM	11.95
/WZJToLLNu-TuneCUETP8M1-13TeV-amcnlo-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-v1/MINIAODSIM	4.708
/ZZTo4L-13TeV-powheg-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-v1/MINIAODSIM	1.212
Data	
Sample	Luminosity (fb⁻¹)
/SingleMuon/Run2016B-03Feb2017-ver2-v2/MINIAOD	5.788
/SingleMuon/Run2016C-03Feb2017-v1/MINIAOD	2.573
/SingleMuon/Run2016D-03Feb2017-v1/MINIAOD	4.248
/SingleMuon/Run2016E-03Feb2017-v1/MINIAOD	4.009
/SingleMuon/Run2016F-03Feb2017-v1/MINIAOD	3.102
/SingleMuon/Run2016G-03Feb2017-v1/MINIAOD	7.540
/SingleMuon/Run2016H-03Feb2017-ver2-v1/MINIAOD	8.606
/SingleMuon/Run2016H-03Feb2017-ver3-v1/MINIAOD	see above
Additional samples used in Machine Learning analysis	
Sample	Cross section (pb)
/QCD-Pt-20toInf-MuEnrichedPt15-TuneCUETP8M1-13TeV-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-v1/MINIAODSIM	302672.16

/QCD-Pt-20to30-MuEnrichedPt5-TuneCUETP8M1-13TeV-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-v1/MINIAODSIM	2960198
/QCD-Pt-30to50-MuEnrichedPt5-TuneCUETP8M1-13TeV-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-v1/MINIAODSIM	1652471
/QCD-Pt-50to80-MuEnrichedPt5-TuneCUETP8M1-13TeV-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-v1/MINIAODSIM	437504
/QCD-Pt-80to120-MuEnrichedPt5-TuneCUETP8M1-13TeV-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-ext1-v3/MINIAODSIM	106034
/QCD-Pt-80to120-MuEnrichedPt5-TuneCUETP8M1-13TeV-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-v1/MINIAODSIM	106034
/QCD-Pt-120to170-MuEnrichedPt5-TuneCUETP8M1-13TeV-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-v1/MINIAODSIM	25191
/QCD-Pt-170to300-MuEnrichedPt5-TuneCUETP8M1-13TeV-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-v1/MINIAODSIM	8655
/QCD-Pt-170to300-MuEnrichedPt5-TuneCUETP8M1-13TeV-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-ext1-v1/MINIAODSIM	8655
/QCD-Pt-300to470-MuEnrichedPt5-TuneCUETP8M1-13TeV-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-v1/MINIAODSIM	797
/QCD-Pt-300to470-MuEnrichedPt5-TuneCUETP8M1-13TeV-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-ext1-v1/MINIAODSIM	797
/QCD-Pt-300to470-MuEnrichedPt5-TuneCUETP8M1-13TeV-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-ext2-v1/MINIAODSIM	797
/QCD-Pt-470to600-MuEnrichedPt5-TuneCUETP8M1-13TeV-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-v1/MINIAODSIM	79.0
/QCD-Pt-470to600-MuEnrichedPt5-TuneCUETP8M1-13TeV-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-ext1-v1/MINIAODSIM	79.0
/QCD-Pt-470to600-MuEnrichedPt5-TuneCUETP8M1-13TeV-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-ext2-v1/MINIAODSIM	79.0
/GluGluHToTauTau-M125-13TeV-powheg-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-v2/MINIAODSIM	48.58*0.0627
/VBFHToTauTau-M125-13TeV-powheg-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-v2/MINIAODSIM	3.782*0.0627
/VBFHToTauTau-M125-13TeV-powheg-pythia8/RunIISummer16MiniAODv2-PUMoriond17-80X-mcRun2-asymptotic-2016-TrancheIV-v6-v2/MINIAODSIM	3.782*0.0627

Appendix B

Technical Appendix for Chapter 4

Section 4.1	
<i>data samples</i>	All samples (collision data and MC) are produced centrally by the CMS Collaboration and are available via CMS Data Aggregation Service (DAS) system.
<i>runs (high quality) certified</i>	The quality of the data is controlled by Data Quality Monitoring (DQM) system and Data Certification (DC) team [161]. The valid runs are then collected in JSON file. We have used data for runs defined in Cert_271036-284044_13TeV_23Sep2016ReReco_Collisions16_JSON.txt file.
<i>mass of Z boson</i>	The Drell-Yan MC samples are binned in the event invariant mass variable (inv. mass of γ^*/Z'). Inclusive samples used in the analysis are generated with invariant mass above 50 GeV. The low-mass samples are DYJetsToLL_M-10to50-* and high-mass samples are DYJetsToLL_M-150-*. The CMS simulated dataset names are described in [162].
<i>global tags</i>	The following global tags (or GlobalTags) were used: <ul style="list-style-type: none"> • for data samples Run B to G: 80X_dataRun2_2016SeptRepro_v7; • for data samples Run H: 80X_dataRun2_Prompt_v16; • for MC samples: 80X_mcRun2_asymptotic_2016_TrancheIV_v8.
Section 4.2	
<i>defining the Higgs boson candidates</i>	The pairs were created with a dedicated module. We used CandViewShallowCloneCombiner, which is a standard CMSSW utility. It allows to pick an arbitrary particles collection and define simple kinematical conditions on component/combined particle(s).
<i>muons and taus</i>	The muon and taus collections are slimmedMuons and slimmedTaus miniAOD data format [163].
<i>jets collection</i>	slimmedJets
<i>re-apply them</i>	The original corrections are reverted and the new ones are applied. JEC labels are: L1FastJet, L2Relative, L3Absolute and L2L3Residual.
<i>loose jet ID</i>	The usual name for it is loosePFJetID
<i>MET collection</i>	slimmedMET
<i>cuts on discriminator value</i>	We are using two working points corresponding to 80% and 90% efficiencies with cuts: <ul style="list-style-type: none"> • For 80% eff. WP: >0.941 ($0 \leq \eta_e < 0.8$); >0.899 ($0.8 \leq \eta_e < 1.48$); >0.758 ($1.48 \leq \eta_e$). • For 90% eff. WP: >0.837 ($0 \leq \eta_e < 0.8$); >0.715 ($0.8 \leq \eta_e < 1.48$); >0.357 ($1.48 \leq \eta_e$)
<i>generator matching</i>	The generator matching procedure uses a standard collection of pruned generator particles (prunedGenParticles, 50-100/event), which are pointed by the MC matching of the high level objects. The pruned generator particles are equipped with a set of <i>status flags</i> which are robust, generator-independent functions for categorizing mainly final state particles, but also intermediate hadrons/taus.

Section 4.2	
<i>Prompt electron</i>	<code>prunedGenParticle</code> with <code>pdgId</code> equal to -11/11 and status flag <code>isPrompt</code>
<i>Prompt muon</i>	<code>prunedGenParticle</code> with <code>pdgId</code> equal to -13/13 and with status flag <code>isPromptFinalState</code>
<i>Direct prompt electron</i>	<code>prunedGenParticle</code> with <code>pdgId</code> equal to -11/11 and with status flag <code>isDirectPromptTauDecayProduct</code>
<i>Direct prompt muon</i>	<code>prunedGenParticle</code> with <code>pdgId</code> equal to -13/13 and with status flag <code>isDirectPromptTauDecayProductFinalState</code>
<i>status flags</i>	<p>The ones we used in analysis are:</p> <ul style="list-style-type: none"> • <i>isPrompt</i>: not from hadron, muon or tau decay (n.b. in standard samples muons are not allowed to decay at generator level); • <i>isPromptFinalState</i>: final state (status 1) particle satisfying <code>isPrompt</code>; • <i>isDirectPromptTauDecayProduct</i>: is the direct decay product from a tau decay (i.e. no intermediate hadron), where the tau does not come from hadron decay; • <i>isDirectPromptTauDecayProductFinalState</i>: final state (status 1) particle satisfying <code>isDirectPromptTauDecayProduct</code>.
<i>generator matching recipe</i>	<p>First the <i>TauGenJetProducer</i> fed with <code>prunedGenParticles</code> and with <code>includeNeutrinos==False</code> flag is used to generate collection of generator level hadronic taus. Then the three collections of generator particles are produced with <i>TauGenJetDecayModeSelector</i>:</p> <ul style="list-style-type: none"> • category 3 generator hadronic taus: <code>select</code> flag set to "electron"; • category 4 generator hadronic taus: <code>select</code> flag set to "muon"; • category 5 generator hadronic taus: <code>select</code> flag set to "one-Prong0Pi0, oneProng1Pi0, oneProng2Pi0, oneProngOther, threeProng0Pi0, threeProng1Pi0,threeProng1Pi0, rare"; <p>At the end the matched generator category is found by choosing the closest generator particle from between the collections above and the <code>prunedGenParticles</code> for categories 1,2 and 6 (fulfilling conditions from Table 4.2).</p>
<i>Three WPs</i>	The names for those in the muon collection are: <i>isLooseMuon</i> , <i>isMediumMuon</i> and <i>isTightMuon</i> .
<i>medium2016 muon ID</i>	<code>bool goodGlobal = isGlobalMuon and globalTrack.normalizedChi2<3 and combinedQuality.chi2LocalPosition<12 and combinedQuality.trkKink<20;</code> <code>bool medium2016ID = isLooseMuon and innerTrack.validFraction>0.49</code> and <code>segmentCompatibility > (goodGlobal ? 0.303 : 0.451);</code>
<i>reconstructed as 1- or 3-prong tau decay</i>	<code>decay mode finding tau ID > 0.5</code>
<i>MVA-based tau ID discriminant</i>	<code>byIsolationMVArun2v1DBoldDMwLTraw</code>
<i>boson p_T value (full and visible) and jet multiplicity</i>	<p>A boson is reconstructed from its decay products taken from <code>pruned</code> particles collection. The criteria for particle to be a boson constituent are (Particle Data Group IDs (<code>pdgId</code>) can be found, e.g. [164]):</p> <ul style="list-style-type: none"> • For boson full p_T calculation: <code>isDirectHardProcessTauDecayProduct or (fromHardProcessFinalState and (pdgId in {11,12,13,14,16}))</code> • For boson visible p_T calculation: <code>(fromHardProcessFinalState and (pdgId in {11,13})) or (isDirectHardProcessTauDecayProduct and (pdgId not in {12,14,16}))</code> <p>Jets are required to be $\Delta R > 0.5$ away from both pairs of leptons. In selected W+Jets events one of the leptons is faked by a hadronic jet and this jet should be counted as a part of hadronic recoil to the W boson. Therefore, when processing the W+Jets MC sample the number of jets is increased by one.</p>

Section 4.2	
HLT triggers for muon $p_T > 23$.	
HLT trigger	Filters to match
HLT-IsoMu22	<code>hltL3crIsoL1sMu20L1f0L2f10QL3f22QL3trkIsoFiltered0p09</code>
HLT-IsoTkMu22	<code>hltL3fL1sMu20L1f0Tkf22QL3trkIsoFiltered0p09</code>
HLT-IsoMu22-eta2p1	<code>hltL3crIsoL1sSingleMu20erL1f0L2f10QL3f22QL3trkIsoFiltered0p09</code>
HLT-IsoTkMu22-eta2p1	<code>hltL3fL1sMu20erL1f0Tkf22QL3trkIsoFiltered0p09</code>
HLT triggers for muon $20. < p_T \leq 23$.	
HLT trigger	Filters to match
HLT-IsoMu19-eta2p1-LooseIsoPFTau20	<code>hltL3crIsoL1sMu18erTauJet20erL1f0L2f10QL3f19QL3trkIsoFiltered0p09</code> or <code>hltOverlapFilterIsoMu19LooseIsoPFTau20</code> <code>hltPFTau20TrackLooseIsoAgainstMuon</code> or <code>hltOverlapFilterIsoMu19LooseIsoPFTau20</code>
HLT-IsoMu19-eta2p1-LooseIsoPFTau20-SingleL1	<code>hltL3crIsoL1sSingleMu18erlorSingleMu20erL1f0L2f10QL3f19QL3trkIsoFiltered0p09</code> or <code>hltOverlapFilterSingleIsoMu19LooseIsoPFTau20</code> <code>hltPFTau20TrackLooseIsoAgainstMuon</code> or <code>hltOverlapFilterSingleIsoMu21LooseIsoPFTau20</code>

Table B.1: HLT triggers with filters to match. Muon filters to match are colored in purple and tau filters in blue.

Section 4.3	
<i>tight WP tau MVA isolation ID</i>	<code>byTightIsolationMVArun2v1DBoldDMwLT</code>
Section 4.4	
<i>pileup distributions</i>	They are calculated for "Moriond17" data processing and saved in 800 bins histograms (for data and MC) in $0 < \text{NPU} < 80$ range (NPU is a number of primary vertexes).
<i>loose anti-electron</i>	<code>againstElectronMVA Loose</code>
<i>tight anti-muon</i>	<code>againstMuonTight3</code>
<i>weights in the analysis</i>	Muon tracking weight is <code>m_trk_ratio</code> and muon trigger weights are <code>m_trgMu22OR_eta2p1_desy_ratio</code> ($p_T^\mu > 23$.) or <code>m_trgMu19leg_eta2p1_desy_ratio</code> ($20. < p_T^\mu \leq 23$.) from <code>htt_scalefactors_v16_5.root</code> file. The cross trigger weight is <code>t_genuine_TightIso_mt_ratio</code> (for $20. < p_T^\mu \leq 23$.) from <code>htt_scalefactors_sm_moriond_v2.root</code> file. Both files are from CMS Correction Workspace repository.
<i>Lepton ID scale factors</i>	The values of corrections applied are taken out of the CMS Correction Workspace repository. The muon ID weight used is taken as <code>m_idiso0p15_desy_ratio</code> value from <code>htt_scalefactors_v16_5.root</code> file.
<i>Z boson is reconstructed</i>	The reconstruction is the same as for <i>boson p_T value (full and visible) and jet multiplicity</i> row in the previous table.
<i>weights are presented in Fig. 4.2</i>	These are taken from <code>zpt_weight_nom</code> variable from <code>htt_scalefactors_v16_5.root</code> file in CMS Correction Workspace repository.
Section 4.5	
<i>partons</i>	is a number of particles in the event with <code>status=1</code> and <code>pdgID ∈ {21, 1, 2, 3, 4, 5}</code> . Particles are taken out of the <code>externalLHEProducer</code> collection.
<i>leptons</i>	Leptons are taken out of the <code>externalLHEProducer</code> collection here as well. If there is a different number of leptons then two than the invariant mass is set to zero.

Appendix C

Control plots

The reference plots from official $H \rightarrow \tau_\mu \tau_h$ CMS analysis [82] and after the global fit are given (for all analysis categories) in Fig. C.1, C.2 and C.3

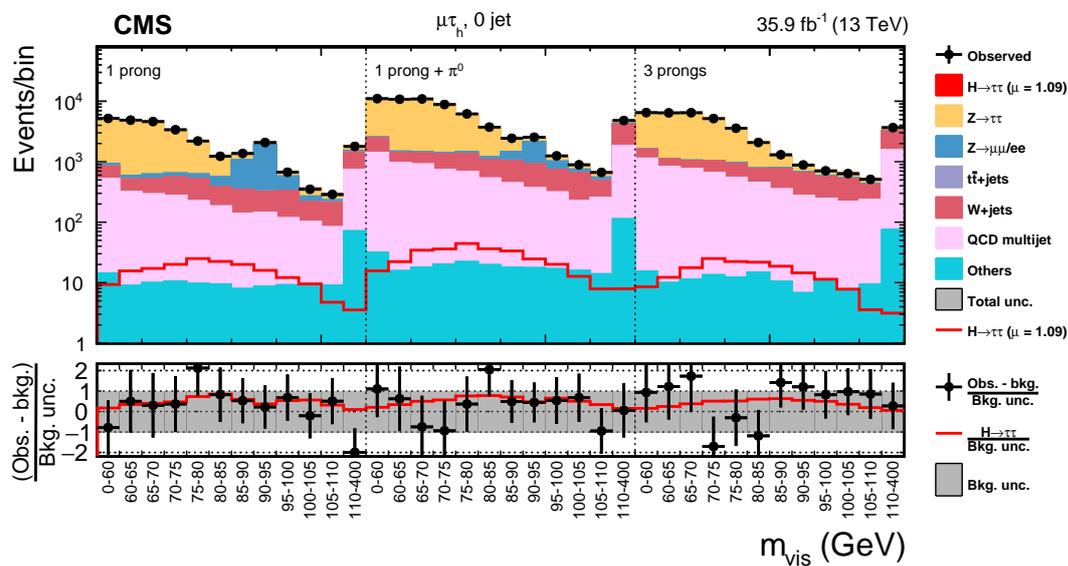


Figure C.1: Observed and predicted 2D distributions in the 0jet category of the $\mu\tau_h$ decay channel. The normalization of the predicted background distributions corresponds to the result of the global fit. The signal distribution is normalized to its best fit signal strength. The background histograms are stacked. The "Others" background contribution includes events from diboson and single top quark production, as well as Higgs boson decays to a pair of W bosons. The background uncertainty band accounts for all sources of background uncertainty, systematic as well as statistical, after the global fit. The signal is shown both as a stacked filled histogram and an open overlaid histogram. [82]

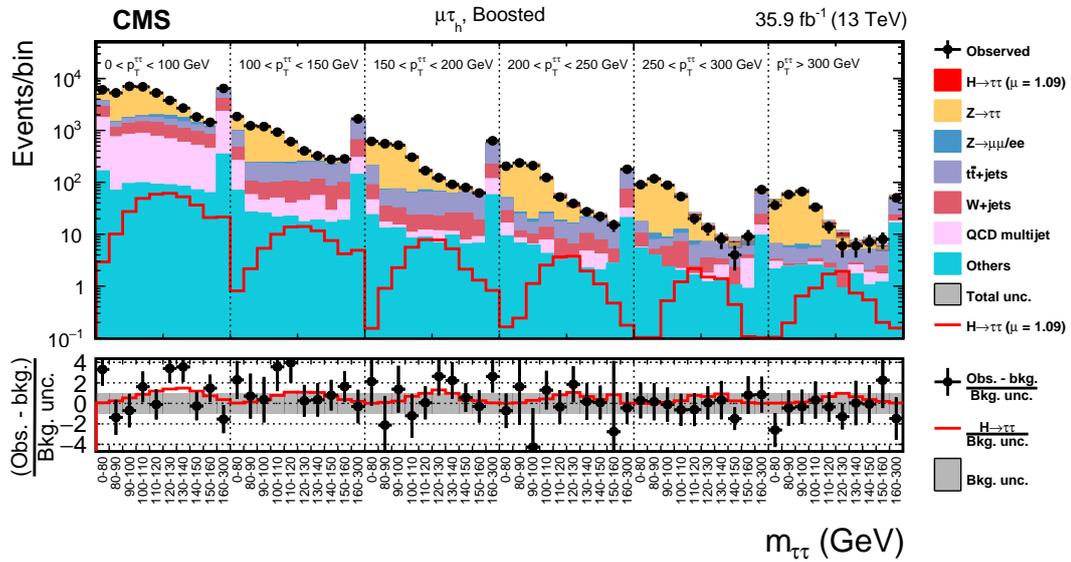


Figure C.2: Observed and predicted 2D distributions in the boosted category of the $\mu\tau_h$ decay channel. The description of the histograms is the same as in Fig. C.1 [82]

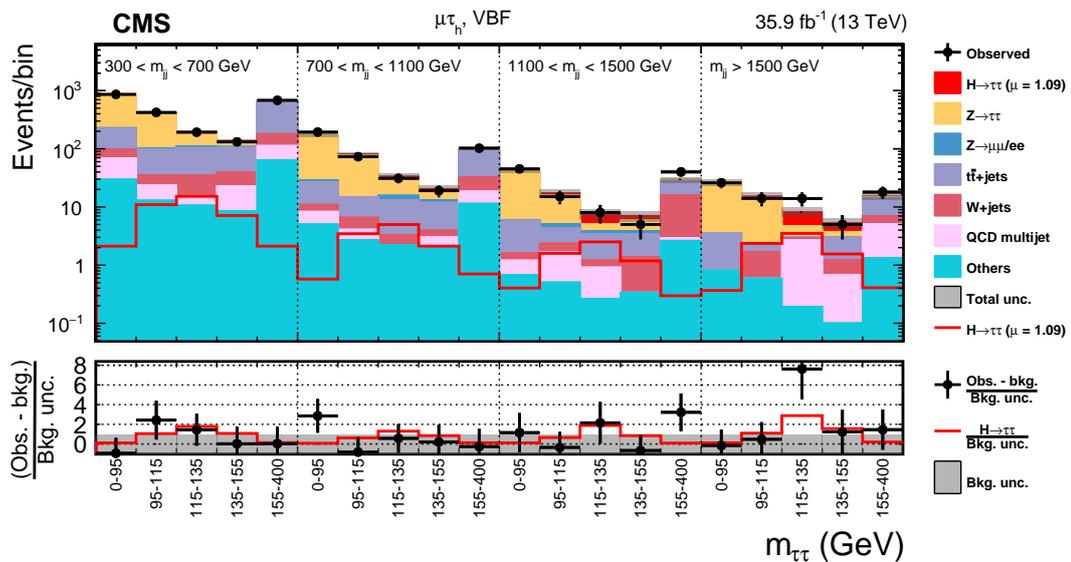


Figure C.3: Observed and predicted 2D distributions in the vbf category of the $\mu\tau_h$ decay channel. The description of the histograms is the same as in Fig. C.1. [82]

Appendix D

Features list

Name	Meaning
svfmass	pair SVFit mass
svfpt	pair SVFit transverse momentum
svfmassErr	error of SVFit mass
svftransverseMass	SVFit transverse mass
svftransverseMassErr	error of SVFit transverse mass
pth	Higgs boson visible transverse momentum (lepton pair + MET)
ptvis	lepton pair transverse momentum
m_vis	lepton pair visible mass
mt	lepton pair transverse mass
mjj	mass of the pair of leading jets (Nan if $n_{jets} < 0$)
njets	number of jets
njetsingap	number of jets between (in η) the pair of leading jets
jjdr	ΔR for two leading jets
mu_pt	muon transverse momentum
mu_phi	muon azimuthal angle
mu_eta	muon pseudorapidity
mu_mt	muon transverse mass
mu_d0	muon displacement from PV in $x - y$ plane
mu_dz	muon displacement from PV in z -axis (beam-axis)
mu_isLooseMuon	muon loose ID
mu_isTightMuon	muon tight ID
mu_isHighPtMuon	muon high- p_T ID
mu_isMediumMuon	muon medium ID
mu_iso	muon isolation
mu_charge	muon charge
tau_pt	tau transverse momentum
tau_phi	tau pseudorapidity
tau_eta	tau azimuthal angle
tau_mass	tau mass
tau_energy	tau energy
tau_decayMode	tau decay mode
tau_charge	tau charge
tau_mt	tau transverse mass
tau_d0	tau displacement from PV in $x - y$ plane
tau_dz	tau displacement from PV in x -axis
tauID_1	byCombinedIsolationDeltaBetaCorrRaw3Hits tau ID

tauID_2	chargedIsoPtSum tau ID
tauID_3	neutralIsoPtSum tau ID
tauID_4	puCorrPtSum tau ID
tauID_5	footprintCorrection tau ID
tauID_6	photonPtSumOutsideSignalCone tau ID
tauID_7	againstElectronMVA6Raw tau ID
tauID_8	againstElectronMVA6category tau ID
tauID_9	byIsolationMVArun2v1DBdR03oldDMwLTraw tau ID
tauID_10	byIsolationMVArun2v1DBoldDMwLTraw tau ID
tauID_11	byIsolationMVArun2v1PWdR03oldDMwLTraw tau ID
tauID_12	byIsolationMVArun2v1PWoldDMwLTraw tau ID
met_pt	MET transverse momentum
met_phi	MET azimuthal angle
met_eta	MET pseudorapidity
met_px	MET p_x
met_py	MET p_y
bptag_1	leading jet pfJetBProbabilityBJetTags discriminator
bptag_2	sub-leading jet pfJetBProbabilityBJetTags discriminator
csvtag_1	leading jet pfCombinedInclusiveSecondaryVertexV2BJetTags discriminator
csvtag_2	sub-leading jet pfCombinedInclusiveSecondaryVertexV2BJetTags discriminator
jet_energy_1	leading jet energy
jet_energy_2	sub-leading jet energy
jet_eta_1	leading jet pseudorapidity
jet_eta_2	sub-leading jet pseudorapidity
jet_phi_1	leading jet azimuthal angle
jet_phi_2	sub-leading jet azimuthal angle
jet_pt_1	leading jet transverse momentum
jet_pt_2	sub-leading jet transverse momentum
delta_mu_tau_u	deltaR(mu,tau)
delta_mu_met	deltaR(mu,met)
delta_tau_met	deltaR(tau,met)
delta_mu_tau_met	deltaR(mu,tau,met)

Table D.1: Features used in Machine Learning based analysis

Appendix E

Statistics

In HEP Higgs analyses a usual way of dealing with significance computation starts with defining the a hypotheses [165]:

- *null hypothesis* H_0 , where we assume the observed (discriminating variables) distributions stems from existence of only background;
- *alternative hypotheses* H_1 , where there exists an addition of a signal events contributing to the observed phenomena.

The likelihood-ratio based hypothesis test [166] between those two hypotheses can be performed. The Neyman–Pearson lemma states that the most powerful test statistics in such scenario is:

$$Q(\vec{x}) = \frac{L(\vec{x}|H_1)}{L(\vec{x}|H_0)}, \quad (\text{E.1})$$

where L is a probability density function for the distribution of (possibly multi-dimensional) discriminating variable $\vec{x} = x_j$. In the case of number-counting experiment, the likelihood of getting n events is given by the Poisson distribution. If additionally the distribution of a discriminating observable $f(\vec{x})$ is known (and determined for every event j), then total likelihood can be factorized [167]:

$$Q(\vec{x}) = \frac{\text{Pois}(n|s+b) \prod_{j=1}^n f_{s+b}(x_j)}{\text{Pois}(n|b) \prod_{j=1}^n f_b(x_j)}. \quad (\text{E.2})$$

Assuming no interference between signal "s" and background "b" events, the distribution for alternative hypothesis can be written down as: $f_{s+b}(x) = [sf_s(x) + bf_b(x)]/(s+b)$.

Finally, Eq. E.2 can be further extended, taking into account different contributions N ("channels"), e.g. different event topology selections, data taken at different center-of-mass energies or data collected in different experiments:

$$Q(\vec{x}) = \frac{\prod_{k=1}^N \text{Pois}(n|s+b, k) \prod_{j=1}^n f_{s+b}(x_j, k)}{\prod_{k=1}^N \text{Pois}(n|b, k) \prod_{j=1}^n f_b(x_j, k)}.$$

Substituting the actual functions we have:

$$Q(\vec{x}) = \frac{\prod_{k=1}^N (\exp[-s_k - b_k] (s_k + b_k)^{n_k} / n_k!) \prod_{j=1}^n [s_k f_s(x_j, k) + b_k f_b(x_j, k)] / (s_k + b_k)}{\prod_{k=1}^N (\exp[-b_k] (b_k)^{n_k} / n_k!) \prod_{j=1}^n f_b(x_j, k)},$$

where s and b are the number of signal and background events. The above equation can be simplified into:

$$Q(\vec{x}) = \frac{\prod_{k=1}^N (\exp[-s_k]) \prod_{j=1}^n [s_k f_s(x_j, k) + b_k f_b(x_j, k)]}{\prod_{k=1}^N \prod_{j=1}^n b_k f_b(x_j, k)} = \prod_{k=1}^N \exp[-s_k] \prod_{j=1}^n \left[1 + \frac{s_k f_s(x_j, k)}{b_k f_b(x_j, k)} \right]. \quad (\text{E.3})$$

The usage of $t(\vec{x}) = -2 \log Q(\vec{x})$ statistics is convenient, since its value grows together with incompatibility between data and (\vec{x}) [168]. Substituting in Eq. E.3 we have:

$$t(\vec{x}) = -2 \left\{ -\sum_{k=1}^N s_k + \sum_{j=1}^n \log \left[1 + \frac{s_k f_s(x_j, k)}{b_k f_b(x_j, k)} \right] \right\}$$

Let us suppose we only have one channel (e.g. 0jet). The expression above can be simplified into:

$$t(\vec{x}) = -2 \left\{ -s + \sum_{j=1}^n \log \left[1 + \frac{s \cdot f_s(x_j)}{b \cdot f_b(x_j)} \right] \right\}$$

Furthermore, let the $f_{s/b}$ describe the discriminating variable (e.g. the score) p.d.f. that is given in a form of two-bins histogram (for score $w \leq 0.5$ and $w > 0.5$). Then we have:

$$t(\vec{x}) = -2 \left\{ -s + \sum_{\text{events}(w \leq 0.5)} \log \left[1 + \frac{s \cdot f_{s,1}}{b \cdot f_{b,1}} \right] + \sum_{\text{events}(w > 0.5)} \log \left[1 + \frac{s \cdot f_{s,2}}{b \cdot f_{b,2}} \right] \right\}$$

From now on we will follow the convention adopted in [168], and define the "Asimov data set", as the one for which the estimators evaluation for all parameters end up in obtaining their true values. Thanks to doing that it is possible to estimate (the median) significance using a single representative (MC) dataset, instead of an ensemble of simulated datasets. For the Asimov dataset, the observed number of events n is equal to the simulated (expected) sum of the signal and background, namely

$$n = s + b.$$

Further, since $f_{s/b}$ is p.d.f. (is normalized into one), we have $s \cdot f_{s,1} = s_1$, where s_1 is a number of events for which $w \leq 0.5$. This reasoning works for all $f_{s/b}$, therefore we have:

$$t(\vec{x}) = -2 \left\{ -s + (s_1 + b_1) \log \left[1 + \frac{s_1}{b_1} \right] + (s_2 + b_2) \log \left[1 + \frac{s_2}{b_2} \right] \right\}$$

The generalization of this equation for arbitrary number of $f_{s/b}$ distribution bins can be written as:

$$t(\vec{x}) = -2 \left\{ -s + \sum_{\text{x bins}} (s_i + b_i) \log \left[1 + \frac{s_i}{b_i} \right] \right\} = -2 \sum_{\text{x bins}} \left\{ -s_i + (s_i + b_i) \log \left[1 + \frac{s_i}{b_i} \right] \right\}$$

The Taylor expansion of this formula for $y = \frac{s}{b}$ around $y = 0$ gives:

$$\begin{aligned}
 t(\bar{x}) &= - \sum_{\text{x bins}} 2b_i \{-y_i + (y_i + 1) \log [1 + y_i]\} \sim - \sum_{\text{x bins}} 2b_i \left\{ 0 + \frac{y_i}{1!} * 0 + \frac{y_i^2}{2!} * 1 \right\} \\
 &= \sum_{\text{x bins}} \left\{ \frac{s_i^2}{b_i} \right\}
 \end{aligned}$$

Finally for a sufficiently large b , the significance S can be approximated by the formula [168]:

$$S = \sqrt{t},$$

hence:

$$S = \sqrt{2 \left\{ \sum_{\text{x bins}} -s_i + (s_i + b_i) \log \left[1 + \frac{s_i}{b_i} \right] \right\}},$$

or in the Taylor-expanded form:

$$S = \sqrt{\sum_{\text{x bins}} \left\{ \frac{s_i^2}{b_i} \right\}}$$