

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

**PROSPECTS OF NONRESONANT HIGGS BOSON PAIR PRODUCTION
MEASUREMENT IN THE $WW\gamma\gamma$ CHANNEL AT THE HL-LHC
WITH THE PHASE-II CMS DETECTOR**

M.Sc. THESIS

Ahmet Oğuz GÜZEL

Department of Physics Engineering

Physics Engineering Programme

JUNE 2022

CERN-THESIS-2022-059
02/06/2022



ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

**PROSPECTS OF NONRESONANT HIGGS BOSON PAIR PRODUCTION
MEASUREMENT IN THE $WW\gamma\gamma$ CHANNEL AT THE HL-LHC
WITH THE PHASE-II CMS DETECTOR**

M.Sc. THESIS

**Ahmet Oğuz GÜZEL
(509181120)**

Department of Physics Engineering

Physics Engineering Programme

Thesis Advisor: Prof. Dr. M. Altan ÇAKIR

JUNE 2022

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

**STANDART MODEL HIGGS BOZONU ÇİFTİ ÜRETİMİNİN
 $WW\gamma\gamma$ KANALINDA CMS FAZ II DEDEKTÖRÜ İLE
HL-LHC KOŞULLARINDA ARAŞTIRILMASI**

YÜKSEK LİSANS TEZİ

**Ahmet Oğuz GÜZEL
(509181120)**

Fizik Mühendisliği Anabilim Dalı

Fizik Mühendisliği Programı

Tez Danışmanı: Prof. Dr. M. Altan ÇAKIR

HAZİRAN 2022

Ahmet Oğuz GÜZEL, a M.Sc. student of ITU Graduate School student ID 509181120, successfully defended the thesis entitled “PROSPECTS OF NONRESONANT HIGGS BOSON PAIR PRODUCTION MEASUREMENT IN THE $WW\gamma\gamma$ CHANNEL AT THE HL-LHC WITH THE PHASE-II CMS DETECTOR”, which he prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor : **Prof. Dr. M. Altan ÇAKIR**
Istanbul Technical University



Jury Members : **Prof. Dr. Cenap Şahabettin ÖZBEN**
Istanbul Technical University

Assoc. Prof. Dr. Bora IŞILDAK
Ozyegin University



Date of Submission : 24 May 2022

Date of Defense : 02 June 2022

To my family,

FOREWORD

I would like to express my deep gratitude to my supervisor Prof. Dr. Altan akır for his excellent guidance in my pursuit in particle physics and in life. He always tried his best to afford what I needed and showed his support in everything I have done. I would also like to thank the Upgrade Performance Studies Group (UPSG) of CMS for their kind behaviour towards anyone. Each colleague I met at UPSG has been very caring and very intellectual, with most of them I became friends. Last but not least, I would like to thank my family, friends and beloved one for all the support they gave me so far. I couldn't have achieved this without them.

Finally, I acknowledge financial support from Turkish Atomic Energy Authority (TAEK) under the project number 2018TAEK-A5.H6.F2-21.

June 2022

Ahmet Oğuz GÜZEL

TABLE OF CONTENTS

	<u>Page</u>
FOREWORD.....	ix
TABLE OF CONTENTS.....	xi
ABBREVIATIONS	xiii
SYMBOLS	xv
LIST OF TABLES	xvii
LIST OF FIGURES	xix
SUMMARY	xxv
ÖZET	xxvii
1. INTRODUCTION TO THE STANDARD MODEL	1
1.1 The Standard Model Lagrangian and the Higgs Mechanism	5
1.1.1 Quantum chromodynamics	6
1.1.2 Electroweak theory	7
1.1.3 Higgs mechanism.....	10
1.2 The Higgs Boson Phenomenology and Experimental Status	14
1.2.1 Higgs boson pair production.....	20
1.2.2 BSM models	24
2. THE LHC AND THE CMS EXPERIMENT	27
2.1 The Large Hadron Collider.....	27
2.1.1 Operation of the LHC	27
2.1.2 The accelerator complex	29
2.1.3 Design and specifications.....	31
2.2 The Compact Muon Solenoid Experiment	33
2.2.1 Structure and the CMS coordinate system.....	34
2.2.2 Solenoid magnet.....	35
2.2.3 Inner tracking system.....	36
2.2.4 Electromagnetic calorimeter	38
2.2.5 Hadron calorimeter	41
2.2.6 Muon system.....	42
2.2.7 Trigger system and storage	46
2.3 The HL-LHC Programme and the Phase II Upgrade of the CMS Detector.....	49
3. DATA ANALYSIS	51
3.1 Simulation and Reconstruction of Proton-Proton Collisions	51
3.1.1 Event generation	51
3.1.2 Detector simulation.....	54
3.1.3 Particle flow algorithm.....	56
3.1.4 Reconstruction of physics objects.....	57
3.1.5 High level corrections and validation	58
3.2 Higgs Boson Pair Analysis	59

3.2.1 Signal and background samples.....	59
3.2.2 Object selection.....	62
3.2.3 Event selection and categorisation.....	63
3.2.4 Systematic uncertainties.....	68
4. RESULTS & CONCLUSION.....	71
REFERENCES.....	73
APPENDICES.....	83
APPENDIX A.1	85
APPENDIX A.2	105
APPENDIX A.3	110
APPENDIX A.4	117
APPENDIX A.5	118
APPENDIX A.6	125
APPENDIX A.7	127
CURRICULUM VITAE.....	129

ABBREVIATIONS

ALICE	: A Large Ion Collider Experiment
ATLAS	: A Toroidal LHC Apparatus
BEH	: Brout-Englert-Higgs mechanism
BSM	: Beyond the Standard Model
CERN	: European Organisation for Nuclear Research
CMS	: Compact Muon Solenoid
CSC	: Cathode Strip Chamber
DNN	: Deep Neural Network
DT	: Drift Tube
ECAL	: Electromagnetic Calorimeter
EWSB	: Electroweak Symmetry Breaking
fb	: Femtobarn
GeV	: Giga electron volts
ggF	: Gluon Fusion
HCAL	: Hadronic Calorimeter
HH	: Double Higgs boson or Higgs boson pair
HL-LHC	: The High-Luminosity Large Hadron Collider
LEP	: Large Electron Positron Collider
LHC	: Large Hadron Collider
LHCb	: Large Hadron Collider Beauty Experiment
MeV	: Mega electron volts
NLO	: Next-to-Leading Order
NNLO	: Next-to-Next-to-Leading Order
pb	: Picobarn
PDF	: Parton Distribution Function
pp	: Proton-proton
QCD	: Quantum Chromodynamics
QED	: Quantum Electrodynamics
QFT	: Quantum Field Theory
RPC	: Resistive Plate Chambers
SSB	: Spontaneous symmetry breaking
SM	: Standard Model
SU	: Special unitary group
TeV	: Tera electron volts
ttH	: Top quark associated production
VBF	: Vector Boson Fusion
VH	: Vector boson associated production
2HDM	: Two-Higgs-Doublet Model

SYMBOLS

D_μ	: Covariant derivative
E_T	: Transverse Energy
K	: Kelvin
\mathcal{L}	: Lagrangian density
p_T	: Transverse Momentum
Q	: Electric charge
\sqrt{s}	: Centre-of-mass-energy
T_H	: Hagedorn temperature
T_i	: Weak isospin
Y_W	: Hypercharge
η	: Pseudo-rapidity
γ	: Photon
θ_w	: Weak mixing angle

LIST OF TABLES

	<u>Page</u>
Table 1.1 : Fermion fields under $SU(2)_L$ group representation.	10
Table 1.2 : Free parameters of the Standard Model.	15
Table 1.3 : Cross section (in pb) of the Higgs boson production at different centre-of-mass (\sqrt{s}) energies for $m_H = 125$ GeV. The theoretical uncertainties can be found in the reference.	16
Table 1.4 : The branching ratios and the relative uncertainty for a SM Higgs boson with $m_H = 125$ GeV.	19
Table 1.5 : Cross section (in pb) of the Higgs boson pair production at different centre-of-mass (\sqrt{s}) energies for $m_H = 125$ GeV. The theoretical uncertainties including the QCD factorisation and renormalisation scale (α_s) uncertainties, Parton Distribution Function (PDF) uncertainty and effects from the top quark mass at NNLO can be found in the reference.	22
Table 3.1 : Object selections.	63
Table 3.2 : Semi-leptonic final state DNN score categories.	65
Table 3.3 : Post-selections of the Fully-leptonic channel of $HH \rightarrow WW\gamma\gamma$	67
Table 3.4 : Single τ final state DNN categories.	67
Table 3.5 : Experimental uncertainties considered in this study. The values are recommended in the Yellow Report for HL-LHC studies [1].	70
Table 3.6 : Theoretical uncertainties considered on ggHH signal and single Higgs processes.	70
Table 4.1 : Significance numbers extracted in each category of one lepton final state, two leptonic final state and their combination.	71
Table 4.2 : Significance numbers extracted in each category of one tau final state, two tau's final state and their combination.	72
Table 4.3 : Full Phase-II results of $WW\gamma\gamma$ and $\tau\tau\gamma\gamma$ processes with combination.	72
Table A.1 : MC samples used in the analysis.	106
Table A.2 : MC samples used in the analysis (cont'd).	107
Table A.3 : Cut-flow report showing number of events, before selections, in the semi-leptonic channel and in its categories. Percentages in brackets show the total selection efficiency that is number of events of the final state divided by the total number of events in the samples $\times 100$	108
Table A.4 : Cut-flow report showing number of events, before selections, in the semi-leptonic channel and in its categories. Percentages in brackets show the total selection efficiency that is number of events of the final state divided by the total number of events in the samples $\times 100$. (cont'd).	109
Table A.5 : Input variables used to train semi-leptonic final state DNN.	110

Table A.6 :	Hyper-parameter settings for the DNN performed for semi-leptonic channel.	116
Table A.7 :	Cut-flow report showing number of events, before selections, in the semi-leptonic channel and in its categories. Percentages in brackets show the total selection efficiency.	117
Table A.8 :	Input variables used to train single τ final state DNN.	118
Table A.9 :	Cut-flow report showing number of events, before selections, in the single τ final state and in its categories. Percentages in brackets show the total selection efficiency.	123
Table A.10 :	Cut-flow report showing number of events, before selections, in the single τ final state and in its categories (cont'd).	124
Table A.11 :	Cut-flow report showing number of events, before selections and in the double τ final state of $HH \rightarrow \tau\tau\gamma\gamma$	126

LIST OF FIGURES

	<u>Page</u>
Figure 1.1 : The elementary particles of the Standard Model. Fermions and bosons are grouped in columns where the quarks, leptons, gauge bosons and scalar Higgs boson are shown in different colours. The three generations of matter are indicated with roman numerals. The mass, charge and spin values corresponding to each particle are indicated on the upper left of each box. The vector bosons are connected in faint yellow areas to the particles they interact with.....	3
Figure 1.2 : The Higgs potential V in the case that $\mu^2 > 0$. Choosing any point at the bottom of the potential breaks spontaneously the rotational $U(1)$ symmetry.....	12
Figure 1.3a : m_{4l}	16
Figure 1.3b : $m_{\gamma\gamma}$	16
Figure 1.3 : Four-lepton mass distribution, m_{4l} obtained from the data collected at the CMS Detector on the left, and diphoton invariant mass distribution obtained from the data collected at the ATLAS Detector on the right, both at $\sqrt{s} = 13$ TeV in Run II.	16
Figure 1.4 : Summary of the CMS and ATLAS Higgs mass measurements in the $\gamma\gamma$ and ZZ channels in Run I and Run II.....	17
Figure 1.5a : ggF	18
Figure 1.5b : VBF	18
Figure 1.5c : VH	18
Figure 1.5d : $t\bar{t}H$	18
Figure 1.5 : Leading order Feynman diagrams of Higgs boson production at the LHC. The gluon fusion (ggF), vector boson fusion (VBF), Higgs strahlung (VH) and associated production with top quarks ($t\bar{t}H$) are shown. In the diagrams, q denotes any quark, t denotes the top quark, V denotes any of the Z or W^\pm bosons.	18
Figure 1.6a : σ_H	19
Figure 1.6b : m_H	19
Figure 1.6 : Higgs boson production cross section as a function of different production mechanisms (left). Branching fractions the Higgs boson as a function of m_H . The theoretical uncertainties are indicated as bands in both plots.	19
Figure 1.7 : Branching ratio of the Higgs boson decaying into various final states as a function of the Higgs boson mass.....	20
Figure 1.8a : Triangle type	21
Figure 1.8b : Box type	21

Figure 1.9	: Invariant mass distribution of Higgs boson pairs at leading order for the triangle and box-type gluon fusion production mechanism and their interference.	21
Figure 1.10a:	$c_{2\nu}$	22
Figure 1.10b:	c_ν	22
Figure 1.10c:	κ_λ	22
Figure 1.11	: Total cross section of double Higgs production mechanisms as a function of the centre-of-mass energy with $m_H = 125$ GeV within the SM. The gluon fusion remains the dominant production mode at high energies. The VBF continues to be the second dominant production mode while $t\bar{t}HH$ contribution starts to be in the order of VBF at about 100 TeV. The width of the lines shows indicate the total uncertainties coming from the scale and PDF + α_S uncertainties.	23
Figure 1.12	: Branching ratio of major double Higgs boson decay channels assuming a Standard Model Higgs boson. From top to bottom and left to right, less possible decay mode combinations are depicted. The logarithmic scale bar on the left indicates the branching ratio along with colours.	24
Figure 1.13	: The 95% CL upper limits on the signal strength $\mu = \sigma_{HH}/\sigma_{HH}^{SM}$ for $b\bar{b}(VV/b\bar{b}/\tau^+\tau^-/\gamma\gamma)$ channels along with with combined result. The inner green bands show 68% and the outer yellow bands show the 95% CL upper limits on μ under the background-only hypothesis.	25
Figure 1.14	: Feynman diagrams that contribute to HH non-resonant production at LO with pure BSM effects.	26
Figure 2.1	: A detailed schedule of LHC and HL-LHC showing the integrated luminosity and the beam energy corresponding to each period.	28
Figure 2.2	: The LHC accelerator complex. The acceleration of the protons start in the LINAC2 and ends in LHC through Booster, PS and SPS.	30
Figure 2.3	: Cross sections of Standard Model processes as function of collider energy at pp collisions.	32
Figure 2.4	: Total integrated luminosity delivered by the LHC to the CMS detector for pp collisions at Run I-II.	33
Figure 2.5	: Three dimensional view of the CMS detector showing each component. A human shape is positioned on lower right to compare its size. It is sometimes called as the cylindrical onion due to its shape and concentric layers of components.	34
Figure 2.6	: The coordinate system of the CMS detector. IP denotes the interaction point and the momentum of a particle is shown by \vec{p} in red arrow. Pseudo-rapidity values are shown by η	35
Figure 2.7	: Schematic view of a transverse slice of the barrel region of the CMS detector. Particles are shown in their trajectories and signatures in the sub-detectors. Photons and electrons create showers in the ECAL but photons don't leave traces in the tracking system which allows to distinguish these two types. The same recognition can be made between neutral and charged hadrons.	36

Figure 2.8	: Schematic cross section of the tracker in the r-z plane with one half shown since the tracker is symmetric about the $r = 0$. The star indicates the centre of the tracker and the expected collision point. The detector modules are kept in green dashed lines, namely; Tracker inner barrel (TIB) and Tracker inner disks (TID) surrounded by the Tracker outer barrel (TOB) and tracker endcaps (TEC) on the sides. Thin black lines show the strip tracker modules that create 2-D hits, while thick blue lines show the one allowing providing 3-D hit positions which is actually two back-to-back strip modules. The pixel modules provide 3-D hits and they are shown in red curly lines. Each module in a given layer is shifted slightly cover all the gaps providing best acceptance.	37
Figure 2.9	: The material budget contribution from sub-systems as a function of η , showing the tracker sub-detectors, beam pipe and the super tube that surrounds the tracker.	39
Figure 2.10	: Schematic view showing one quadrant of the calorimetry and tracking system.	40
Figure 2.11	: Schematic view of the CMS detector in the y-z plane with the HCAL sub-detectors labelled.	42
Figure 2.12	: R-z cross section of a quadrant of the CMS detector. The bottom left corner indicates the interaction point. Various muon stations and the steel flux-return disks are shown. The drift tube stations (DTs) are labelled MB (Muon Barrel) and the cathod strip chambers (CSCs) are labelled ME (Muon Endcap). The resistive plate chambers in the barrel and in the endcaps are shown RB and RE, respectively.	43
Figure 2.13	: Schematic view of a drift cell showing the electric field lines in the gas volume as well as the isochrones (contours of equal drift times) and the drift lines.	44
Figure 2.14a:	CSC	45
Figure 2.14b:	CMS Endcap	45
Figure 2.14	: Schematic view of a CSC (a) and a photo of the ME+2 disk with its stations of CSCs.	45
Figure 2.15	: Schematic view of a dual RPC detector.	46
Figure 3.1	: Graphical representation of a pp collision event. The proton beams come from the either sides. The red diagram shows the hard interaction and the consequent decay of the products. A secondary interaction before the final state partons hadronise, is shown in purple. The hadronisation is indicated in green.	53
Figure 3.2	: Energy resolution of photons and electrons as a function of energy for a CMS detector configuration. The energy resolution formula of the CMS ECAL, and the samples used for the study along with their generator are given. The CMS electron distribution is taken from . The electron and photon resolutions are identical at high energies since both objects are dominated by ECAL, however at low energies, the electron resolution differs widely because of the finer tracker resolution.	60
Figure 3.3	: Analysis flow chart.	61

Figure 3.4	: The flat tree format of the DELPHES simulated samples. The reconstructed objects include photons, electrons, muons, taus, jets and E_T^{miss} . Each object has the multiplicity, mass, p_T , η , ϕ values except E_T^{miss} which do not have mass and η values. Identification and isolation variables are available for most of the objects and b-originated jets information is available for jets. There is also the possibility to work on the generator level with generated particles (genpart), jets (genjet) and E_T^{miss} (genmet) collections. Genparts collection possess mother (m1, m2) and daughter (d1, d2) particles information along with particle status codes (PID). Jets and E_T^{miss} have different versions produced used different algorithm, i.e. pile-up per particle identification (PUPPI), charged hadrons subtracted (CHS) or particle flow (PF). Additionally, number of events, vertex information, generator level weight are among the accessible information.	62
Figure 3.5	: DNN score distribution in the semi-leptonic final state.	65
Figure 3.6a	: Semi-leptonic final state	66
Figure 3.6b	: DNN categories	66
Figure 3.7	: $m_{\gamma\gamma}$ distribution in the fully-leptonic final state.	67
Figure 3.8	: DNN score distribution in the single τ final state.	68
Figure 3.9	: $m_{\gamma\gamma}$ distributions in the DNN categories of single τ final state and double τ final state.	69
Figure A.1a	: Leading Photon $p_T/m_{\gamma\gamma}$	111
Figure A.1b	: Leading Photon $E/m_{\gamma\gamma}$	111
Figure A.1c	: Leading Photon η	111
Figure A.1d	: Leading Photon ϕ	111
Figure A.1e	: Sub-leading Photon $p_T/m_{\gamma\gamma}$	111
Figure A.1f	: Sub-leading Photon $E/m_{\gamma\gamma}$	111
Figure A.1	: DNN input distributions for the semi-leptonic final state.	111
Figure A.2a	: Sub-leading Photon η	112
Figure A.2b	: Sub-leading Photon ϕ	112
Figure A.2c	: Leading Electron p_T	112
Figure A.2d	: Leading Electron η	112
Figure A.2e	: Leading Electron ϕ	112
Figure A.2f	: Leading Electron energy	112
Figure A.2	: DNN input distributions for the semi-leptonic final state (continued).	112
Figure A.3a	: Leading Muon p_T	113
Figure A.3b	: Leading Muon η	113
Figure A.3c	: Leading Muon ϕ	113
Figure A.3d	: Leading Muon energy	113
Figure A.3e	: Jet multiplicity	113
Figure A.3f	: E_T^{miss}	113
Figure A.3	: DNN input distributions for the semi-leptonic final state (continued).	113
Figure A.4a	: Leading jet p_T	114
Figure A.4b	: Leading jet η	114
Figure A.4c	: Leading jet ϕ	114
Figure A.4d	: Leading jet energy	114
Figure A.4e	: Sub-leading jet p_T	114
Figure A.4f	: Sub-leading jet η	114

Figure A.4 : DNN input distributions for the semi-leptonic final state (continued).	114
Figure A.5a : Sub-leading jet ϕ	115
Figure A.5b : Sub-leading jet energy	115
Figure A.5c : m_{j_0,j_1}	115
Figure A.5d : m_{j_1,j_2}	115
Figure A.5e : Odd training	115
Figure A.5f : Even training	115
Figure A.5 : DNN input distributions (a,b,c,d) and the ROC curves (e,f) for the semi-leptonic final state (continued).	115
Figure A.6a : Odd training	116
Figure A.6b : Even training	116
Figure A.6 : DNN evaluations for the semi-leptonic channel final state.	116
Figure A.7a : Leading Photon $p_T/m_{\gamma\gamma}$	119
Figure A.7b : Leading Photon $E/m_{\gamma\gamma}$	119
Figure A.7c : Leading Photon η	119
Figure A.7d : Leading Photon ϕ	119
Figure A.7e : Sub-leading Photon $p_T/m_{\gamma\gamma}$	119
Figure A.7f : Sub-leading Photon $E/m_{\gamma\gamma}$	119
Figure A.7 : DNN input distributions for the single τ final state.	119
Figure A.8a : Sub-leading Photon η	120
Figure A.8b : Sub-leading Photon ϕ	120
Figure A.8c : Leading Tau p_T	120
Figure A.8d : Leading Tau η	120
Figure A.8e : Leading Tau ϕ	120
Figure A.8f : Leading Tau energy	120
Figure A.8 : DNN input distributions for the single τ final state (continued).	120
Figure A.9a : Jet multiplicity	121
Figure A.9b : B-tagged jet multiplicity	121
Figure A.9c : Leading jet p_T	121
Figure A.9d : Leading jet η	121
Figure A.9e : Sub-leading jet p_T	121
Figure A.9f : Sub-leading jet η	121
Figure A.9 : DNN input distributions for the single τ final state (continued).	121
Figure A.10a E_T^{miss}	122
Figure A.10b Odd ROC curve	122
Figure A.10c Even ROC curve	122
Figure A.10d Odd training weight	122
Figure A.10e Even training weight	122
Figure A.10 : DNN input distributions for the single τ final state (continued) (a), corresponding ROC curves (b,c) and DNN evaluation on odd (d) and even (e) data.	122
Figure A.11a Cat. 1	127
Figure A.11b Cat. 2	127
Figure A.11c Cat. 3	127
Figure A.11d Semi-leptonic	127
Figure A.11e Fully-leptonic	127
Figure A.11 : Fitted distributions of semi-leptonic channel (a,b,c) and fully-leptonic (d,e).	127

Figure A.12a Single τ	128
Figure A.12b Single τ Cat. 1	128
Figure A.12c Double τ	128
Figure A.12 : Fitted distributions of single τ (a,b) and double τ final states.....	128

PROSPECTS OF NONRESONANT HIGGS BOSON PAIR PRODUCTION MEASUREMENT IN THE $WW\gamma\gamma$ CHANNEL AT THE HL-LHC WITH THE PHASE-II CMS DETECTOR

SUMMARY

Since the discovery of the Higgs boson in 2012 by the CMS and ATLAS experiments at the CERN's Large Hadron Collider in Geneva, Switzerland, physicists have tried to measure accurately its properties and to understand better the underlying electroweak symmetry breaking mechanism. In this pursuit, the search for Higgs boson pair production is crucial to test our understanding of the Higgs potential and to search for clues for the Beyond the Standard Model searches. This thesis describes the search for the Higgs boson pair production in decays to a W boson pair and a photon pair. The $\tau\tau\gamma\gamma$ channel of the Higgs boson pair decay is analysed alongside since an overlap is expected in the final states. Monte Carlo simulations of proton-proton collisions corresponding to an integrated luminosity of 3000 fb^{-1} at a centre-of-mass energy of 14 TeV are used. The gluon-gluon fusion production mode of the Higgs boson pair is considered only. The DELPHES fast detector simulation is used with an average pile-up of 200 per interaction with a dedicated card for Phase-2-upgraded CMS detector. A Python-based analysis library called BAMBOO is used to perform the object selections and event categorisation in the data analysis of the study. Cut-flow tables reporting the number of events at each final state of interest are shown. Two binary Deep Neural Networks (DNN) are employed using the KERAS API for TENSORFLOW machine learning library in order to increase the signal and background discrimination in the semi-leptonic final state of $WW\gamma\gamma$ and single τ final state of $\tau\tau\gamma\gamma$ channels. DNN score cuts are applied to each final state and the di-photon invariant mass distributions are obtained. The results are then used in the HIGGS COMBINE TOOL with the statistical and systematic uncertainties applied. The significance levels are reported for each final state along with their combination.

STANDART MODEL HIGGS BOZONU ÇİFTİ ÜRETİMİNİN $WW\gamma\gamma$ KANALINDA CMS FAZ II DEDEKTÖRÜ İLE HL-LHC KOŞULLARINDA ARAŞTIRILMASI

ÖZET

2012 yılında, İsviçre'nin Cenevre kenti ile Fransa sınırında konuşlanmış olan Avrupa Nükleer Araştırma Merkezi'nin (CERN) Büyük Hadron Çarpıştırıcısı (LHC) etrafındaki CMS ve ATLAS deneyleri, 125 GeV kütleinde bir parçacığın gözlemlendiğini açıkladı. İlk defa varlığı deneyler tarafından kanıtlanan bu parçacık, 1964 yılında Peter Higgs ve 5 diğer bilim adamı tarafından ortaya atılan Higgs bozonuydu. Bir sene sonra Peter Higgs ve François Englert Nobel Fizik Ödülü'ne layık görüldüler. 2012'den itibaren, elektro-zayıf simetri kırınımı mekanizmasını açıklayan ve Standart Model ötesi kuramlar için önemli bir parçacık olan Higgs bozonunun tüm özellikleri detaylı bir şekilde çalışılmaya başlandı.

27 km uzunluğunda çevreye sahip çember şeklinde bir yeraltı tüneline kullanan LHC deneyi, 4 adet büyük detektör barındırmaktadır. Bu detektörler; ATLAS, ALICE, CMS ve LHCb detektörleridir. Bunlardan ikisi, ATLAS ve CMS, 13 TeV kadar yüksek bir kütle merkezi enerjisinde çarpışan protonlardan saçılan parçacıkları gözlemleyerek, Higgs bozonu dahil olmak üzere diğer tüm temel parçacıkları araştırıyor. LHC deneyi, 2027 yılı sonrası için 14 TeV'lik bir çarpışma enerjisine çıkmayı ve toplam lüminositayı 3000 fb^{-1} değerine kadar yükseltmeyi planlarken CMS deneyi de, çarpışma noktasının etrafındaki kapsama alanını artırmak gibi çeşitli güncellemelerle geliştiriliyor.

Higgs bozonu çifti üretimi LHC'deki proton-proton çarpışmalarında görece yüksek bir tesir kesitine sahiptir. 13 TeV enerjide gluon-gluon füzyonu üretim yöntemi yaklaşık 30 fb 'dir. Yüksek enerjideki proton-proton çarpışmalarında Higgs bozonu çifti üretiminin önemi ise, Higgs potansiyelinin şeklini belirleyen Higgs tri-lineer öz etkileşim terimine, vektör bozon üretim mekanizması ile doğrudan erişiminin olmasıdır. LHC deneyinde üretilen Higgs bozonu çiftleri ATLAS ve CMS deneyleri tarafından bir çok bozunma kanalında çalışılmıştır. Fakat Yüksek Lüminosite LHC programı (HL-LHC), yani LHC deneyinin 2027 ve sonrası için planlanan güncellenmiş halindeki Higgs bozonu çifti üretimi henüz tüm bozunma kanallarında çalışılmamıştır.

Bu tezde, HL-LHC programı kapsamında yapılması planlanan güncelleme çalışmaları sonrasındaki lüminositada Higgs bozonu çifti üretimi, gluon-gluon füzyonu üretim yöntemi ile CMS detektörünün Faz-2 için güncellenmiş halinde, proton-proton çarpışması simülasyonları ile çalışılmıştır. 1. bölümde, parçacık fiziklerinin başarılı teorisi Standart Model'e giriş niteliğinde açıklamalar verilmiştir. Standart Model Lagrange denklemleri incelenmiş, temel kuvvetler, parçacıklar ve özellikle Higgs

bozonundan ve Higgs bozonu çift üretiminden bahsedilmiştir. Standart Model ötesi araştırmalardan ve bu araştırmalarda Higgs bozonunun öneminden bahsedilmiştir.

2. bölümde, Büyük Hadron Çarpıştırıcısı deneyinin yapısı açıklanmış ve simülasyon programları ile davranışı taklit edilen Kompakt Müon Solenoidi (CMS) detektörünün özellikleri ve alt detektörleri incelenmiştir. CMS detektörünün Faz-2 Güncelleme Çalışmaları ve HL-LHC programı açıklanmıştır.

Higgs bozonu çiftinin $WW\gamma\gamma$ bozunma kanalındaki analizi 3. bölümde detaylıca incelenmiştir. Öncelikle analiz stratejisinden bahsedilmiş ve simülasyon ile üretilen proton-proton çarpışma verileri; bu verilerin üretiminde kullanılan programlar ve bu programların arka planında çalışan metotlar, sinyal ve artalan çarpışma proseslerinin detayları ile CMS detektörünün simülasyonu için kullanılan bilgisayar programının (DELPHES) özellikleri açıklanmıştır. Analizde kullanılan fizik objeleri; fotonlar, leptonlar, jetler ve kayıp enerji, CMS detektörünün kapsama alanı içerisinde ve yeterli enerjiye sahip olacak şekilde tanımlanmıştır. Higgs bozonu çifti üretiminin istenilen kanallarda ($WW\gamma\gamma$, $\tau\tau\gamma\gamma$) incelenmesini sağlamak amacıyla çarpışma olaylarına belirli kriterler uygulanmıştır. Olay seçimi adı verilen bu işlem ve fizik objesi tanımlamaları, BAMBOO adı verilen Python temelli bir analiz kütüphanesi yardımıyla yapılmıştır.

TENSORFLOW makine öğrenmesi kütüphanesi ile KERAS arayüzü kullanılarak, analiz için 2 adet yapay sinir ağı geliştirilmiştir. Bu çalışmadaki modelin, analizi öğrenip tahminlerde bulunması için fizik objelerinin; olay başına miktarları, enine momentumları (*transverse momentum*), dik eksenindeki (η) ve azimut eksenindeki (ϕ) açı değerleri, fotonlar ve tau leptonları için; değişmez kütlelerinin (*invariant mass*) enerji ve enine momentumlarına oranları gibi değerler girdi olarak verilmiştir. Bu girdileri kullanarak 0 ile 1 arasında yapay zeka ağı puanı dağılımı çıkarılmıştır. Bu dağılımda sinyal ile artalan süreçlerinin birbirinden ayrılması ROC eğrileri ve istatistiksel önem değerleri olay bazında hesaplanarak sağlanmış ve yapay zeka ağı puanının farklı aralıklardaki değerleri için kategoriler oluşturulmuştur. Belirlenen bu kategoriler analiz kodlamasında yeni birer seçim olarak uygulanmış ve bu seçim sonrasında elde edilen foton çifti değişmez kütlesi dağılımları çıkarılmıştır.

Sistemik belirsizlikler 3. bölümün son konusu olarak açıklanmıştır. Bu kısımda CMS ve ATLAS deneyleri tarafından önerilen deneysel ve teorik sistemik belirsizlikler uygulanmıştır. Deneysel belirsizlikler; lüminosite, foton çifti tetiklemesi, foton çifti değişmez kütlesi, foton ID, elektron ID, muon ID, tau ID ve jet enerji ölçeği belirsizlikleridir. Teorik belirsizlikler ise, tesir kesiti belirsizlikleri, QCD skalası ve parton dağılım fonksiyonları (PDF) olarak her bir proses veri setine uygulanmıştır. Farklı olarak sinyal proseslerine top kuark kütlelerinin belirsizliği de eklenmiştir. Daha sonra HIGGS COMBINE TOOL adı verilen, Higgs analizleri için özelleştirilmiş bir analiz kütüphanesi kullanılarak istatistiksel önem düzeyleri (*statistical significance*) çıkarılmıştır. HIGGS COMBINE TOOL kütüphanesine, foton çifti değişmez kütlesi dağılımı girdi olarak verilmiş ve sistemik belirsizlikler bu aşamada uygulanmıştır.

Sonuçların sunulduğu 4. bölümde ise, anlamlılık düzeyleri ve yakınlık taramaları verilmiş ve analiz açısından önemleri tartışılmıştır. Bazı son bozunma durumlarındaki foton çifti kütlesi dağılımları artalan ve sinyal proses sonuçlarına uyarlanıp grafik olarak sunulmuştur. Ekler bölümünde, analizdeki kullanılan tüm veri setleri listelenmiş, tüm bozunma durumlarında seçim yapılmadan önceki ve sonraki olay

sayıları ve seçilim verimleri yüzdelik olarak tablo halinde verilmiştir. Ayrıca yapay zeka sinir ağlarına girdi olarak kullanılan tüm dağılımlar verilmiş, ve performans grafikleri eklenmiştir.

1. INTRODUCTION TO THE STANDARD MODEL

The Particle Physics is probing the smallest objects that are known as elementary particles and tries to extend our knowledge of the subatomic world. These elementary particles are accelerated, collided and detected at very high energies in the experiments around the globe - one of them being the largest experimental setup ever built for science - and studied after being detected. *The Standard Model* of particle physics (SM) is the theory of the fundamental interactions in this pursuit. It is a quantum field theory developed with the contribution of many scientists around the world mainly in the second half of the 20th century, and over the last few decades it has been shown to be an accurate description of the picture. It describes all known particles but is a mathematical description of three of the four known fundamental forces of the nature. These are the electromagnetic interaction, and the weak and the strong nuclear interactions. The gravitational force, due to difficulties in combining general relativity with quantum mechanics, does not take place in the Standard Model. The gravity is known to be 10^{40} times weaker than the electromagnetic force, thus its effects are expected to be negligible in the theory.

The Standard Model, being a quantum field theory, was formulated in the 1960's. It is grounded on the mathematical concept of the local gauge invariance under the symmetry groups of its Lagrangian equations. It has predicted successfully a wide range of phenomena which have been observed in the experiments. The discovery of the W^{\pm} and Z^0 bosons at CERN's Super Proton Synchrotron (SPS) by the UA1 and UA2 Collaborations in 1983 [2–6], the observation of the top quark at FNAL's Tevatron by $D\phi$ and CDF Collaborations in 1995 [7, 8], the observation of the Higgs boson at CERN's Large Hadron Collider Experiment (LHC) by CMS and ATLAS Collaborations in 2012 [9, 10] and many other discoveries along with verification can be shown as the biggest achievements of the Standard Model.

However, this successful theory had an imperfection; it could not explain the reason for the masses of the W^{\pm} and Z^0 bosons. The solution came from R. Brout, F. Englert

and P. Higgs in 1964 called the Brout-Englert-Higgs (BEH) Mechanism [11–13]. They introduced a complex scalar field into the SM and assumed a spontaneous symmetry breaking, hence giving mass to the vector bosons while preserving the unitarity of the Standard Model. This mechanism anticipated the existence of a scalar boson field. Later on, the BEH mechanism was incorporated in the electroweak model by Salam and Weingberg in 1964 [14], and proven that it is renormalisable by 't Hooft and Veltman in 1972 [15]. The predicted boson is called the Higgs boson and searched by high energy physics experiments at the forefront of scientific research nearly for 50 years.

In 1995, the technical design report for the LHC Experiment was published [16]; it involved constructing a new particle accelerator in the Franco-Swiss border in Geneva and was intended to explore the TeV energy scale up to centre-of-mass energies of 14 TeV. This experiment has lead the front line of the experimental particle physics research for many years and discovered the Higgs boson at a mass of 125 GeV which confirmed the BEH mechanism. The details of the LHC is explained in section 2.1.

In the SM, the elementary particles of matter and the ones that carry forces between them are named as fermions and bosons, respectively. The distinction shows itself in their spin properties; fermions have a spin value of half an integer whereas bosons have integer spin values. Besides, fermions obey the Pauli exclusion principle meaning that any two fermions cannot be at the same quantum state in a bound structure, whereas bosons do not need to obey the same rule. The first generation of fermions include up and down quarks, electron and its neutrino counterpart; and bosons include photons, W and Z bosons, gluons and the Higgs boson. When added the second and the third generation of particles - whose existence is one of the mysteries of nature - along with their anti-particles, they form the fundamental particles that are known today. A tabular form of these particles can be seen in Figure 1.1.

Fermions make up the ordinary matter that we see around us everyday. A subgroup of fermions are **leptons** and they include electron, muon, tau and their neutrino counterparts. Neutrinos have a very small mass compared to other leptons and quarks. They do not have electromagnetic nor colour charge which makes them obey only the weak force and they barely interact with matter. Flavours of leptons other than the

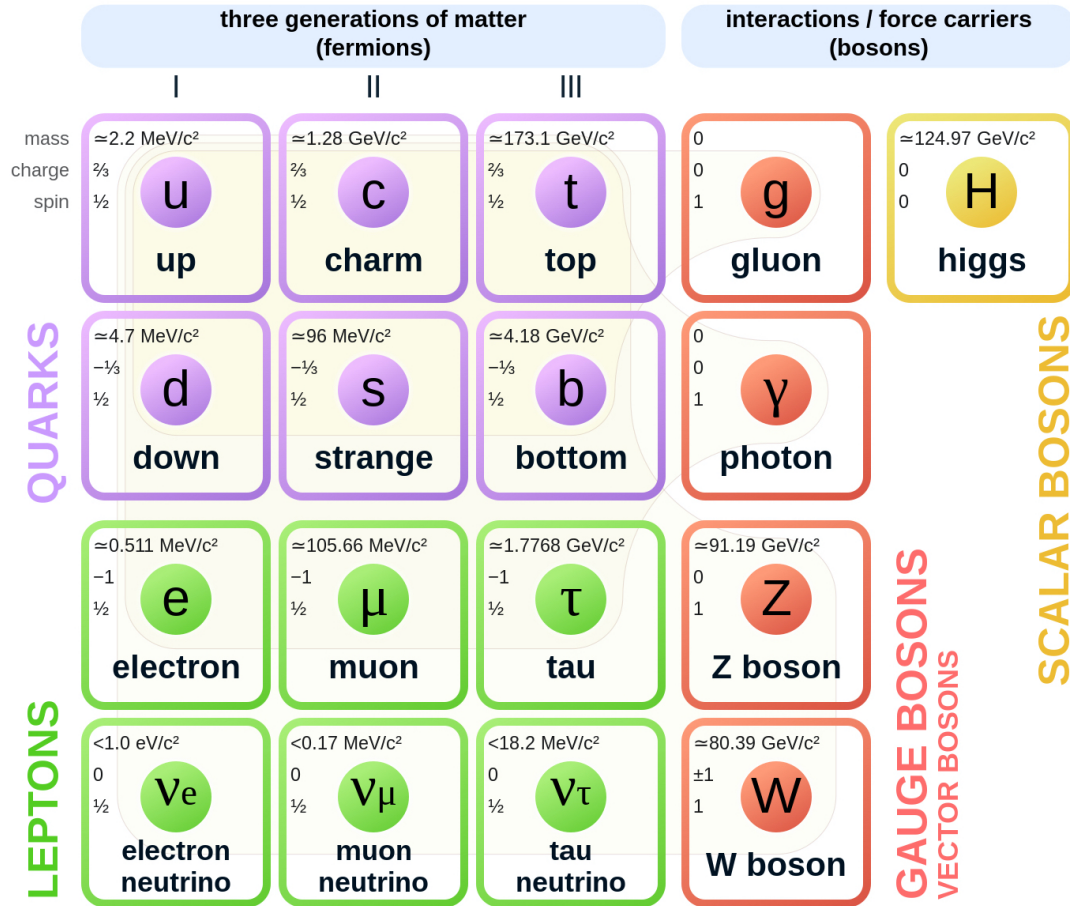


Figure 1.1 : The elementary particles of the Standard Model. Fermions and bosons are grouped in columns where the quarks, leptons, gauge bosons and scalar Higgs boson are shown in different colours. The three generations of matter are indicated with roman numerals. The mass, charge and spin values corresponding to each particle are indicated on the upper left of each box. The vector bosons are connected in faint yellow areas to the particles they interact with [17].

neutrinos (electrons, muons and taus) have sizeable mass and charge and they are the members of the three generations of matter.

Quarks are the other type of fermions. They have three generations as leptons do, and they include six different flavours, namely up, down, charm, strange, top and bottom quarks. They interact with electromagnetic, weak and strong nuclear forces since they have colour charges in addition to their hypercharge.

Bosons are the mediator particles of the three forces described in the SM. The particles of this type are called the gauge bosons and they include photon (γ), W^\pm bosons, Z boson, gluons and the Higgs boson. Photons and gluons do not have mass where W

bosons have about $80 \text{ GeV}/c^2$, Z boson $91 \text{ GeV}/c^2$ and the Higgs boson has $125 \text{ GeV}/c^2$ of mass. Among all bosons, only W^\pm bosons have electric charge. Moreover, only the Higgs boson has a spin value of zero making it a scalar boson where all other bosons are spin-1 gauge bosons, making them vector bosons. Fermions and bosons interact with fundamental forces which we will look at now.

The ***electromagnetic interaction***, whose quantum field theory is established by the *Quantum Electrodynamics (QED)*, is mediated by photons and acts only on the particles that have electric charges. Almost everything we see in the daily life is governed by the electromagnetic force. The carrier of this force, photons, do not carry electric charge, nor self-interacts. Since photons have zero mass, their interaction range is unlimited.

The ***weak nuclear force***, which is responsible for the decay of the particles that is a flavour changing interaction, acts only on the fermions. It has a very short interaction range. The neutral and the charged current interactions are mediated by vector bosons of the this force; Z boson and W^\pm bosons, respectively. The Higgs boson in this picture, plays the role of generating the masses of the W^\pm , and Z bosons through the Higgs mechanism explained in subsection 1.1.3, and of the quarks and leptons (except neutrinos) through Yukawa interaction [18].

The ***Quantum Chromodynamics (QCD)*** is the theory of the ***strong nuclear force*** and describes the interaction between quarks mediated by the massless gluons. It has an interaction range of about 10^{-15} m . Unlike the electromagnetic interaction, the gauge bosons of the strong nuclear force possess the charge of the interaction, namely colours, hence can couple to itself. There are 3 types of colour; red, green and blue and each quark in the universe carry one of these; they carry anticolour if they are antiquarks. In this fashion, gluons can be thought of colour carrier particles.

A phenomenon called the ***colour confinement*** states that colour-charged particles cannot be isolated, meaning that only colour-neutral particles can be observed in nature¹. This imposes that gluons carry a pair of charge consisting a colour and an anticolour. Likewise, the hadrons are colour-neutral in two ways; i) with a pair consisting a colour and an anticolour or ii) with 3 different types of colour or anticolour.

The first combination makes up mesons; consisting of a quark pair (a quark and an anti-quark), and the second combination forms baryons.

So far, we have taken a glance at the fundamental interactions and particles of the Standard Model. Many of the predictions of the Standard Model are successfully observed in the experiments, and no other particle nor force (except gravity) is found beyond the description given by the SM. However, the theory does not provide answer to the unsolved problems in the fundamental physics such as the non-zero masses of neutrinos [20], or the dark matter and the dark energy [21] which are the dominated energy content of the universe. Therefore the SM is seen as an effective field theory (EFT)².

1.1 The Standard Model Lagrangian and the Higgs Mechanism

The Standard Model describes the behaviour of the mentioned fundamental interactions and elementary particles in a single *Lagrangian density* formulation, often abbreviated as Lagrangian. It is often considered in two parts: the strong sector offers a description for the particles with colour charges, while the electroweak part consists of the electromagnetism and the weak force. The SM is a renormalised gauge field theory with the $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ gauge form and the charges are colour, weak isospin and hypercharge, respectively. The $SU(2)_L$ and $U(1)_Y$ groups mix and create the W^\pm , Z and γ bosons where $SU(3)_C$ gauge group describes the strong force. The Lagrangian here, does not involve the particle masses but they are introduced to the theory via the *spontaneous symmetry breaking* (SSB) of the the electroweak gauge group $SU(2)_L \otimes U(1)_Y$, which we will address after studying the SM Lagrangian.

Mathematical interpretation of the SM is provided by the *Quantum Field Theory (QFT)*, where each particle is represented by a quantum field that is pervaded across the space-time. Most of the field theories normally starts with defining a set of symmetries of the system and continues with writing down the renormalisable Lagrangian of the particles or fields that obey these symmetries. The QFT also follows this path; it

¹ Below the Hagedorn temperature (T_H) of about 0.15 GeV. At the energies higher than 1.7×10^{12} K, hadrons become unstable and it can be thought of the boiling point of the hadronic matter [19].

²An effective field theory is a form of estimation for an underlying physical phenomena at a given energy or length scale.

consists of translational symmetry, rotational symmetry and a boost symmetry that is the invariance of an inertial reference frame. Under these symmetries, the Lagrangian of the SM can be interpreted in three parts; Quantum Chromodynamics, Electroweak theory and the Higgs Mechanism.

1.1.1 Quantum chromodynamics

Starting from the free-field Dirac Lagrangian density for a massless spin-1/2 fermion, which is the quark field in the QCD case, reads,

$$\mathcal{L} = \bar{\psi}(x) (i\gamma^\mu \partial_\mu) \psi(x) , \quad (1.1)$$

where ψ is the free-field for the fermion and γ^μ represents the Dirac matrices. The explanation below is valid with a mass term of $m\bar{\psi}\psi$, and $\not{\partial}_\mu \equiv \partial_\mu \gamma^\mu \equiv \partial^\mu \gamma_\mu$ is implied. The reason to discuss massless fermions is explained in the next two sections. The transformation of the fermion field under the $SU(3)_C$ group reads,

$$\psi(x) \rightarrow e^{ig\frac{\lambda^a}{2}\theta_a(x)} \psi(x) , \quad (1.2)$$

where $\frac{\lambda^a}{2}$ are the 8 Gell-Mann matrices and generators of the group. It is worth noting that $\partial_\mu \psi(x)$ does not transform in the same way and needs a redefinition, such that,

$$D_\mu = \partial_\mu - igA_\mu^a(x) \frac{\lambda^a}{2} , \quad (1.3)$$

where the gauge vector fields A_μ^a represent the eight gluons. These gluon fields must transform in a specific way to satisfy the local gauge invariance, such that,

$$A_\mu^a \rightarrow A_\mu^a + \partial_\mu \theta^a + gf^{abc} A_\mu^c \theta^b , \quad (1.4)$$

where f^{abc} denotes the structure constants according to the following commutation $\left[\frac{\lambda^a}{2}, \frac{\lambda^b}{2}\right] = if^{abc} \frac{\lambda^c}{2}$. We have achieved the invariance of the Lagrangian density by introducing the vector fields. Lastly, adding the kinetic term, whose form is

$$-\frac{1}{4}F_a^{\mu\nu}F_{\mu\nu}^a , \quad (1.5)$$

where

$$F_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + gf^{abc} A_\mu^b A_\nu^c . \quad (1.6)$$

The complete QCD Lagrangian density becomes,

$$\mathcal{L}_{QCD} = \bar{\psi} (i\gamma^\mu \partial_\mu) \psi(x) - g\bar{\psi}(x)\gamma^\mu \frac{\lambda_a}{2} \psi(x) A_\mu^a - \frac{1}{4}F_a^{\mu\nu}F_{\mu\nu}^a , \quad (1.7)$$

where a summation convention is implied over all the quark fields. In the Lagrangian density, the first term is the same free-field propagation as in the original equation and the second term emerges from the interaction of the quark field with the introduced vector field A_μ along with the covariant derivative. The constant g parametrises the interaction strength and is usually redefined as $\alpha_s = g^2/4\pi$ which is the strong coupling constant. The third term in the Equation 1.6, when substituted in Equation 1.7 in the kinetic term, results in the cubic and quartic self coupling of the gluon fields which are a general feature of the non-Abelian gauge theories. The gluons and the quark fields are produced via imposing the local gauge invariance, and $SU(3)_C$ group provides the 8 generators that are the 8 gluons of the strong force.

1.1.2 Electroweak theory

Electroweak interactions are explained with the same local gauge invariance but under the $SU(2)_L \otimes U(1)_Y$ group. The parity violating nature of the weak interactions requires defining different interactions for fermions of opposite chiralities. The left hand chirality projection operator is defined as $\frac{1-\gamma^5}{2}$ where the corresponding operator for the right hand chirality projection is $\frac{1+\gamma^5}{2}$, where γ^5 is defined as

$$\gamma^5 \equiv \gamma^0 \gamma^1 \gamma^2 \gamma^3 ; \quad (1.8)$$

and for a massless particle, the chirality corresponds to the helicity of that particle which is the normalised projection of the spin vector on the momentum vector.

The first part of the symmetry group, $SU(2)_L$, is associated with the weak isospin (I), and is a non-Abelian gauge group. The gauge invariance under the $SU(2)_L$ group introduces W_μ^i , the gauge fields where i runs over the three generators of this group. Fermion fields are stated with chirality components as left-handed doublets and right-handed singlets, and right handed particles do not interact with the W_μ^i gauge fields.

The second part, $U(1)_Y$ is an Abelian group. B_μ is the gauge field of this group, associated with weak hypercharge (Y) and results from the local gauge invariance. This field interacts with both the left and right handed particles. Here, the weak hypercharge symmetry is defined different than the Quantum Electrodynamics (QED) in order to unify the electrodynamics with the weak interactions. The relation between the electric

charge Q and the hypercharge Y_W is given by the Gell-Mann-Nishijima formula;

$$Q = I_3 + \frac{1}{2}Y_W , \quad (1.9)$$

where I_3 is the third component of the weak isospin. The fermion fields can now be written as one doublet ψ_L field and, two singlet ψ_R and ψ'_R fields, such that,

$$\psi_L = \frac{1 - \gamma^5}{2} \begin{pmatrix} \psi \\ \psi' \end{pmatrix} = \begin{pmatrix} \psi_L \\ \psi'_L \end{pmatrix} , \quad (1.10)$$

$$\psi_R = \frac{1 + \gamma^5}{2} \begin{pmatrix} \psi \\ \psi' \end{pmatrix} = \begin{pmatrix} \psi_R \\ \psi'_R \end{pmatrix} , \quad (1.11)$$

where ψ and ψ' denotes either neutrino, charged lepton fields or up, down-type quark fields. A leptoquark coupling is not predicted in the SM nor observed, meaning that these two sectors are separate. In other words, these two sectors cannot transform lepton fields into quark fields, vice versa.

The Lagrangian density so far, consists of 3 main parts,

$$\begin{aligned} \mathcal{L} &= \mathcal{L}_{kin} + \mathcal{L}_{CC} + \mathcal{L}_{NC} \\ &= i\bar{\psi}_L \not{D} \psi_L + i\bar{\psi}_R \not{D} \psi_R + i\bar{\psi}'_R \not{D} \psi'_R , \end{aligned} \quad (1.12)$$

where

$$D_\mu = \partial_\mu - igW_\mu^i T_i - ig' \frac{Y_\psi}{2} B^\mu , \quad (1.13)$$

and $T_i = \frac{\sigma_i}{2}$ which represents the Pauli matrices that are the generators of the $SU(2)_L$ group with eigenvalues that are actually the weak isospin values. $\mathcal{L}_{CC/NC}$ formalism here corresponds to the charged-current and neutral-current interactions, respectively. The Lagrangian density can then be rewritten as,

$$\mathcal{L}_{kin} = i\bar{\psi}_L \not{\partial} \psi_L + i\bar{\psi}_R \not{\partial} \psi_R + i\bar{\psi}'_R \not{\partial} \psi'_R , \quad (1.14)$$

$$\begin{aligned} \mathcal{L}_{CC} &= gW_\mu^1 \bar{\psi}_L \gamma^\mu \frac{\sigma_1}{2} \psi_L + gW_\mu^2 \bar{\psi}_L \gamma^\mu \frac{\sigma_2}{2} \psi_L \\ &= \frac{g}{\sqrt{2}} W_\mu^+ \bar{\psi}_L \gamma^\mu \sigma^+ \psi_L + \frac{g}{\sqrt{2}} W_\mu^- \bar{\psi}_L \gamma^\mu \sigma^- \psi_L , \end{aligned} \quad (1.15)$$

$$\begin{aligned} \mathcal{L}_{NC} &= \frac{g}{\sqrt{2}} W_\mu^3 (\bar{\psi}_L \gamma^\mu \psi_L - \bar{\psi}'_L \gamma^\mu \psi'_L) + \frac{g'}{2} B_\mu [Y_{\psi_L} (\bar{\psi}_L \gamma^\mu \psi_L + \bar{\psi}'_L \gamma^\mu \psi'_L) \\ &\quad + Y_{\psi_R} (\bar{\psi}_R \gamma^\mu \psi_R) + Y_{\psi'_R} (\bar{\psi}'_R \gamma^\mu \psi'_R)] , \end{aligned} \quad (1.16)$$

where

$$W_\mu^\pm = \frac{1}{\sqrt{2}} (W_\mu^1 \mp iW_\mu^2) \sigma_\mu^\pm = \frac{1}{2} (\sigma^1 \pm i\sigma^2) , \quad (1.17)$$

which is the charged current interaction between the ψ_L and ψ'_L fields mediated by charged weak boson with W_μ^\pm fields. The neutral current field is described by the neutral Z boson field Z_μ and the photon field A_μ . This interactions could not be mediated by the W_μ^3 field nor by B_μ , since both couples to neutral fields. The relation between these fields are given by,

$$B_\mu = A_\mu \cos \theta_W - Z_\mu \sin \theta_W , \quad (1.18)$$

$$W_\mu^3 = A_\mu \sin \theta_W + Z_\mu \cos \theta_W , \quad (1.19)$$

where θ_W is the Weinberg angle. The neutral current interactions are derived by substituting Equation 1.18 and Equation 1.19 in Equation 1.16, resulting in the coupling strength of $g \sin \theta_W I_3 + g' \cos \theta_W \frac{Y}{2}$ for A_μ field. The electroweak unification is achieved by equating this coupling strength to the photon field's coupling constant. Since the hypercharge is only multiplied by g' , we can set an arbitrary value for it as $Y_{\psi_L} = -1$. This process then gives, along with $Q = -1$ for leptons and 0 for neutrinos,

$$g \sin \theta_W = g' \cos \theta_W = e , \quad (1.20)$$

The Lagrangian density for the electroweak interactions then becomes,

$$\mathcal{L}_{EW} = i \bar{\psi}_L \not{D} \psi_L + i \bar{\psi}_R \not{D} \psi_R + \bar{\psi}'_R \not{D} \psi'_R - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} - \frac{1}{4} W_{\mu\nu}^i W_i^{\mu\nu} , \quad (1.21)$$

where

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu , \quad (1.22)$$

$$W_{\mu\nu}^i = \partial_\mu W_\nu^i - \partial_\nu W_\mu^i + g \epsilon^{abc} W_\mu^b W_\nu^c , \quad (1.23)$$

Equation 1.21 includes the charged and neutral interactions of fermions. Expanding the kinetic terms of the weak bosonic field yields the self couplings of the gauge bosons including the trilinear and quartic couplings. The mass terms for the bosonic fields violates the local gauge invariance as the same is valid for the strong force. Also, they are not allowed for the fermionic fields since the left and right chiralities transform individually which we will see in the next section. A list of fermionic fields is shown in Table 1.1.

In brief, the fermion fields share the same composition; the left and right chirality fields are doublet and singlets under the $SU(2)_L$ group, respectively, resulting in the

Table 1.1 : Fermion fields under $SU(2)_L$ group representation.

Type	1 st gen.	2 nd gen.	3 rd gen.	I_3	Y	Q	$SU(3)_C$
Quarks	$\begin{pmatrix} u_L \\ d_L \end{pmatrix}$	$\begin{pmatrix} c_L \\ s_L \end{pmatrix}$	$\begin{pmatrix} t_L \\ b_L \end{pmatrix}$	$\begin{pmatrix} 1/2 \\ -1/2 \end{pmatrix}$	1/3	$\begin{pmatrix} 2/3 \\ -1/3 \end{pmatrix}$	triplet
	u_R	c_R	t_R	0	4/3	2/3	
	d_R	s_R	b_R	0	-2/3	-1/3	
Leptons	$\begin{pmatrix} \nu_{e,L} \\ e_L \end{pmatrix}$	$\begin{pmatrix} \nu_{\mu,L} \\ \mu_L \end{pmatrix}$	$\begin{pmatrix} \nu_{\tau,L} \\ \tau_L \end{pmatrix}$	$\begin{pmatrix} 1/2 \\ -1/2 \end{pmatrix}$	-1	$\begin{pmatrix} 0 \\ -1 \end{pmatrix}$	singlet
	e_R	μ_R	τ_R	0	-2	1	

possession of the weak isospin in the doublets. This interaction is mediated by the W^\pm bosons where neutral interactions is mediated by the Z boson; it interacts with both chiralities but this time with a different coupling strength due to the mixing of the gauge fields shown in Equation 1.18 and Equation 1.19. The electromagnetic interaction is mediated by photons that is only sensitive to electric charge Q, hence to hypercharge Y and to weak isospin I_3 with the relation given in Equation 1.9.

The interactions between fields may differ the quantum numbers by carrying charges; the strong interactions can change the colour charge of quarks, and charged weak interactions can change the weak isospin, hence the electric charge of fermions. The Standard Model concept, explaining the matter fields by quantum numbers and the interactions by symmetry applications is an elegant mathematical theory, however the experimentally observed massive W^\pm and Z bosons requires an addition to the theory, explained in the next section.

1.1.3 Higgs mechanism

So far, we have built the SM on the assumption that the interactions are gauge invariant. This requires the vector bosons W^\pm and Z to be massless. In addition, for a fermion field ψ satisfying the Dirac equation $(i\hbar\gamma^\mu\partial_\mu - mc)\psi = 0$, the mass term $-m\bar{\psi}\psi$ arises which is actually not invariant under the $SU(2)_L$ gauge symmetry. This can be seen by expanding the mass term with left and right handed fermion fields, that is

$$-m\bar{\psi}\psi = -m(\bar{\psi}_L\psi_R + \bar{\psi}_R\psi_L), \quad (1.24)$$

and we have seen that the left-handed fields are doublets under $SU(2)_L$ while right-handed fields are singlet. This means that their gauge quantum numbers are different and this kind of mass term is forbidden. The solution to these theoretically massless but experimentally massive particles is given by *the Brout-Englert-Higgs*

Mechanism [11–13] (BEH), usually called the Higgs Mechanism, showing that the electroweak symmetry can be broken spontaneously under specific conditions.

The solution starts with introducing a complex scalar field ϕ to the theory, which is a doublet under $SU(2)_L$,

$$\phi = \begin{pmatrix} \phi_+ \\ \phi_0 \end{pmatrix}, \quad (1.25)$$

This field has a hypercharge value $Y = 1$ and the corresponding covariant derivative,

$$D_\mu = \partial_\mu - igW_\mu^i \frac{\sigma_i}{2} - \frac{1}{2}ig'B_\mu. \quad (1.26)$$

The Lagrangian describing the Higgs field becomes,

$$\mathcal{L}_{Higgs} = (D_\mu \phi)^\dagger (D^\mu \phi) - V(\phi^\dagger \phi). \quad (1.27)$$

Since the potential V must satisfy $SU(2)$ and $U(1)$ symmetries, a good solution is a Mexican hat potential;

$$V(\phi) = -\mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2. \quad (1.28)$$

A spontaneous symmetry breaking happens when both the constants μ^2 and λ are positive values. A convenient choice for the minimum would be

$$\langle \phi \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}. \quad (1.29)$$

The minimum of this potential is non-zero with the potential shape in Figure 1.2 and given by,

$$\begin{aligned} \frac{dV}{d(\phi^\dagger \phi)} &= \mu^2 + 2\lambda \phi^\dagger \phi = 0 \\ \Rightarrow |\phi_{min}| &= \sqrt{\frac{\mu^2}{2\lambda}} \equiv \frac{v}{2}, \end{aligned} \quad (1.30)$$

where v is called the **vacuum expectation value** (vev). The symmetry is broken once a particular ground state is chosen but the Lagrangian density remains gauge invariant.

Now let us expand the complex doublet field around the minimum $\phi(x)$,

$$\langle \phi \rangle = \frac{1}{\sqrt{2}} e^{\frac{i\sigma_i \theta^i(x)}{v}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}. \quad (1.31)$$

This corresponds to a real scalar massive field, to three massless **Goldstone boson fields** θ^i , which are not observed in nature. These bosons may be created as many

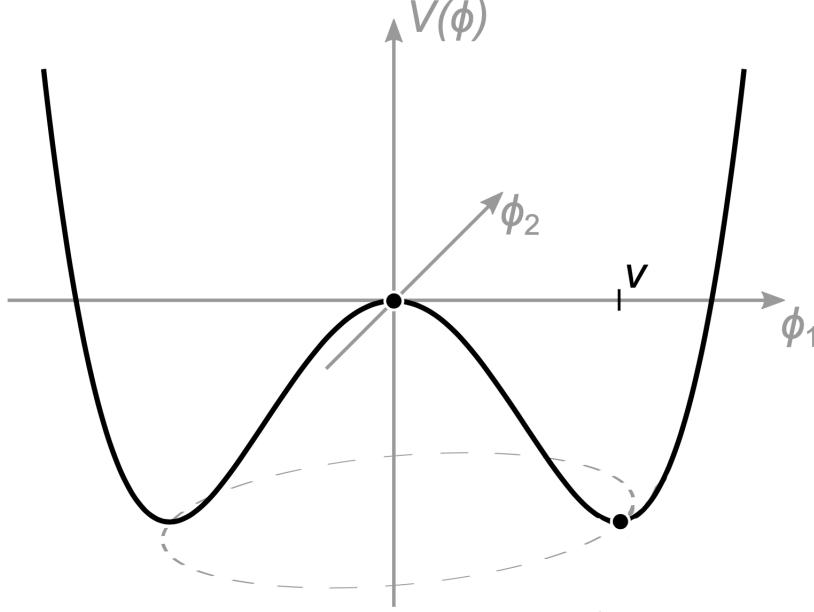


Figure 1.2 : The Higgs potential V in the case that $\mu^2 > 0$. Choosing any point at the bottom of the potential breaks spontaneously the rotational $U(1)$ symmetry.

as the broken generators and can be removed with an $SU(2)_L$ transformation with a unitary gauge choice:

$$\phi(x) \rightarrow \phi'(x) = e^{-\frac{i\sigma_i \theta^i(x)}{v}} \phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}. \quad (1.32)$$

This transformation yields only the real scalar field. Substituting Equation 1.26 and Equation 1.32, the Lagrangian density with the Higgs field h , becomes

$$\begin{aligned} \mathcal{L}_{Higgs} = & \frac{1}{2} \partial^\mu h \partial_\mu h - \frac{1}{2} (2\lambda v^2) h^2 \\ & + \left[\left(\frac{gv}{2} \right)^2 W^{\mu+} W_\mu^- + \frac{1}{2} \frac{(g^2 + g'^2) v^2}{4} Z^\mu Z_\mu \right] \left(1 + \frac{h}{v} \right)^2 \\ & - \lambda v h^3 - \frac{\lambda}{4} h^4 + \frac{\lambda}{4} v^4. \end{aligned} \quad (1.33)$$

The first two terms represent the kinetics of the Higgs field with the mass value $m_H^2 = 2\lambda v^2 = 2\mu^2$. The third term represents the weak boson masses with these values;

$$m_W^2 = \frac{g^2 v^2}{4} \Rightarrow m_W = \frac{gv}{2}, \quad (1.34)$$

$$m_Z^2 = \frac{(g^2 + g'^2) v^2}{4} = \frac{m_W^2}{\cos^2 \theta_W} \Rightarrow m_Z = \frac{m_W}{\cos \theta_W}. \quad (1.35)$$

It can be seen that the removed Goldstone bosons returned as extra degrees of freedom for the W^\pm and Z bosons. The weak bosons gained mass via this mechanism. The

third term of the Lagrangian defines the interactions between the Higgs and the weak bosons. It can be seen that the contributions of $H - WW/ZZ$ interactions come from $2h/v$ term and of $HH - WW/ZZ$ interactions from h^2/v^2 term when the square bracket is expanded. The final terms demonstrate the cubic and quartic self-couplings of the Higgs field. The potential now can be rewritten with the self-couplings as,

$$V(h) = \frac{1}{2}m_h^2 h^2 + \lambda_{hhh} v h^3 + \frac{1}{4}\lambda_{4h} h^4 - \frac{\lambda}{4} v^4, \quad (1.36)$$

where the self interaction couplings are defined as,

$$\lambda_{hhh} = \lambda_{4h} = \frac{m_h^2}{2v^2}. \quad (1.37)$$

Here, it is worth noting that these two self-couplings are dependent on the Higgs boson mass and the vacuum expectation value. Their experimental measurement provides a practical test for the SM and the electroweak symmetry breaking. There is one last term in the Higgs potential, $-\lambda v^4/4$. This constant, being irrelevant in the SM, contributes to the cosmological constant and is not compatible with the astronomical observations [22]. This is yet another mystery to be solved in the SM.

There is now the mass of the Higgs boson and the vacuum expectation value as free parameters to SM. The vacuum expectation value can be calculated precisely with the muon lifetime [23], such that,

$$\frac{G_F}{\sqrt{2}} = \left(\frac{g}{2\sqrt{2}} \right)^2 \frac{1}{m_W^2} \Rightarrow v = \sqrt{\frac{1}{\sqrt{2}G_F}} \approx 246 \text{ GeV}, \quad (1.38)$$

where G_F is the Fermi constant. Mass terms for the fermions are generated by the Higgs field's Yukawa interaction between the left and right chiral fields. The interaction's Lagrangian density can be written as,

$$\mathcal{L}_{Yukawa} = -y_{f'} \left(\bar{\psi}_L \phi \psi_R' + \bar{\psi}_R' \phi^\dagger \psi_L \right) - y_f \left(\bar{\psi}_L \tilde{\phi} \psi_R + \bar{\psi}_R \tilde{\phi}^\dagger \psi_L \right), \quad (1.39)$$

with the broken symmetry,

$$\tilde{\phi} = i\sigma_2 \phi^* = \begin{pmatrix} \phi_0^* \\ -\phi_+^* \end{pmatrix} \xrightarrow{\text{EWSB}} \frac{1}{\sqrt{2}} \begin{pmatrix} v + h(x) \\ 0 \end{pmatrix}, \quad (1.40)$$

where ψ denotes up and ψ' denotes down-type fermions. When the **Electroweak Symmetry Breaking (EWSB)** applied, the Lagrangian density reads,

$$\mathcal{L}_{Yukawa} = -\sum_f m_f (\bar{\psi}_L \psi_R + \bar{\psi}_R \psi_L) \left(1 + \frac{h}{v} \right), \quad (1.41)$$

where up-type and down-type fermions are summed with the mass terms,

$$m'_f = y'_f \frac{v}{\sqrt{2}}. \quad (1.42)$$

Now we have given the explanation for the fermion masses generated via the interaction with the Higgs field. The strengths of the interactions depend on the fermion masses; stronger interactions with heavier particles.

Finally the Standard Model Lagrangian becomes,

$$\begin{aligned} \mathcal{L}_{SM} = & -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} - \frac{1}{4}W_{\mu\nu}W^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} \\ & + \bar{\psi}i\gamma^\lambda D_\lambda \psi + (D_\mu \phi)^\dagger (D^\mu \phi) - V(\phi^\dagger \phi) \\ & + \mathcal{L}_{Yukawa} + h.c. . \end{aligned} \quad (1.43)$$

The EWSB is shown later to be a renormalisable theory 't Hooft and Veltman [15].

The Higgs boson, required for the spontaneous symmetry breaking, was found experimentally by the CMS and ATLAS Experiments at the LHC Experiment at CERN [9, 10] in 2012, almost 50 years after its theoretical assumption. The mass of the Higgs boson is 125 GeV and its parameters; spin, parity and branching ratios are found to be consistent with the Standard Model predictions [24, 25].

The most general form of the SM Lagrangian depends on 19 parameters. These parameters are given in Table 1.2.

1.2 The Higgs Boson Phenomenology and Experimental Status

The Higgs boson, with its zero spin, is different from any other particle in the Standard Model. It does not carry any electric or colour charge hence does not take part in the electromagnetic and strong interactions. Nonetheless, it interacts with fermions and heavy bosons including itself, since it carries weak isospin. The coupling strengths of the Higgs boson to those particles are derived in subsection 1.1.3. These couplings are only dependent on the mass of the particles which Higgs couple to, resulting in stronger coupling strengths to heavier particles.

The discovery of the Higgs Boson in 2012 was made with the data collected at the Run-1 of the CMS and ATLAS Detectors at $\sqrt{s} = 7$ and 8 TeV with an integrated luminosity of about 5 fb^{-1} in its decays to $H \rightarrow b\bar{b}$, $H \rightarrow ZZ$, $H \rightarrow \gamma\gamma$, $H \rightarrow WW$ and $H \rightarrow \tau\tau$. Further data at $\sqrt{s} = 13$ TeV collected with the same detectors at Run-2

Table 1.2 : Free parameters of the Standard Model.

#	Symbol	Name	Value
1	m_e	Electron mass	0.511 MeV
2	m_μ	Muon mass	105.7 MeV
3	m_τ	Tau mass	1.78 GeV
4	m_u	Up quark mass	1.9 MeV
5	m_d	Down quark mass	4.4 MeV
6	m_s	Strange quark mass	87 MeV
7	m_c	Charm quark mass	1.32 GeV
8	m_b	Bottom quark mass	4.24 GeV
9	m_t	Top quark mass	173.5 GeV
10	θ_{12}	CKM 1-2 Mixing angle	13.1°
11	θ_{23}	CKM 2-3 Mixing angle	2.4°
12	θ_{13}	CKM 1-3 Mixing angle	0.2°
13	δ	CKM CP violation Phase	0.995
14	g_1 or g'	U(1) gauge coupling	0.357
15	g_2 or g	SU(2) gauge coupling	0.652
16	g_3 or g_s	SU(3) gauge coupling	1.221
17	θ_{QCD}	QCD vacuum angle	~ 0
18	v	Higgs vacuum expectation value	246 GeV
19	m_H	Higgs mass	125 GeV

confirmed the discovery, shown in Figure 1.3. The mass of the Higgs boson is given by the Particle Data Group [26] as, $m_H = 125.25 \pm 0.17$, and a summary of the mass measurements resulting to this value is shown in Figure 1.4. The Higgs boson was also shown to have a spin-parity, $J^P = 0^+$ [27]. The results of the Higgs boson searches showed good agreement with the Standard Model predictions.

The phenomenology of the Higgs boson at the LHC (the LHC is explained in chapter 2), has been thoroughly studied [30–32]. Four main production modes of the Higgs boson at the hadron colliders have been discovered. The Feynman diagrams corresponding to these production mechanism are shown in Figure 1.5, and the cross sections at the proton-proton (pp) collisions at the LHC is shown in Table 1.3 in 4 different production mechanism. The production modes other than these four have very small cross sections.

In the ***gluon fusion production*** mode (ggF), shown at Figure 1.5a, two gluons produce a Higgs boson via a fermion loop but this fermion is usually the top quark since the Higgs boson couples to massive particles more often. This production mode has a cross section of 48.6 pb at $\sqrt{s} = 13$ TeV and is one order of magnitude larger than

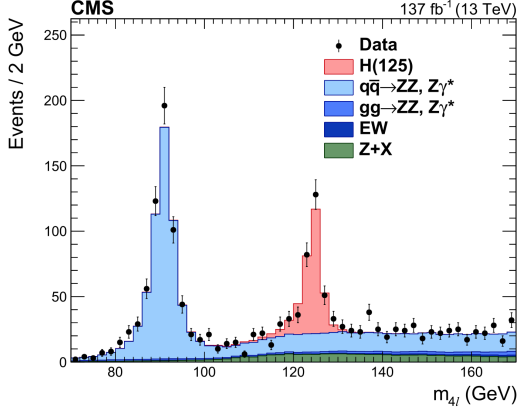


Figure 1.3a : m_{4l}

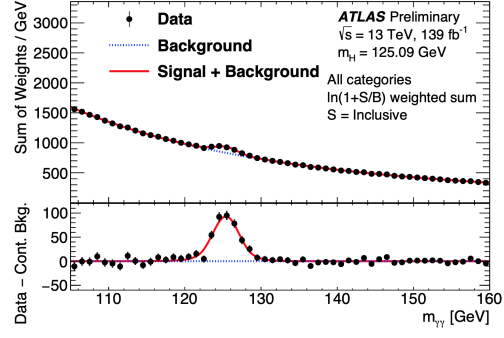


Figure 1.3b : $m_{\gamma\gamma}$

Figure 1.3 : Four-lepton mass distribution, m_{4l} obtained from the data collected at the CMS Detector [28] on the left, and diphoton invariant mass distribution obtained from the data collected at the ATLAS Detector [29] on the right, both at $\sqrt{s} = 13$ TeV in Run II.

Table 1.3 : Cross section (in pb) of the Higgs boson production at different centre-of-mass (\sqrt{s}) energies for $m_H = 125$ GeV. The theoretical uncertainties can be found in [26].

\sqrt{s} (TeV)	Production mode				
	ggF	VBF	VH	$t\bar{t}H$	Total
1.96	0.95	0.065	0.209	0.004	1.23
7	16.9	1.24	0.92	0.09	19.1
8	21.4	1.60	1.12	0.13	24.2
13	48.6	3.78	2.25	0.5	55.1
14	54.7	4.28	2.5	0.61	62.1

the second likely production mode at the proton-proton collisions. The Higgs boson produced in this mechanism is expected to have a small transverse momentum which can ease the event selection to separate the Higgs signal from background processes.

The **vector boson fusion production** mode (VBF) has a cross section of 3.78 pb at $\sqrt{s} = 13$ TeV, whose Feynman diagram is shown in Figure 1.5b. This production mode scatters the quarks out of the protons and these quarks help characterising this production mode with two additional jets in the final state. This process can be observed in the particle detectors as two back-to-back jets that carry high transverse momentum at the forward regions.

WH and ZH associated production mechanism of the Higgs boson (VH), or the **Higgs strahlung**, is the process of producing a W^\pm or Z boson where a Higgs boson is

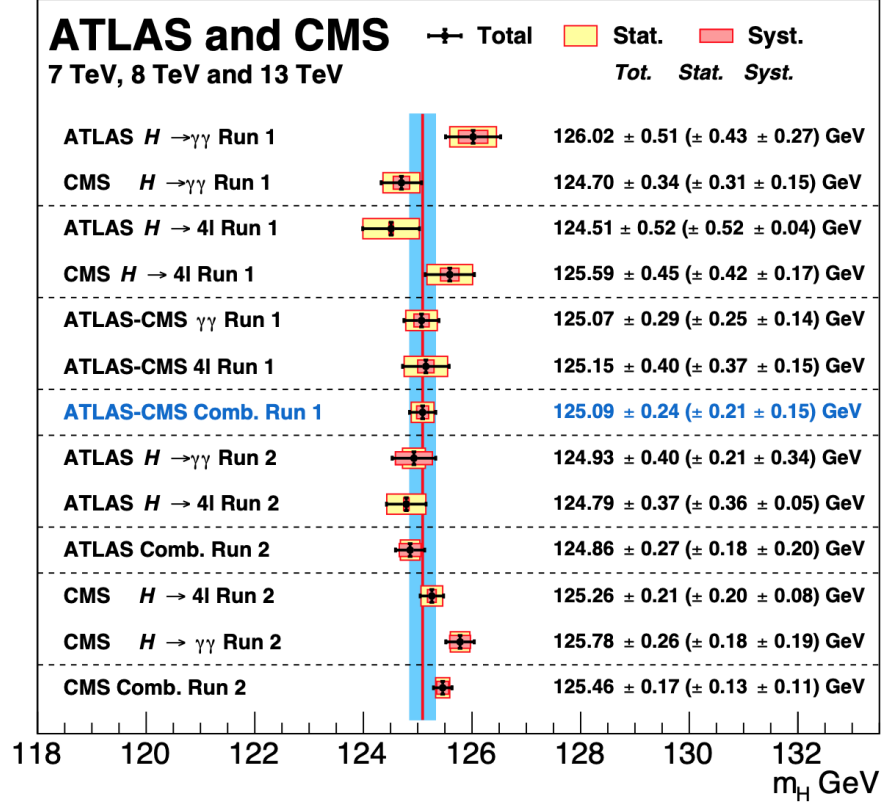


Figure 1.4 : Summary of the CMS and ATLAS Higgs mass measurements in the $\gamma\gamma$ and ZZ channels in Run I and Run II [26].

radiated away, as shown in Figure 1.5c. This production mode has a cross section of 2.25 pb at $\sqrt{s} = 13$ TeV in pp collisions. The vector boson in the final state is a good signature for signal and background separation when decayed leptonically.

The fourth and final production mode is the *associated production with top quark pair* ($t\bar{t}H$) whose leading order Feynman diagram is shown in Figure 1.5d. This production mechanism is a rare process in pp collisions with a cross section of 0.5 pb at $\sqrt{s} = 13$ TeV. It is observed at the CMS Experiment in 2018 [33]. The final state objects in this production mode, the top pair and the Higgs boson, can be identified by the large multiplicity of jets and leptons, however due to the lower cross section and the complex final state objects, this production mode is usually not included in the analyses and the same is valid for this thesis. When all of the standard model processes considered, the total cross section for the Higgs boson production makes it a rare process. The cross sections of the production modes of Higgs as a function of \sqrt{s} is shown in Figure 1.6a.

The decay modes of the standard model Higgs boson around its mass region is shown in Figure 1.6b, and branching ratios are given in Table 1.4. Since the decay channels

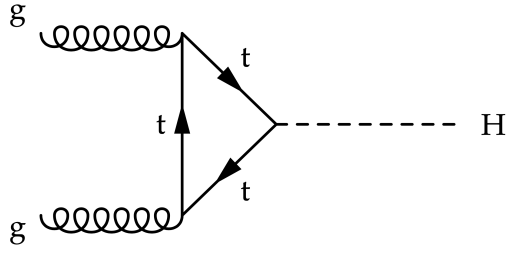


Figure 1.5a : ggF

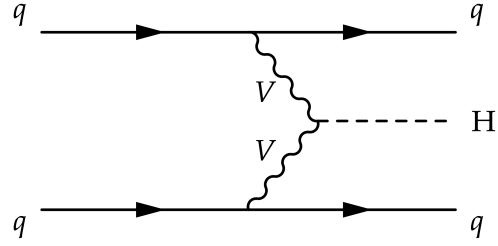


Figure 1.5b : VBF

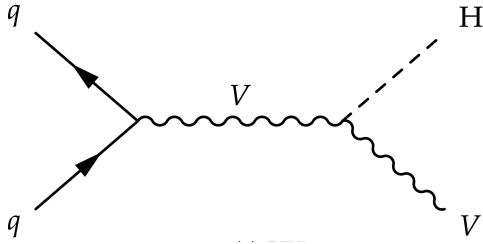


Figure 1.5c : VH

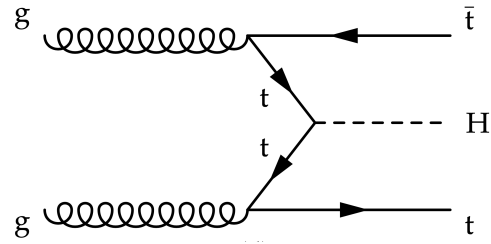


Figure 1.5d : $t\bar{t}H$

Figure 1.5 : Leading order Feynman diagrams of Higgs boson production at the LHC. The gluon fusion (ggF), vector boson fusion (VBF), Higgs strahlung (VH) and associated production with top quarks ($t\bar{t}H$) are shown. In the diagrams, q denotes any quark, t denotes the top quark, V denotes any of the Z or W^\pm bosons.

of the Higgs is strictly dependent on its mass, the branching ratios vary too much, as can be seen in Figure 1.7. At high Higgs masses, the decays into the W^\pm and Z bosons dominate the spectrum. On the other hand, the decays in the lower mass region where $m_H \leq 200 \text{ GeV}$ needs extensive searches since the decays become complicated. The decays in this region is dominated by the $b\bar{b}$ channel and the decay to $t\bar{t}$ is not observed since the top quark pair is too heavy. An interesting result is that the decays to W^+W^- and to ZZ channels have non-negligible branching ratios much below their masses. In such cases, one of the decay products is produced off-shell (V^*) while the other is expected to have its nominal mass. Regarding the decays to vector bosons, the $H \rightarrow 4l$ is often called *the golden channel* because of the leptonic decay of the intermediate vector boson, where the background contribution from QCD interactions is highly reduced.

The diphoton decay of the Higgs boson occurs via a top quark or W boson loops. Although this decay channel is rare - its branching ratio is 2.27×10^{-3} - it is a favourable decay mode since photons do not decay and can be reconstructed with a high accuracy in detectors.

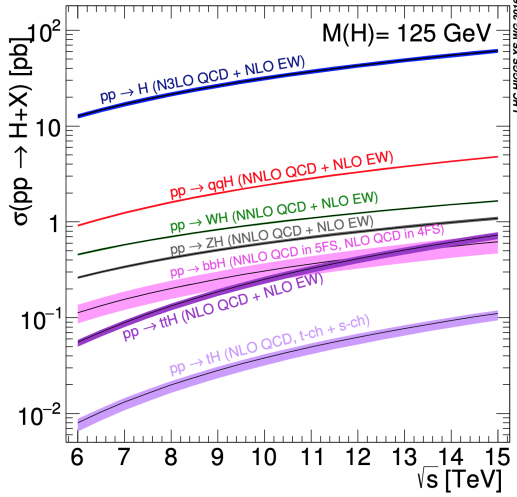


Figure 1.6a : σ_H

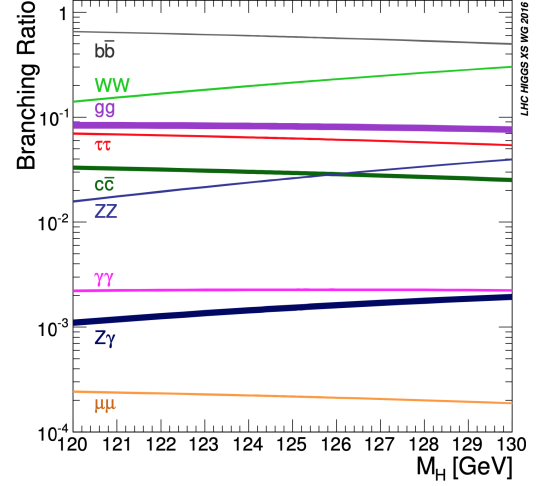


Figure 1.6b : m_H

Figure 1.6 : Higgs boson production cross section as a function of different production mechanisms (left) [32]. Branching fractions the Higgs boson as a function of m_H (right) [34]. The theoretical uncertainties are indicated as bands in both plots.

Table 1.4 : The branching ratios and the relative uncertainty for a SM Higgs boson with $m_H = 125$ GeV [32, 34].

Decay channel	Branching ratio	Rel. Uncertainty
$H \rightarrow \gamma\gamma$	2.27×10^{-3}	2.1%
$H \rightarrow ZZ$	2.62×10^{-2}	$\pm 1.5\%$
$H \rightarrow W^+W^-$	2.14×10^{-1}	$\pm 1.5\%$
$H \rightarrow \tau^+\tau^-$	6.27×10^{-2}	$\pm 1.6\%$
$H \rightarrow b\bar{b}$	5.82×10^{-1}	+1.2% -1.3%
$H \rightarrow c\bar{c}$	2.89×10^{-2}	+5.5% -2.0%
$H \rightarrow Z\gamma$	1.53×10^{-3}	$\pm 5.8\%$
$H \rightarrow \mu^+\mu^-$	2.18×10^{-4}	$\pm 1.7\%$

Another type of the decay modes of the Higgs boson is the leptonic decay; it has been first established in the $\tau^+\tau^-$ decay mode [35], further confirming the previous combined result from the ATLAS and CMS Experiments [36]. In the recent years, the decay of the Higgs boson to a muon pair has been studied and an excess of events is observed in data corresponding to an integrated luminosity of 137 fb^{-1} with a significance of 3.0 standard deviations [37].

The tau pair decay of the Higgs boson, with a branching ratio of 6.27×10^{-2} , can be observed in four production mechanisms that we have covered hence can have access to Yukawa couplings to fermions. Another importance of the $H \rightarrow \tau\bar{\tau}$ channel is to be the pioneering channel amongst the leptonic decays of the Higgs since the tau lepton is the heaviest of all known leptons. This feature of the $\tau\bar{\tau}$ channel may pave the way

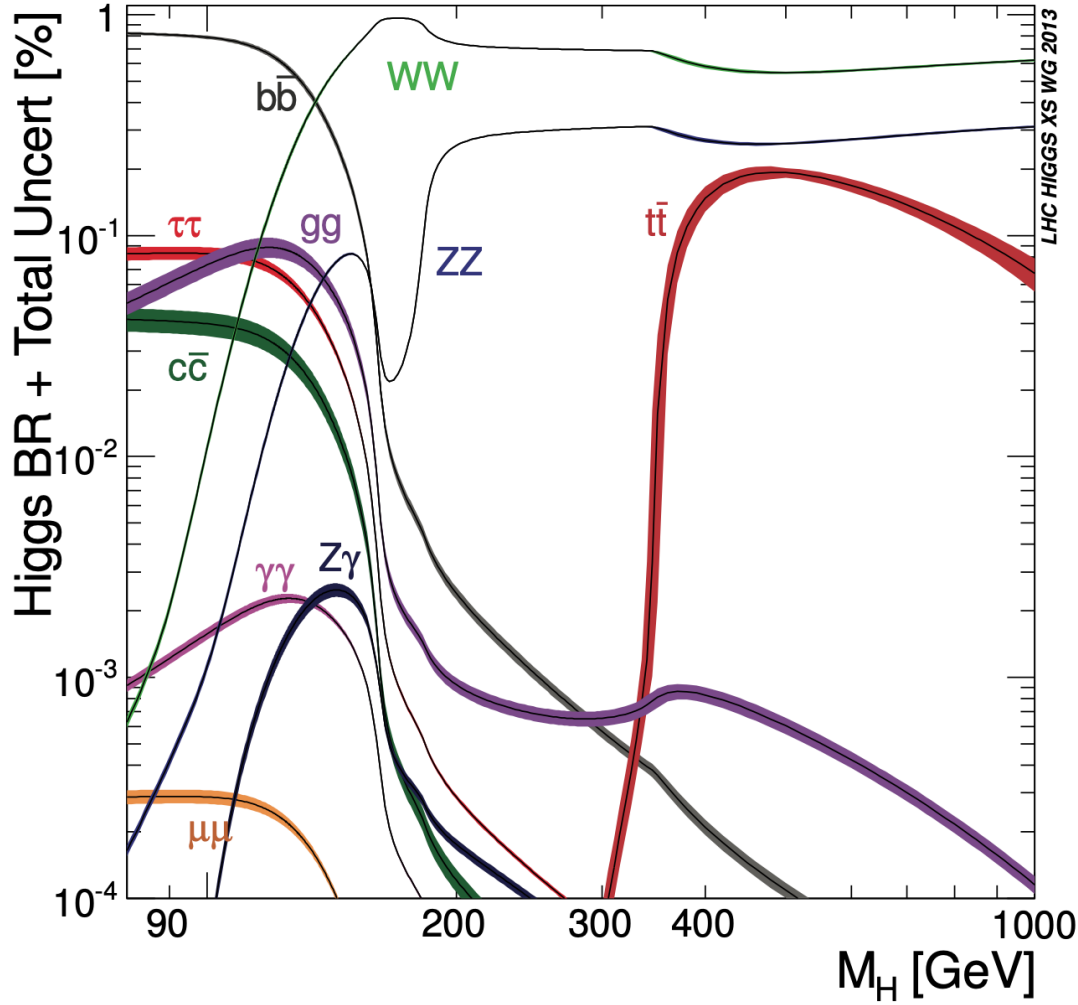


Figure 1.7 : Branching ratio of the Higgs boson decaying into various final states as a function of the Higgs boson mass [32].

for any difference between the Higgs mechanisms to lepton and to quarks. Also, the final state particles of this decay mode have access to the Higgs properties such as spin and parity, but the reconstruction of the hadronic taus can be elaborate.

1.2.1 Higgs boson pair production

The trilinear self-coupling of the Higgs boson was studied long before its discovery in 1988 [38], emphasising that it can be measured from the Higgs boson pair production. There are other processes that contribute to the double Higgs boson production via different production mechanisms. The Higgs boson pairs are produced at the LHC by the similar mechanism as the Higgs boson productions.

The **gluon fusion**, similar to the to the single Higgs gluon fusion production mechanism explained in section 1.2, starts with two gluons producing a Higgs boson

via either i) a heavy quark loop, which finally produces a Higgs boson pair through the trilinear self-coupling (triangle type) or ii) two on-shell Higgs bosons emerge from the heavy quark loop (box type), as can be seen in Figure 1.8. The relative contribution of the two processes can be seen in Figure 1.9. The cross section for this production mode is calculated up to next-to-next-to-leading order (NNLO) in QCD perturbation theory [39].

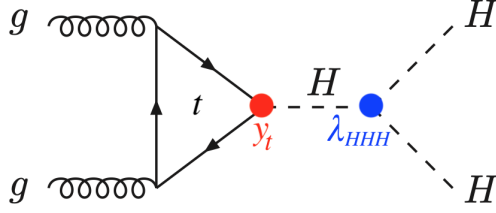


Figure 1.8a : Triangle type

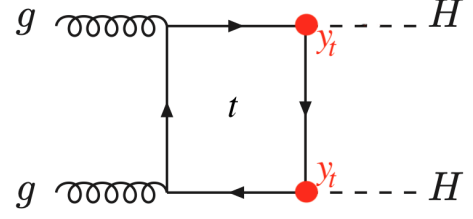


Figure 1.8b : Box type

Figure 1.8 : Leading order Feynman diagrams contributing to Higgs boson pair production through (a) the self coupling of the Higgs boson and (b) through a top or bottom quark loop.

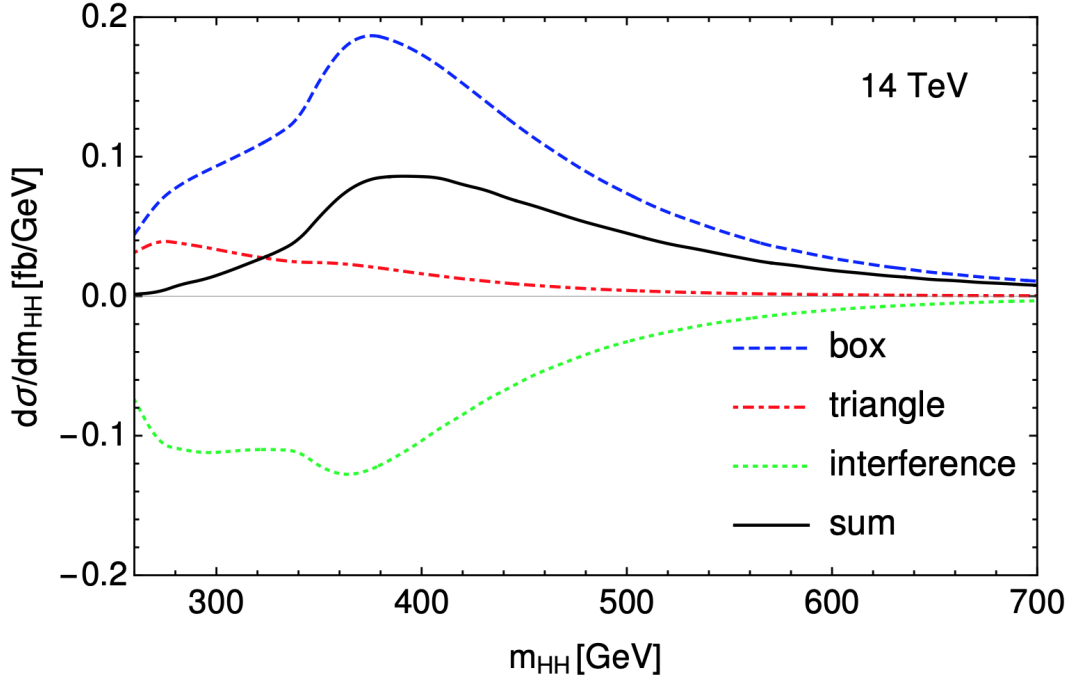


Figure 1.9 : Invariant mass distribution of Higgs boson pairs at leading order for the triangle and box-type gluon fusion production mechanism and their interference. The trilinear self-coupling reduces the cross section about a half w.r.t. the box-type process due to the large destructive interference [40].

The **vector boson fusion**, is the second production mechanism of double Higgs in terms of cross section. This production mechanism is dominated by t-channel W and Z boson exchange, similar to the single Higgs vector boson production mode. Three main Feynman diagrams contribute to this production; two Higgs bosons radiate off from the virtual W or Z boson shown in Figure 1.10a and Figure 1.10b, and a single off-shell Higgs boson splits into a Higgs pair Figure 1.10c. This production mechanism allows, in addition to the Higgs trilinear self-coupling, direct access to the Higgs boson coupling to a vector boson pair ($c_{2\nu}$) as well as to the Higgs boson coupling to two a single vector boson (c_ν). Two back-to-back jets with high transverse momentum is the characteristic of this production mechanism analogous to the single Higgs vector boson fusion production. Other production mechanism of Higgs boson pair involve *double Higgs strahlung* and *associated production* with a top quark or a top quark pair, whose cross sections are very small. The cross sections of these production mechanisms at different centre-of-mass energies are given in Table 1.5 and as a function of the centre-of-mass-energy Figure 1.11.

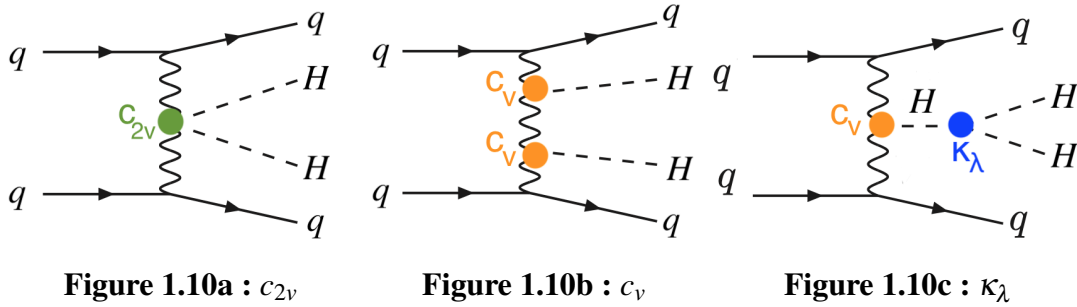


Figure 1.10 : Leading order Feynman diagrams of the double Higgs boson production via vector boson fusion mechanism.

Table 1.5 : Cross section (in pb) of the double Higgs boson production at different centre-of-mass (\sqrt{s}) energies for $m_H = 125$ GeV. The theoretical uncertainties including the QCD factorisation and renormalisation scale (α_s) uncertainties, Parton Distribution Function (PDF) uncertainty and effects from the top quark mass at NNLO can be found in [34].

\sqrt{s} (TeV)	Production mode cross section (fb)				
	ggF	VBF	$t\bar{t}HH$	VHH	Total
8	10.15	0.459	0.174	0.355	11.14
13	33.49	1.62	0.772	0.864	36.74
14	39.59	1.95	0.949	0.979	44.47

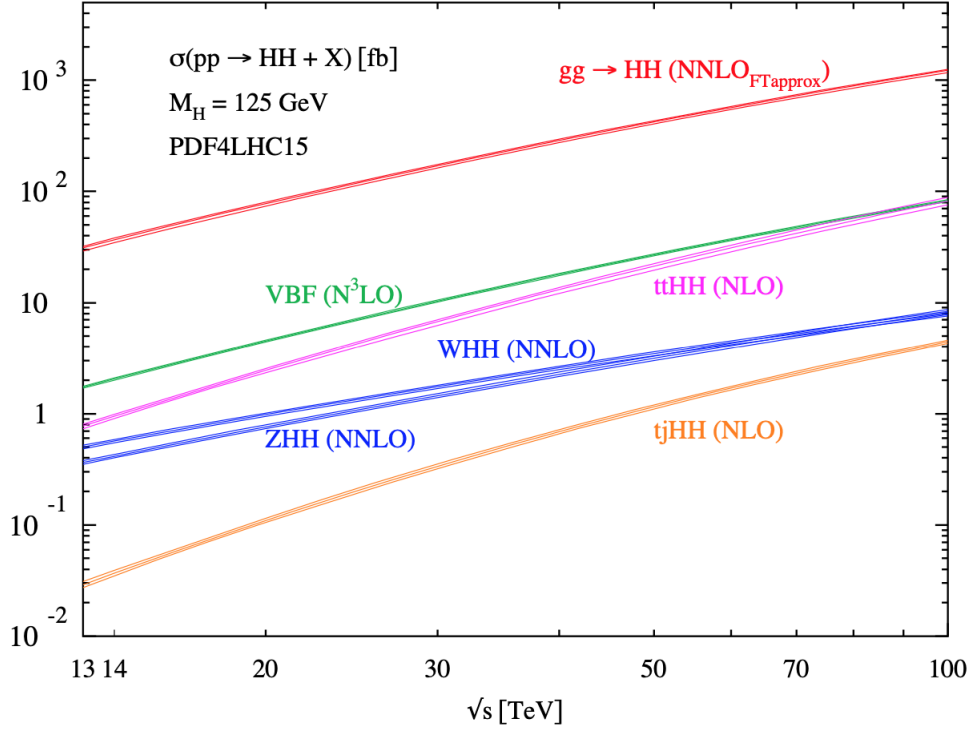


Figure 1.11 : Total cross section of double Higgs production mechanisms as a function of the centre-of-mass energy with $m_H = 125$ GeV within the SM. The gluon fusion remains the dominant production mode at high energies. The VBF continues to be the second dominant production mode while $t\bar{t}HH$ contribution starts to be in the order of VBF at about 100 TeV. The width of the lines shows indicate the total uncertainties coming from the scale and PDF + α_s uncertainties [40].

The double Higgs boson production is very rare at the LHC, hence the experimental searches usually include first two dominant production mechanisms. As mentioned earlier, the VBF production allows the direct access to both the Higgs trilinear self-coupling and the Higgs coupling to vector bosons. The destructive interference of the vector boson fusion production along with the restricted phase space of the double Higgs, makes the double Higgs production very sensitive to BSM physics from which the contributions might require modifications to the theory which can be observed at the LHC data at hand.

The SM expectations or the BSM models can be verified/rejected depending on the results of the decay mode searches of double Higgs production. The ATLAS and CMS Collaborations have discovered a rich variety of channels of double Higgs production. At Run I, these two collaborations have searched for both resonant and non-resonant productions in the following channels: $b\bar{b}\gamma\gamma$, $b\bar{b}\tau^+\tau^-$, $b\bar{b}b$, $b\bar{b}4l$, $WW^*\gamma\gamma$, $\tau^+\tau^-\gamma\gamma$

and in some other final states [26]. Most of these channel are updated in Run II by both ATLAS [41] and CMS [42] Experiments.

The HH production processes are very rich in final states given the SM Higgs boson decays. The possible final states for HH processes are shown in Figure 1.12 by a simple combination calculation using branching ratios of single Higgs boson. Searches for double Higgs at the LHC focus on decay channels with large branching ratios since the HH production process itself has a rather small cross section at the pp colliders. A similar search is performed by the CMS experiment using Run II dataset and resulted in an upper limit as 22.2 times the SM cross section, shown in Figure 1.13. However other decay channels, such as $WW\gamma\gamma$, $WWWW$, $\tau\tau\tau\tau$ and $\tau\tau\gamma\gamma$ can be probed and is needed to set up a tighter upper limit on the cross section calculations using collision data to broaden our knowledge on the Higgs boson pair production.

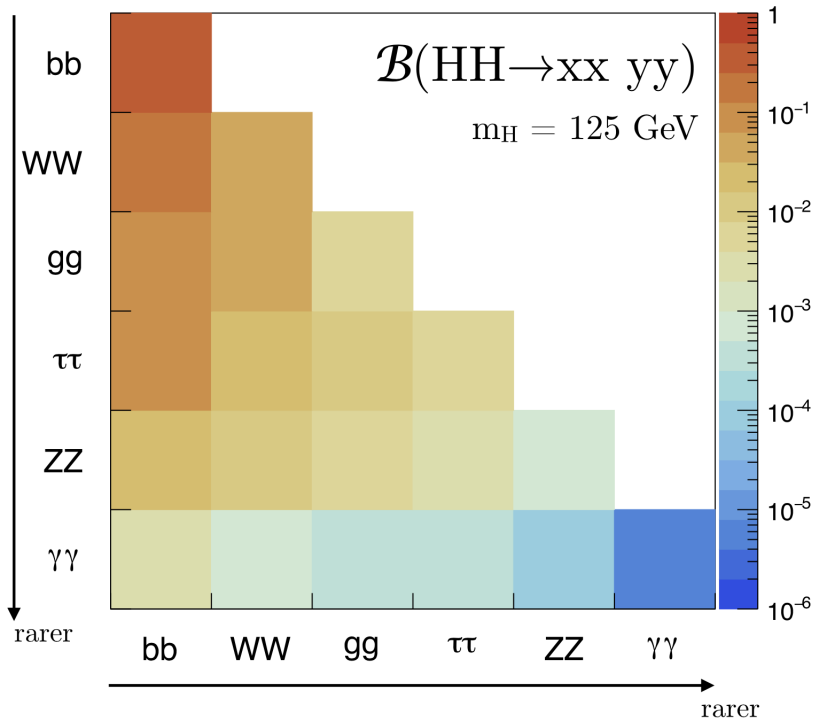


Figure 1.12 : Branching ratios of major double Higgs boson decay channels assuming a Standard Model Higgs boson [43]. From top to bottom and left to right, less possible decay mode combinations are depicted. The logarithmic scale bar on the left indicates the branching ratio along with colours.

1.2.2 BSM models

It is well known that the SM is one of the most successful theories in physics. However, the experiments show that there are some phenomena which the SM cannot explain.

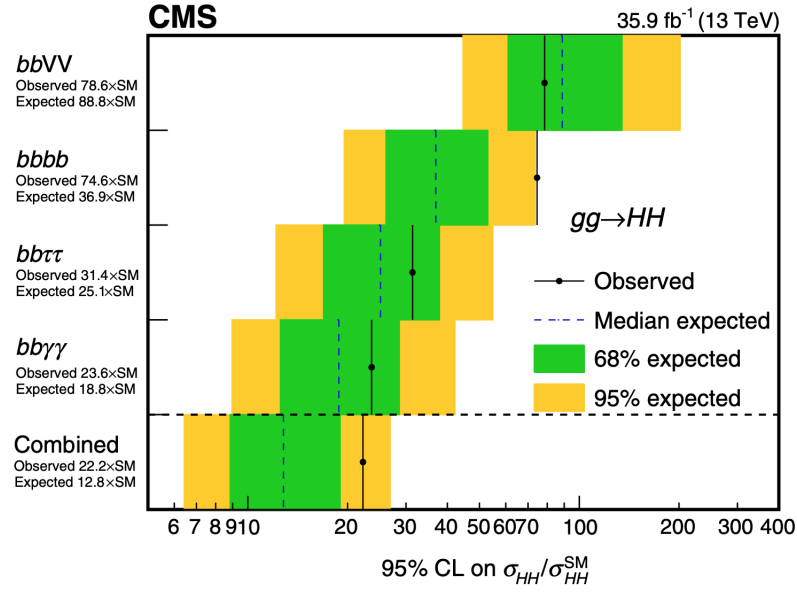


Figure 1.13 : The 95% CL upper limits on the signal strength $\mu = \sigma_{HH}/\sigma_{HH}^{SM}$ for $b\bar{b}(VV/b\bar{b}/\tau^+\tau^-/\gamma\gamma)$ channels along with with combined result. The inner green bands show 68% and the outer yellow bands show the 95% CL upper limits on μ under the background-only hypothesis [42].

The main topics that are beyond the SM regarding the Higgs boson can be given in three main section.

The first phenomenon is the existence of three generations/families of fermions. The particles across the families are identical under many aspects however the Higgs couplings to them differ in several orders of magnitude and the SM is incapable of answering to this asymmetry. The second phenomenon is the mass of the Higgs boson itself, one of the free parameters of the SM shown in Table 1.2. It is dependent on the radiative corrections and is not stabilised by any fundamental theory. Finally, the effectiveness of the theory up to the Planck scale can only be assured as far as the Higgs potential is bounded below for the vacuum to be stable. The stability of the potential at high energies is also associated with its conceivable role in the inflation theory of the Universe [44]. The model of Higgs as an inflation and the stability of the vacuum depend on the shape of the Higgs potential which is determined by the running self-coupling parameters λ . Therefore, it is reasonable to regard the SM as a manifestation of an unfolded embracing theory. The existence of the BSM physics could bring a solution to these problems.

Double Higgs boson production in this regard is a good handle to test our understanding of the Nature. If the new physics is at the energy scale of the LHC,

new heavy states can be observed which then decay to a Higgs boson pair. This type of the production mechanisms is called the ***resonant production*** and it shall manifest itself as an increase in the double Higgs cross section at a particular double Higgs boson invariant mass. The theories involving the existence of such particles are subject to modify the Higgs boson couplings. A considerable outcome is expected in such theories; the multiplet extension of the scalar Higgs field [45], extra Higgs doublet models, namely Two-Higgs-Doublet Model (2HDM) [46, 47], and composite Higgs models [48–50].

The ***non-resonant production*** of Higgs pairs in the BSM models can contribute in the quantum loops in the Feynman diagrams for HH production hence can modify the kinematics of the process and increase the cross section, but without causing a resonance that decays to a Higgs boson pair, shown in Figure 1.14.

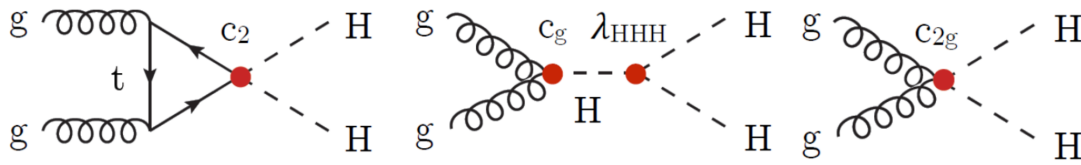


Figure 1.14 : Feynman diagrams that contribute to HH non-resonant production at LO with pure BSM effects.

2. THE LHC AND THE CMS EXPERIMENT

European Organisation for Nuclear Research, best known as CERN, is established in 1954 by 12 European countries and is based at the Franco-Swiss border in northwest Geneva, Switzerland. Today, the organisation operates the largest particle physics laboratory in the world and has 23 member states [51] and many others that make up more than 11 thousand researches and more than 70 countries around the globe. CERN's main activity is to provide the particle accelerators and needed infrastructure for high energy physics research. The organisation hosts the LHC Experiment which is the world's largest particle collider. This chapter introduces the structure and operations of the LHC and the CMS Detector.

2.1 The Large Hadron Collider

The Large Hadron Collider consists of a 27 km ring tunnel where the beams of particles (protons or lead ions) are accelerated with the help of superconducting magnets along with a number of accelerating systems. It was built between 1998 and 2008 and it lies 175 metres beneath the surface with a total cost of the project expected to be on the order of 4.4 billion US\$. The initial design of the LHC aimed at the centre-of-mass energy of $\sqrt{s} = 14$ TeV with a nominal peak luminosity of $\mathcal{L} = 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ [52]. The LHC collides proton-proton beams as well as lead-lead (Pb - Pb), proton-lead (p - Pb) and Xenon-Xenon (Xe - Xe) nuclei to study heavy-ion collisions. The *LHC accelerator complex* consist of a series of accelerators and is used to accelerate protons before being injected in the LHC, which will be explained shortly.

2.1.1 Operation of the LHC

The success of the LHC required two decade-long international collaboration. The initial studies started in the early 1980s when the The Large Electron Positron Collider (LEP) at CERN was not even running. CERN Council approved the construction of the LHC in 1994 and a technical design report was published the next year. Starting

from 1998, the construction of the LHC was completed in 2008 and it succeeded to accelerate proton beams the same year [53]. The first accelerated proton beams had an energy of 450 GeV per beam, and later on the collisions at $\sqrt{s} = 2.14$ TeV with respect to Tevatron at 1.96 TeV made the LHC the highest-energy collider ever built. In 2010, LHC increased the beam energy to 3.5 TeV which was a world record of man-made particle acceleration.

Data collection started in 2010 and finished in 2013; this time period is known as Run I of the LHC. The data collected by the CMS Experiment corresponds to an integrated luminosity of 45 pb^{-1} , 6 fb^{-1} and 23 fb^{-1} at $\sqrt{s} = 7$ TeV in 2010, 2011 and 2012, respectively. The data collected at Run I has lead the discovery of the Higgs boson. LHC entered an upgrade stage called Long Shutdown 1 (LS1) for two years when Run I is finished in 2012. LHC was upgraded in LS1 in the way of achieving its design performances, and started its operation again in 2015 for a period of 3 years known as Run II but his time the beam energy was 6.5 TeV. The Run II is finished in 2018 with a data amounting to about 150 fb^{-1} and LHC entered LS2 with a schedule to be in Run III in 2022 at $\sqrt{s} = 14$ TeV. The LS3 is planned to be the upgrade period where the LHC will have an unprecedented instantaneous luminosity corresponding to an integrated luminosity of 3000 to 4000 fb^{-1} eventually, called the High-Luminosity LHC (HL-LHC) or Phase II. Both the LHC and the HL-LHC plans shown in Figure 2.1.

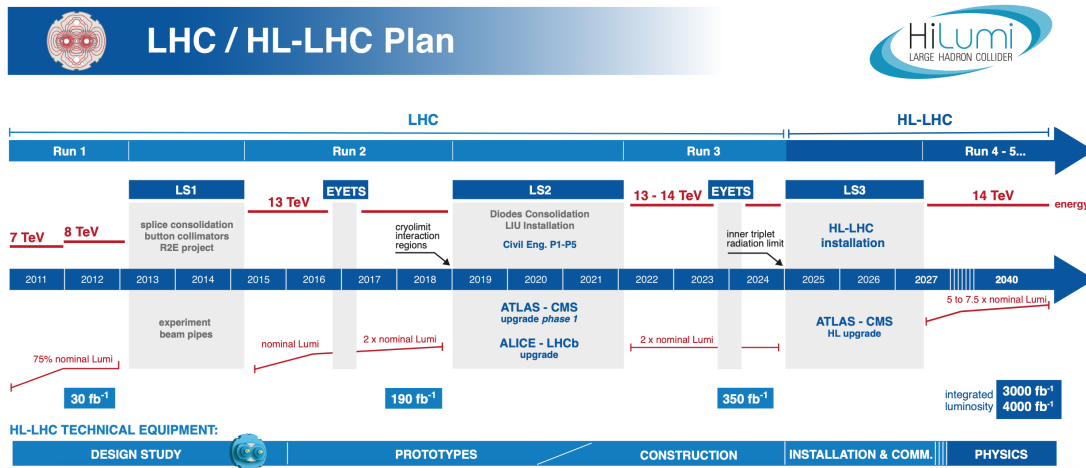


Figure 2.1 : A detailed schedule of LHC and HL-LHC showing the integrated luminosity and the beam energy corresponding to each period [54].

The main scope of the HL-LHC is to increase the collision data which will allow physics searches to be more statistically abundant and be able to perform higher-precision measurements.

2.1.2 The accelerator complex

The accelerator tunnel of the LHC, which was previously used host LEP collider, has two parallel vacuum pipes where two counter-rotating beams are kept inside a magnetic field generated by superconducting niobium-titanium (NbTi) cables. The magnetic field generated to steer the beams are about 8 Tesla which is more than 100 000 times higher than the Earth's magnetic field. This field is generated by 1232 dipole magnets each with a 14 metres of length and 35 tonnes of weight where 11 thousand Amperes of electric current flows. This acceleration system allow the beams to circulate with 7 TeV energy. The focusing of the beams in a narrow area is secured by 392 quadrupole magnets with 5 to 7 m lengths. In order to inject the beams in the collision points, special quadruples are positioned at each entrance to squeeze the beams in a narrower area. These superconducting magnets are cooled down to a temperature of 1.9 K by a cryogenics cooling system supplied with 120 tonnes of Helium-4 fluid.

Before being injected into the LHC, the proton beams are pulled off from hydrogen gas and accelerated by a series of systems gradually increasing their energy, presented in Figure 2.2. Firstly, protons are accelerated to an energy of 50 MeV in the Linear Accelerator (LINAC2) then injected into the Proton Synchrotron Booster (PSB) where beams are re-accelerated to 1.4 GeV. These particles are sent to the Super Proton Synchrotron (SPS) to further accelerate them to 450 GeV energy. The final injection is made from the SPS to LHC's two beam pipes in the counter directions. The filling of the LHC by protons takes about 8 minutes, and 20 minutes for protons to be accelerated to 6.5 TeV of energy in bunches via Radio Frequency (RF) cavities operating. The injected beams circulate the LHC rings for many hours (~ 12 hours) under normal operation. The bunches are spaced by 25 ns or 7.5 metres thus the bunch crossing rate is 40 MHz. Nominal number of protons per bunch is 12×10^{11} and the nominal number of bunches per beam is 2808 contributing to an inelastic collision event in the order of 10^9 per second with about 20 collisions per bunch crossing.

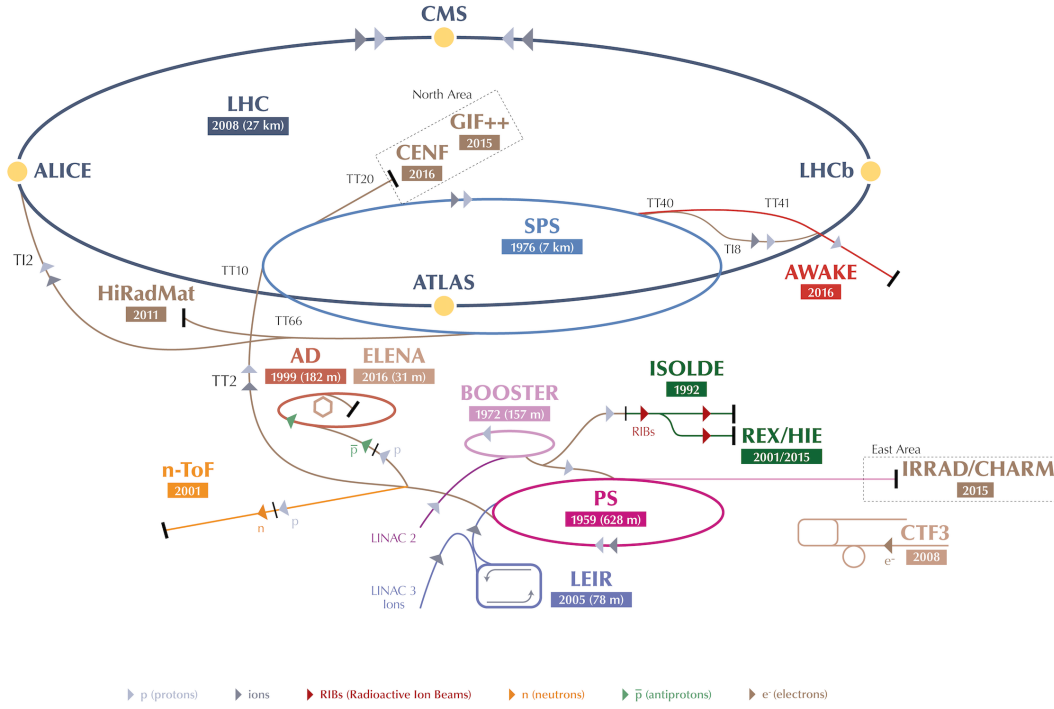


Figure 2.2 : The LHC accelerator complex. The acceleration of the protons start in the LINAC2 and ends in LHC through Booster, PS and SPS [55].

The accelerated beams of particles are collided inside one of the four detectors; ALICE, ATLAS, CMS and LHCb. The ATLAS and CMS detectors are installed in opposite sides, at Point 1 and Point 5 of the LHC. These two detectors are designed as multi-purpose detectors that surround the collision points to detect any out-coming particle. ATLAS experiment is described in the reference [56] and the CMS experiment is explained in detail in section 2.2. ALICE experiment is installed at Point 2 and its main purpose is to study heavy ion collisions and quark-gluon plasmas. Final of the four large experiments, LHCb, located at Point 8, is a forward one sided detector mainly aimed at measuring the charge conjugation and parity symmetry violation in Beauty baryons. The detailed description of the two detectors can be found in [57,58].

LHC host many other experiments; the LHCf [59] and TOTEM [60] experiments positioned at 100 meters away from the both sides of the ATLAS and CMS's collision points. These experiment study mainly the pp interactions and forward physics. Others are the MoEDAL [61] experiment which is dedicated to search for magnetic monopoles at the same experimental cavern with LHCb detector, and the FASER

experiment searching for lighter particles and studying neutrinos situated 480 metres away from the ATLAS's collision point.

2.1.3 Design and specifications

The **collision energy** of the particles inside the detectors is one of the most important parameters at the LHC and is simply the sum of the energy of two colliding beams. When these collisions happen, independent types of interactions may happen between protons. The **soft interactions** signify a small amount of transferred momentum which means the interacting protons barely came close to each other and escapes without decaying at all. The **hard interactions** on the other hand, signify a high transferred momentum which causes the proton to decay. When two protons collide, a fraction of the energy (s') contributes to the interaction since it is actually the partons of the protons that participate in the collision proportional to the fraction of energies (x_1, x_2) of each parton.

$$\sqrt{s'} = \sqrt{x_1 x_2 s} \quad (2.1)$$

Another important parameter is the **instantaneous luminosity** for the LHC apparatus. It is a description of the number of collisions per time and cross section. It is given by the following formula,

$$L = \frac{N_b^2 n_b f_{rev} \gamma}{4\pi \epsilon_n \beta} F, \quad (2.2)$$

where N_b is the number of particles in each bunches of n_b bunches circulating in the accelerator ring with f_{rev} frequency. The parameters γ , ϵ_n , β and the factor F denotes the relativistic Lorentz factor; emittance and focal length describing the shape and focus of the beam, and geometrical reduction which depends on the angle between two beams, respectively. The luminosity lifetime τ on the other hand shows how luminosity decreases with time,

$$L = L_0 e^{-t/\tau}, \quad (2.3)$$

where L_0 is the peak luminosity at time zero.

The integrated luminosity, which is another important parameter, is given by $L_{int} = \int L dt$ and takes part in the total amount of collisions over time.

$$N = L \times \sigma \quad (2.4)$$

is the number of events produced for a particular physics process with the cross section σ . Thus, in order for a rare process to be observed in the detectors, maximising the luminosity is essential. In Figure 2.3, the cross sections as a function of \sqrt{s} is shown. It is obvious that the Higgs processes are several orders of magnitude lower in production cross section than the dominant processes, however since the cross section increases with the increasing centre-of-mass energy, it is necessary to increase also luminosity and \sqrt{s} in order to achieve higher event rates. The integrated luminosity delivered to the CMS experiment since 2010 is shown in Figure 2.4.

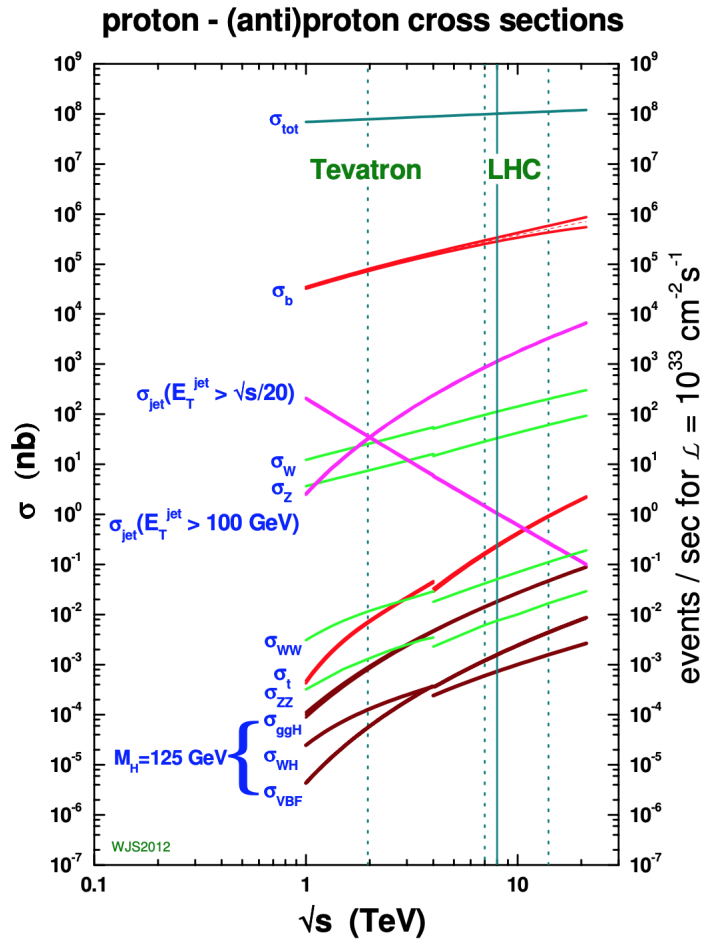


Figure 2.3 : Cross sections of Standard Model processes as function of collider energy at pp collisions [62].

The large instantaneous luminosity of the LHC causes a disadvantage. Multiple pp collisions happen at each bunch crossing and many primary vertices are superimposed. Most of these parton interactions have relatively small centre-of-mass energy, hence they are not of much interest for the experiment, which is often called *in-time pileup (PU)* interactions. Therefore, the detector needs to resolve these pileup interactions

from the hard collisions. Besides, since the collisions take place too frequently (per 25 ns), it is inevitable that new particles reach the detector before the products of the previous bunch crossing escapes the detector. This type of pile up is called *out-of-time pileup* interactions. In order to overcome these demanding conditions, highly granular and fast-response detectors are needed. Moreover, the detector needs to be resistant to radiation as well as be precise in energy measurements. The Compact Muon Solenoid, an appropriate detector designed to overcome these challenges, is explained in the next section.

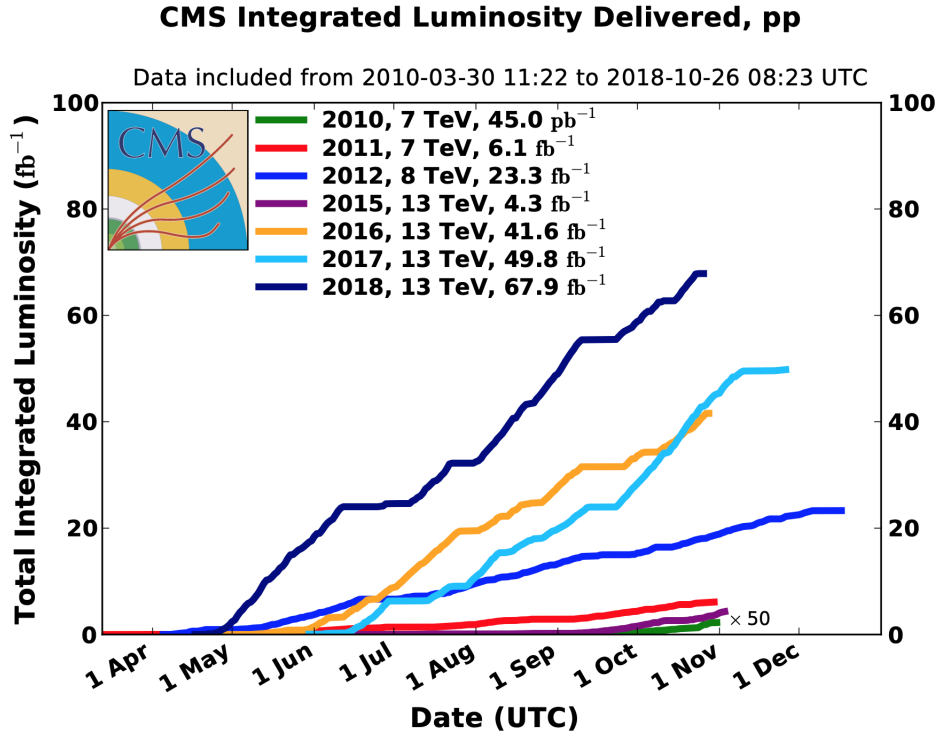


Figure 2.4 : Total integrated luminosity delivered by the LHC to the CMS detector for pp collisions at Run I-II [63].

2.2 The Compact Muon Solenoid Experiment

The CMS Collaboration, consisting of more than 4000 particle physicist, engineers, technicians and student from around 200 institutes and more than 40 countries, is a particle physics community in search for the fundamental building blocks of our universe. The collaboration operates the CMS detector and collects data from the collision of the protons supplied by the LHC. It is one of the two multi-purpose detectors at the LHC facility, and has a broad physics program. Each term in its name means one of the detector's main features; *Compact* signifies its small size for all the

detector material it contains, ***Muon*** emphasises the dedicated muon tracking system and ***Solenoid*** underlines the most powerful solenoid magnet ever made.

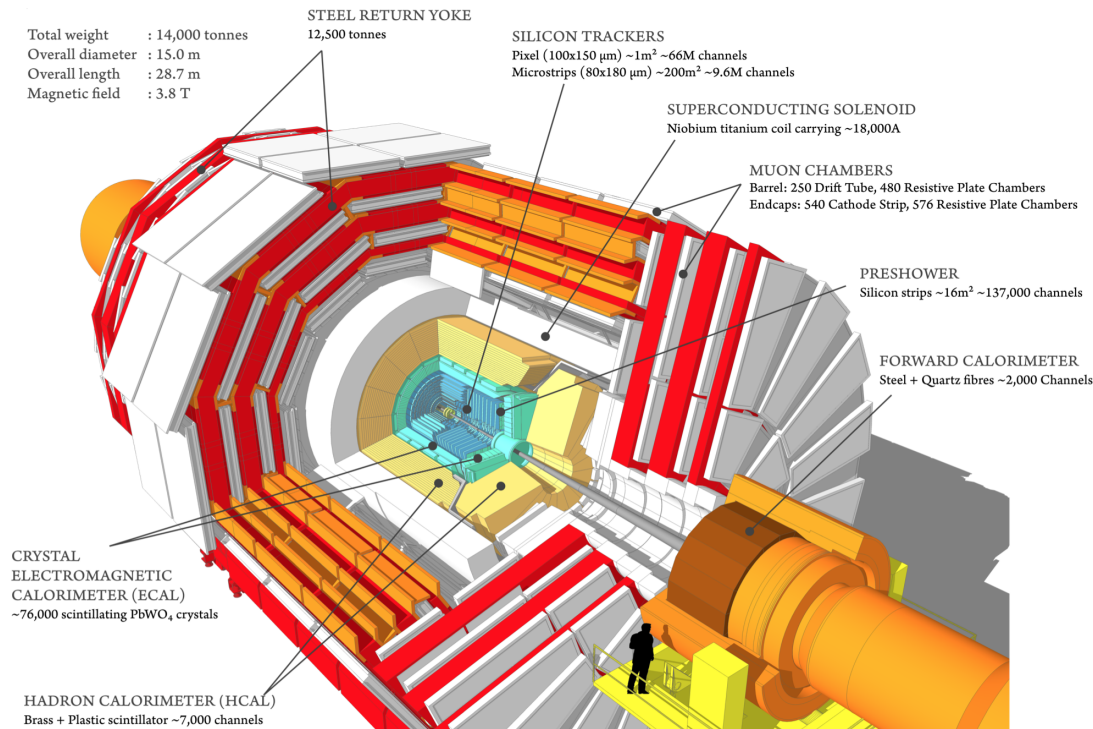


Figure 2.5 : Three dimensional view of the CMS detector showing each component. A human shape is positioned on lower right to compare its size. It is sometimes called as the cylindrical onion due to its shape and concentric layers of components [64].

2.2.1 Structure and the CMS coordinate system

The CMS detector has a cylindrical shape with a diameter of 15 m and a length of 21.5 m. It contains many detector components in that size that they weigh about 12 500 tonnes and that is another reason for calling it compact. The central region is called ***barrel*** and each of the forward regions are called ***endcaps***. It consists of many concentric layers of sub-detectors for many different purposes, shown in Figure 2.5. These layers include, from the innermost part towards outside; a silicon tracker, the electromagnetic calorimeter (ECAL), the hadron calorimeter (HCAL), the superconducting solenoid and the muon chambers. Each component is explained in the following sections.

The coordinate system of the CMS experiment is a right-handed cartesian system. It accepts the collision point as the centre with the x-axis pointing radially inward to the centre of the LHC's circle, y axis is perpendicular to x-axis and to the direction of the

beams which is the z-axis, shown in Figure 2.6. The z-axis points in the anticlockwise direction when looked from above the underground.

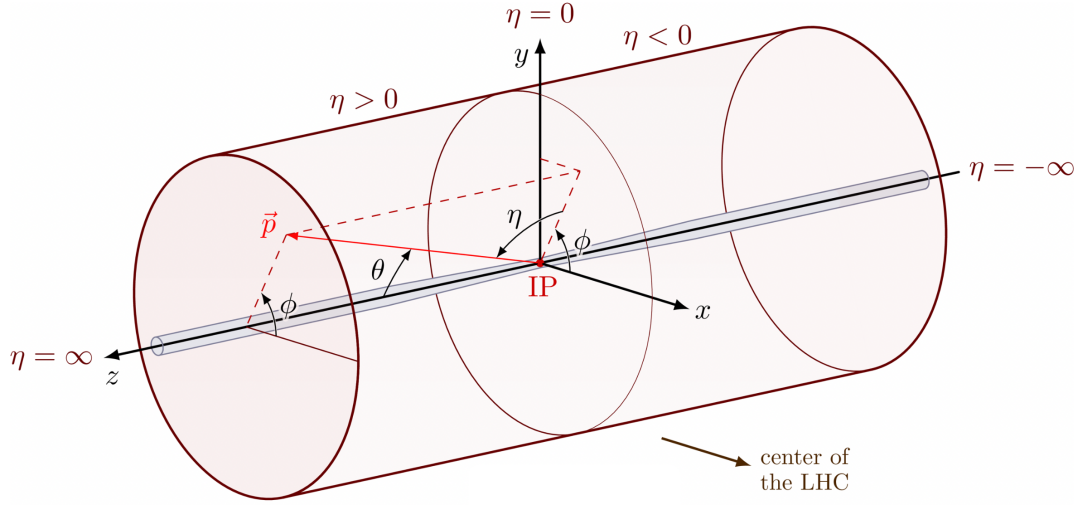


Figure 2.6 : The coordinate system of the CMS detector. IP denotes the interaction point and the momentum of a particle is shown by \vec{p} in red arrow. Pseudo-rapidity values are shown by η [65].

The cylindrical coordinates is often used for the CMS experiment because of the shape of the detector; an azimuth angle ϕ between x and y-axes in $[-\pi, +\pi]$, and a polar angle θ between z and y-axes in $[0, +\pi]$, respectively. Another kinematic quantity, the **pseudorapidity** which is a relativistic reference frame independent kinetic observable widely used at the collider physics and defined as $\eta = -\ln(\tan(\theta/2))$. The barrel region in this reference frame is simply in $|\eta| < 1.2$ region and two endcaps are in $1.2 < |\eta| < 3$ region.

Another kinematic variable, the **angular separation** or *angular distance* is defined as $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ and is used to separate particles using the cones around the direction of particles. It is also a boost invariant variable. The **transverse momentum** p_T is another important kinematic variable in the collider physics, and is defined as the projection of the particle's momentum and energy onto the transverse plane that is $\eta = 0$.

2.2.2 Solenoid magnet

The core feature of the CMS detector is the large niobium-titanium superconducting solenoid with 5.9 m inner diameter and 12.5 of length. It creates 3.8 T uniform magnetic field along the beam line axis while operating at 4.5 K with 18 kA of electric

current. The magnetic field bends the trajectories of emerging charged particles from the collisions in the transverse plane, see Figure 2.7. It is also useful for measuring the transverse momentum of particles up to TeV/c via the curvatures in their tracks.

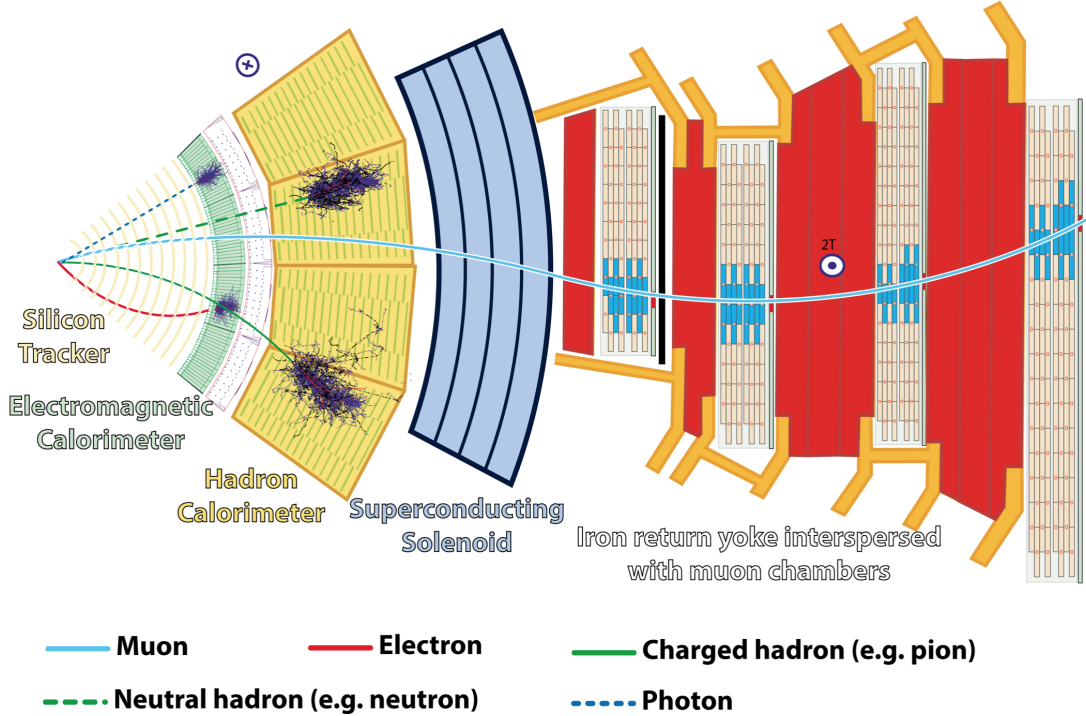


Figure 2.7 : Schematic view of a transverse slice of the barrel region of the CMS detector. Particles are shown in their trajectories and signatures in the sub-detectors. Photons and electrons create showers in the ECAL but photons don't leave traces in the tracking system which allows to distinguish these two types. The same recognition can be made between neutral and charged hadrons [66].

The solenoid hosts the tracker and calorimeter systems at its inner space and this requires these sub-detectors to be very compact. The returning magnetic field of the magnet is used to measure the transverse momentum using the muon chambers positioned inside the iron structure surrounding the solenoid and has a 2 T magnetic field. Hence the muons are bent in opposite directions depending on being inside and outside of the solenoid.

2.2.3 Inner tracking system

The first sub-detector system that the emerging particles encounter, is the inner tracking system in $\eta < 2.5$ region. The hitting points in the tracker system of outgoing particles from the interaction point are combined to create a high resolution charge

and momentum information. The resolution of these measurements decreases with the increasing transverse momentum since the curvature becomes more straight. The measurements in the inner tracking system also helps to determine the interaction point of hard scatterings, usually known as the *primary vertex*, and its separation from pileup in the event. The reconstruction of additional decays are also performed using the inner tracker, such as a B meson decay from the secondary vertex.

In order to provide a radiation durability, the material inside the inner tracking system is kept minimum since the system is the first to encounter the propagating particles from the interaction point, and to limit the energy loss of those particles via coulomb scattering, bremsstrahlung or nuclear interactions before they reach the aimed sub-detector. These requirements are achieved by fast readout on-board electronics for the whole sub-detector consisting two different silicon sensors sensitive to the charged particles shown in Figure 2.8.

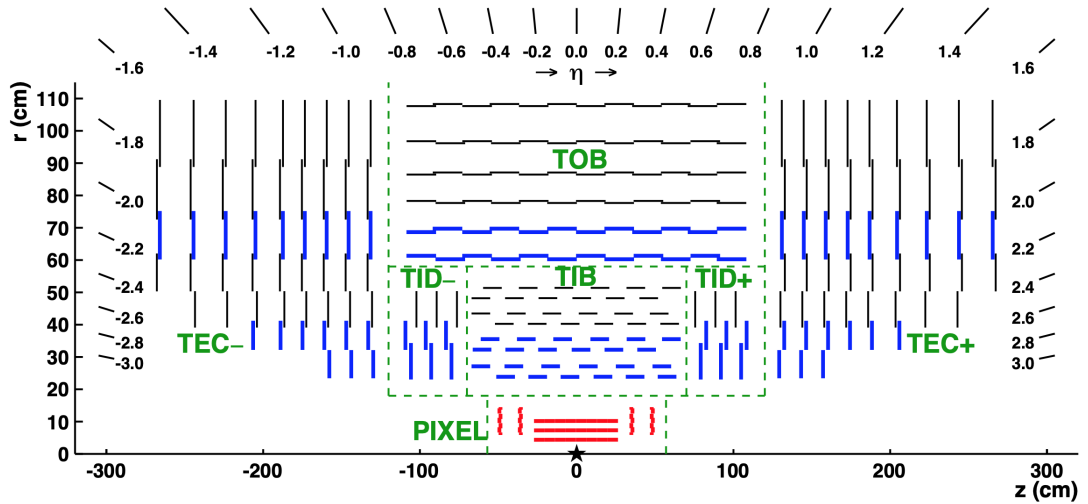


Figure 2.8 : Schematic cross section of the tracker in the r - z plane with one half shown since the tracker is symmetric about the $r = 0$. The star indicates the centre of the tracker and the expected collision point. The detector modules are kept in green dashed lines, namely; Tracker inner barrel (TIB) and Tracker inner disks (TID) surrounded by the Tracker outer barrel (TOB) and tracker endcaps (TEC) on the sides. Thin black lines show the strip tracker modules that create 2-D hits, while thick blue lines show the one allowing providing 3-D hit positions which is actually two back-to-back strip modules. The pixel modules provide 3-D hits and they are shown in red curly lines. Each module in a given layer is shifted slightly to cover all the gaps providing best acceptance. [67].

The inner tracker includes a *silicon pixel* sub-detector consisting of four barrel and three endcap pixel detectors. The endcap pixel detectors are in disk shapes and have pixel cells of $100\ \mu\text{m} \times 150\ \mu\text{m}$ size providing a resolution of $10\ \mu\text{m}$ in the transverse plane, and $20\ \mu\text{m}$ along the beam axis with the third coordinate being the position of each sensors. A total of 66 million pixels allow highly precise measurements and a high resolution of the impact parameter, the closest distance from the primary vertex to a given track, essential for second vertex reconstruction [67,68]. The silicons in these pixels are stimulated when a charged particle passes in them. The pixel detectors and the strip tracker operates at -15°C and -20°C , respectively, which reduces the radiation damage and dissipates the heat generated by the on-board electronics. An upgrade of the pixel detector has been performed during the technical stop in 2016 in order to retain its performance which reduced due to radiation damage [69].

The outermost part of the tracker consists of silicon micro-strips and is called the *silicon strip* sub-detector. It contains about 9.3 million strips covering a $198\ \text{m}^2$ area in the $\eta < 2.5$ region. It has 10 cylindrical layers in the barrel region close to the pixel detector; four layers of TIB and six layers of TOB, shown in Figure 2.8. Additionally, 24 disks of endcap modules are installed each containing concentric disks of silicon strip modules; three TIDs on the sides of TIBs, and nine TEC disks covering $124 < |z| < 282\ \text{cm}$. Figure 2.9 shows the tracking system's materials budget, which is basically the detector thickness in units of radiation length (the amount of material traversed by particles in the detector).

2.2.4 Electromagnetic calorimeter

The calorimeters are designed to measure the energy of both charged and neutral particles by their interaction with the detector material. The electromagnetic calorimeter at the CMS detector [71] performs the energy measurements of electrons and photons. It is placed between the tracker and the HCAL, and is highly granular made up of lead tungstate inorganic crystals (PbWO_4). The energy measurement is made by a photosensitive device in the ECAL, that is excited by the scintillation light coming from the interaction of the crystal material with an electromagnetic shower originating from an incident electron or muon.

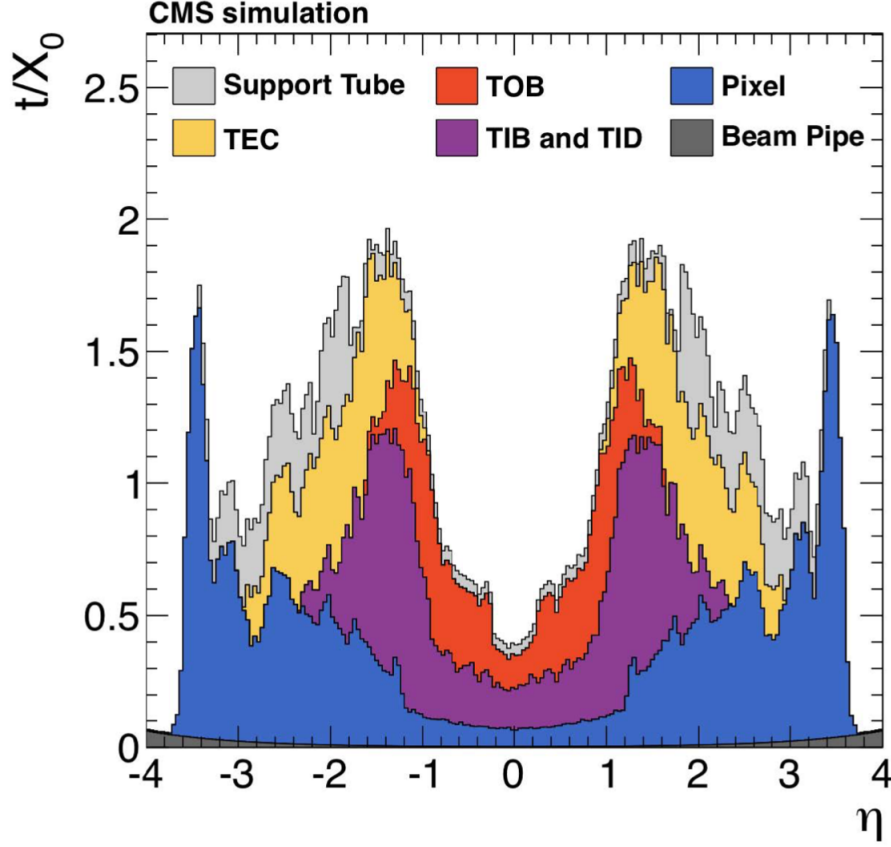


Figure 2.9 : The material budget contribution from sub-systems as a function of η , showing the tracker sub-detectors, beam pipe and the support tube that surrounds the tracker [70].

The ($PbWO_4$) is a good material choice here because of its short Molière radius, that is the radius approximately of a cylinder coaxial with the shower axis which contains 90% of the energy of the shower. It means that the crystal material can contain the EM shower. Besides, ($PbWO_4$) is good scintillating medium with a small radiation length which is useful for determining the amount of depth of material traversed by particles. Additionally, lead tungstate is a radiation hard material; almost 80% of the scintillation happens within 25 ns. The very short scintillation time is also an essential feature given the high luminosity of the LHC and bunch spacing of 25 ns. A drawback of this material is its low light yield, which requires using extra amplification.

The barrel of the ECAL (EB) contains 61200 crystals in $|\eta| < 1.479$ region and two endcaps (EE) are made up of about 7300 crystals covering $|\eta| < 3.0$ shown in Figure 2.10. The crystals in both EB and EE are aligned such that their axes confronts the nominal interaction point ensuring the escape of particles to a minimum. The scintillation light coming from the interaction of the particles with the crystals is

read-out by avalanche photo-diodes in EB and by vacuum photo-triodes in two EEs, and the ECAL as a whole operates at 18 °C.

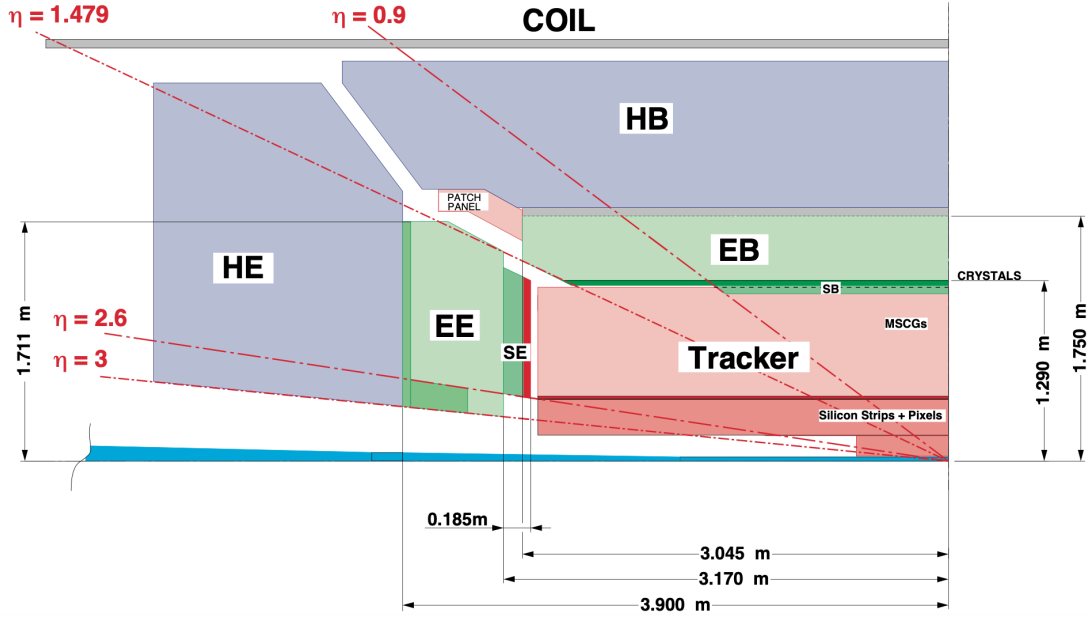


Figure 2.10 : Schematic view of one quadrant of the calorimetry and tracking system [71].

Two EM pre-shower (ES) sampling sub-detectors are installed in the front sides of the EEs, each containing a sampling calorimeter consisting two lead layers to initiate EM showers. These lead layers are followed by 2-mm silicon strip detectors to measure the transverse profile of the showers. They provide 2-D profiling of the showers thanks to their orthogonal placement.

The energy resolution of the ECAL barrel was measured in 7 TeV data using electron from $Z \rightarrow e^+e^-$ decays, resulting an energy resolution better than 2% in $|\eta| < 0.8$ and ranging from 2 to 5% in other regions [72]. Energy resolution formula for a calorimeter reads;

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S(\eta)}{\sqrt{E}}\right)^2 + \left(\frac{N(\eta)}{E}\right)^2 + C(\eta)^2 . \quad (2.5)$$

The energy resolution of a calorimeter increases with the particle energy because of the reduction in the first and the second terms in the equation. The first term is called **stochastic term** and is proportional to the number of photons caused by the scintillation in the interactions. The second term is the **noise term** and signifies the noise originating from the electronic readout systems. Finally, the third term, being a constant, depends on the detector inhomogenities, calibration uncertainties and some

instrumental effects. The values for ECAL are measured for ECAL barrel during a test beam with incident electrons as 2.8%, 41.5 *MeV* and 0.3%, respectively [73].

2.2.5 Hadron calorimeter

The HCAL [74] is designed to absorb hadrons coming from the collisions which usually pass the ECAL without being stopped. The energy of the hadrons are measured from their showers in the HCAL, which is a more difficult task than the one made in ECAL about the electrons and photons because of rare effects in the shower generation where many undetectable particles may be produced from nuclear and hadronic interactions. These effects reduce the energy resolution of the hadrons detected in the HCAL, on the other hand the particle flow reconstruction techniques can improve the resolution offline, explained in subsection 3.1.3.

The HCAL is at the heart of the reconstruction of jets produced in the final states. It is made up of two main sections; a barrel (HB) and two endcaps (HE) covering $|\eta| < 1.3$ and $|\eta| < 3.0$, respectively, shown in Figure 2.10. These two sub-detectors are sampling calorimeters each containing a brass absorber made up of 70% copper and 30% Zinc, and active plastic scintillating tiles between the absorbers. The choice of the brass was due to its paramagnetic character along with other aspects. The light coming from the scintillation is gathered in wavelength shifting fibres (WLF) which are embedded in the tiles. The WFLs alter the frequency of the coming light to dispatch the light to hybrid photo-diodes (HPDs). The read-out units are composed of scintillating tiles with dimensions of 0.087×0.088 ($\Delta\eta \times \Delta\phi$) in the HB and of 0.17×0.17 in the HE.

An outer layer (HO) is placed on the outer surface of the solenoid using iron as absorber which extends the total interaction depth to 11 times of the average interaction depth of hadrons in the HCAL. The interaction depth here indicates the nuclear interaction length meaning the distance travelled by a hadron before being submit to an inelastic nuclear interaction.

Another part of the hadronic calorimetry system is the very forward calorimeter (HF). Two HFs are placed 11.2 m away from the interaction point in each direction along the beam line and cover the pseudorapidity range $3 < |\eta| < 5$. The HF is very useful in reconstructing the very forward jets and is crucial to measure forward jets produced

in VBF Higgs production. Since the radiation level is very high in this region, these sub-detectors are equipped with radiation-hard steel absorbers and quartz fibres. The quartz fibres produce Cherenkov radiation which is measured by photomultiplier tubes (PMTs). Two different lengths of fibres are installed in order to distinguish the EM showers from hadronic showers; an EM shower leaves a large fraction of its energy in the first 22 cm of the calorimeter whereas a hadronic shower leaves equal amount of trace along the HF. A general view of the HCAL can be seen Figure 2.11.

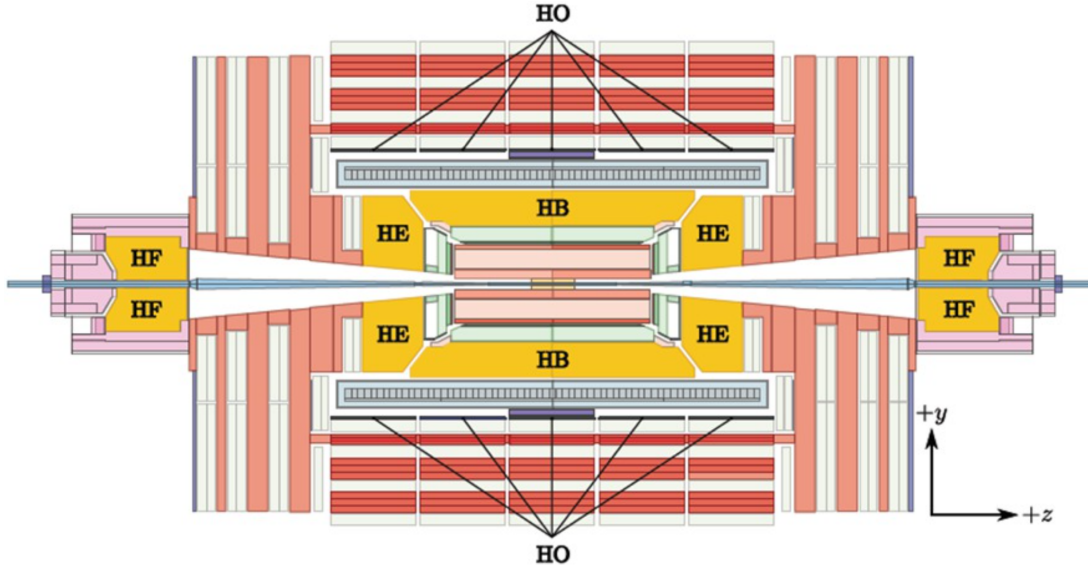


Figure 2.11 : Schematic view of the CMS detector in the y - z plane with the HCAL sub-detectors labelled [75].

2.2.6 Muon system

The typical energy of muons allow them to pass the ECAL, the HCAL and the solenoid without being stopped. The muon sub-detectors constitute the outermost part of the CMS detector. They identify the radiating muons and measure their properties such as p_T and charge. The muon system also provides a powerful trigger for events that involve muons and accurate time measurement of the bunch crossing [76].

The muon system is positioned just outside of the solenoid magnet covering the $|\eta| < 2.4$ region, and is interspersed with the iron return yokes in order to bend the trajectory of muons hence to measure their p_T , with the 3.8 T magnetic field inside the solenoid and the 1.8 T average return field. Additionally, the muon systems are well protected from the charged particles other than muons due to the presence of a large amount of material including the solenoid.

The muon system consists of three types of gaseous detectors; Drift Tubes (DTs), Cathode Strip Chambers (CSCs) and Resistive Plate Chambers (RPCs). The DTs are placed within $|\eta| < 1.2$ region in the barrel, the CSCs in $0.9 < |\eta| < 2.4$ in the endcaps and the RPCs in both barrel and endcaps, shown in Figure 2.12. Each of these gaseous detectors is placed in a particular region in the detector due to their different rate capabilities which is a parameter related with the time needed for the material to receive the next particle, and due to their ability to work inside a magnetic field.

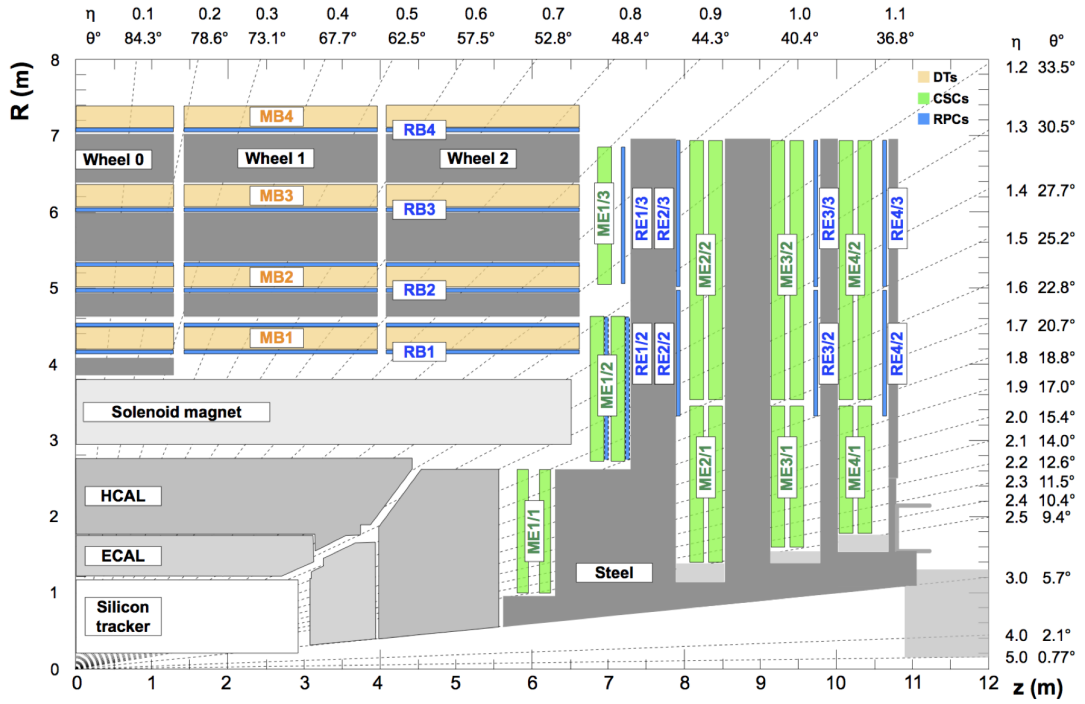


Figure 2.12 : R-z cross section of a quadrant of the CMS detector. The bottom left corner indicates the interaction point. Various muon stations and the steel flux-return disks are shown. The drift tube stations (DTs) are labelled MB (Muon Barrel) and the cathod strip chambers (CSCs) are labelled ME (Muon Endcap). The resistive plate chambers in the barrel and in the endcaps are shown RB and RE, respectively [77].

These limitations require the *Drift Tubes*, which do not have a high rate capability, to be placed only in the barrel region. Additionally, the DTs require the trajectory of the particles traversing them to be modified as small as possible. This makes the barrel is a good choice, where the residual magnetic field is low. Another advantage of placing DTs in the barrel is that the muon and neutron induced background rates are low. The Muon Barrel (MB) detectors consist of four stations in coaxial cylindrical layouts around the beamline interspersed with the iron yoke. These cylindrical stations

make up the five wheels of stations aligned in the beamline direction with five wheels of return yokes placed in between. This region consists 250 drift chambers.

The basic constituent of a DT is the drift cell, shown in Figure 2.13, is a tube of rectangular cross section filled with $Ar/CO_2(85/15)$ gas mixture. An anode wire is placed in each of the cells, while the cathode strips are positioned along the small sides. When a charged particle is passed through the cell, the atoms of the contained gas becomes ionised and the resulting electrons of the gas atoms drift towards the anode wire. The geometry of the cell provides a uniform electric field inside, which allows the drift velocity to be constant and precise. Hence, it is possible to measure the drift time of the electrons therefore the position of the ionising particle [78].

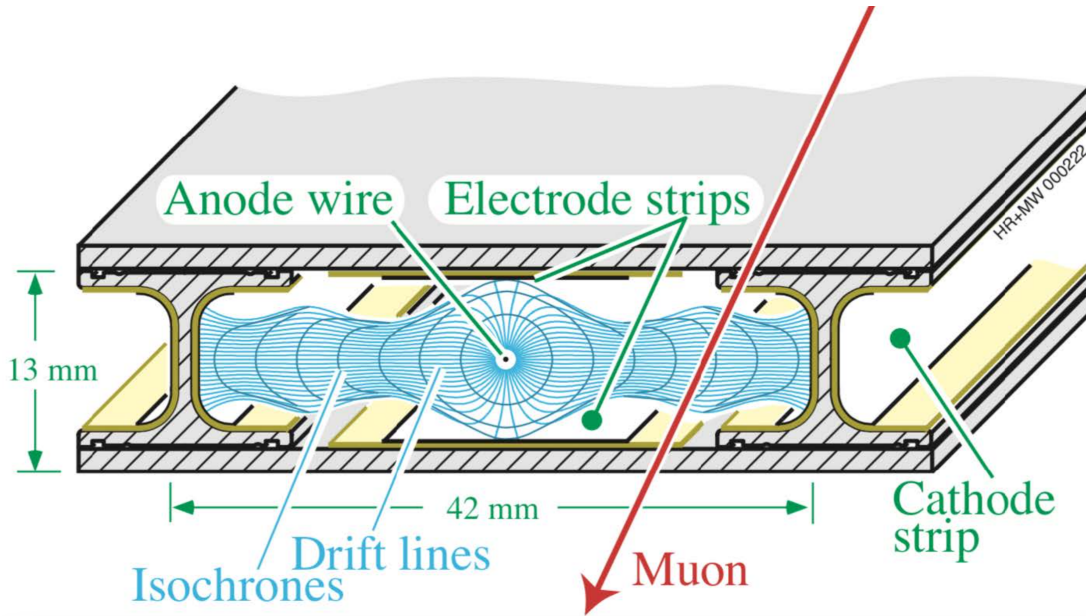


Figure 2.13 : Schematic view of a drift cell showing the electric field lines in the gas volume as well as the isochrones (contours of equal drift times) and the drift lines [79].

A single drift cell, with a $42 \times 13 \text{ mm}^2$ cross section and 2 to 3 metres of wire length, ensures a drift time of about 400 ns and a resolution of about 200 μm resulting in 80 – 120 ns of resolution for the global chamber measurement. Four stacked layers of drift cell constitute a super-layer, and two or three super-layers make up a drift tube. The orientation of the anode wire differs; in the outer super-layers they are parallel to the beamline, while they are orthogonal at the inner part. This arrangement allows to provide information regarding different coordinates. A track measurement is

performed in the (r, ϕ) plane at the outer super-layers where the residual magnetic field is low, while the inner part measures the z coordinate.

Contrary to the DTs, the ***Cathode Strip Chambers*** are placed in the endcaps where a higher residual magnetic field and larger particle rate is present, since they are more robust for the radiative environment and the magnetic field. Both the DTs and the CSCs yield a high spatial resolution for the p_T measurement of the charged particles. The main task of the CSCs is the tracking measurement of the muons, formed as four stations at each endcap. Each CSC is composed of a multi-wire proportional chamber, where the cathode plane is divided into sections of strips that are perpendicular to the wire direction. The chambers are filled with the $Ar/CO_2/CF_4$ (40/50/10) gas mixture. They have a trapezoidal shape that is made up of seven cathode planes stacked with a 10 mm gas gap between and contain planes of anode wires, shown in Figure 2.14a. When a muon passes through one of the CSCs, it produces an avalanche of electron-ion pairs in the gas and induces signals on the cathode strip wires. The information is then used to acquire the position of the ionising particle, with the radial component being measured with the wires and the z component via the cathode planes [78].

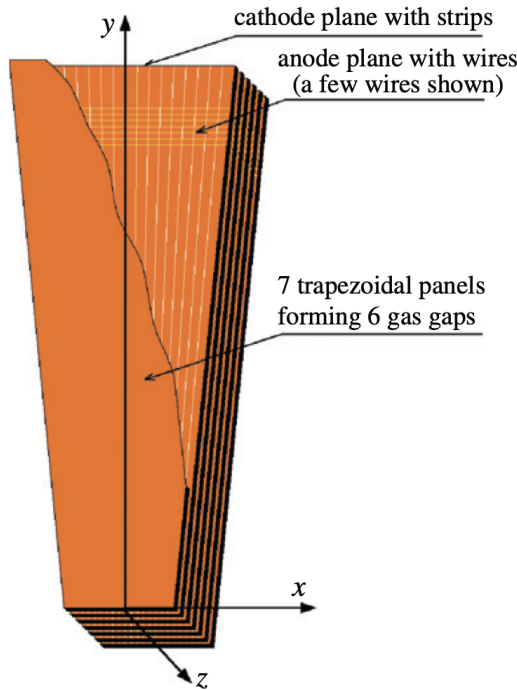


Figure 2.14a : CSC

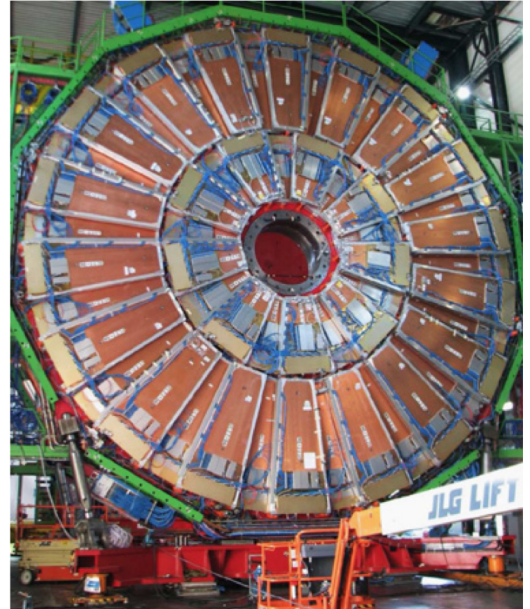


Figure 2.14b : CMS Endcap

Figure 2.14 : (a) Schematic view of a CSC and, (b) a photo of the ME+2 disk with its stations of CSCs [80].

Finally, the **Resistive Plate Chambers** are installed in both barrel and endcaps in order to provide a redundancy in the muon chambers with their fast trigger signal. The RPCs constitute 1056 parallel-plate gas detectors with a moderate spatial resolution of 0.8 - 1.2 cm, but a distinguished time resolution on the orders of ns. However, the electronics system records the RPC hit information at each bunch crossing that is 25 ns, reducing the full timing resolution of the RPC. Each RPC consists of 2 mm of two parallel planes of Bakelite, which is a very resistive resin, coated with graphite. These two plates are separated by 2 mm and filled with the $C_2H_2F_4/i - C_4H_{10}/SF_6$ (96.2,3.5,0.3) gas mixture. The RPC detectors are used as couples in order to improve the efficiency of particle detection. The charged particles that pass through the RPC dual generate an avalanche of ions, and the signal coming from these are collected on a set of readout aluminium strips installed between the dual, shown in Figure 2.15. RPCs can function in two different modes: the *streamer mode* where a strong electric field that produces localised gas discharges near the regions of the trajectory of the ionising particle, and an *avalanche mode*, where the electric field is weaker to allow a higher counting capacity per area because of a reduced charge generated in the ionisation. The RPCs at the CMS muon system operates in avalanche mode allowing to provide higher rates [78].

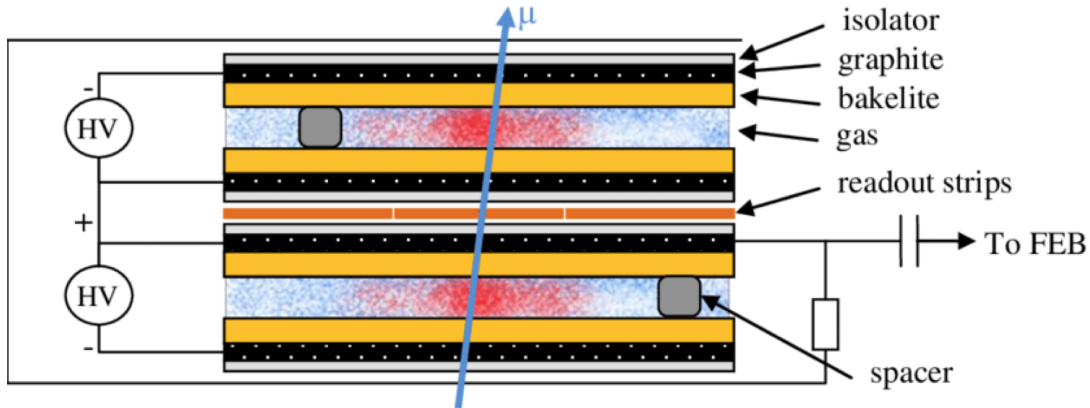


Figure 2.15 : Schematic view of a dual RPC detector [81].

2.2.7 Trigger system and storage

The proton-proton collisions take place at the centre of the CMS detector every 25 ns, corresponding to a data of about 70 Terabytes per second. This much of data, being impossible to store with current electronics technology, mostly contains low-energy

pp interactions that are not of much interest for the physics programme of the CMS experiment; the total pp interaction cross section, that is about 10^{11} pb ($\sim 98.3 \text{ mb}$), is 5 order of magnitude larger than the frequently studied processes at the LHC [82]. The task of the trigger system is to select the events that are interesting for physics analyses, reducing the acquisition rate by a factor of about 10^5 . The trigger must satisfy the technical requirements for the online data-taking and provide a high efficiency for offline data analysis [83]. Specifically, with the nominal bunch crossing of the LHC, every bunch produces about 20 pp collisions resulting in about 800 M collisions per second. The corresponding data is retained and processed in pipelines, with $\sim 1 \text{ MB}$ of memory per event making up to 70 TB per second. The trigger system decides which events to be stored in a very short time $\sim 25 \text{ ns}$ to satisfy the LHC requirements.

The trigger systems includes two independent trigger levels. The ***Level-1 Trigger*** (L1) consists of only hardware components. It reduces the the event rate from 40 MHz (corresponding to 25 ns) to a 100 kHz. The decision at L1 is made using the information from the calorimeter and the muon system. The tracker information is not used since its reconstruction time is longer than the L1 decision limit [84]. The L1 trigger uses the characteristics of trigger objects (or trigger primitives), that are photons, leptons, jets, and p_T . The trigger objects are provided based on the energy deposits in the calorimeters and hit patterns in the muon chambers. The L1 trigger information from these two sub-detectors are then combined in the L1 Global Trigger (L1 GT) which takes the decision of whether to pass or reject the event.

The events that pass the L1 GT is sent to the ***High Level Trigger***, a software-based system containing multi-core computers. The task of the HLT is to reduce the event rate to be able to store the data in disks. It reduces the event rate from 100 kHz to about 1 kHz. The events that pass the requirements of at least one of the HLT trigger paths, is written permanently on disk by the ***Data Acquisition System*** (DAQ) and sent to CERN Tier 0 (T0) storage system. The ***Data Quality Monitoring*** (DQM) observes the recorded data and labels it in terms of quality. Finally, the collision data becomes certified and is stored as data-sets for further physics analyses.

The data that pass the trigger systems makes up to 70 PB of data every year of LHC operation, and in order to satisfy the demanding storage need, a globally distributed storage system called ***World LHC Computing Grid*** (WLCG) is established in 1990s

[85]. The WLCG consists of the computing resources of over 170 sites in 42 countries with a total of about 900 000 processor cores and 1 Exabyte of storage. This tremendous distributed computing infrastructure grants more than 12 000 physicist around the world the access to the resources nearly the real-time. Specifically, at the end of the Run II, the transfer rates exceeded 60 GB/s globally with over 2 million tasks per day.

The WLCG consists four layers called Tiers, namely T0-1-2-3 and each tier includes several computing sites with particular tasks. Tier 0 is the CERN Data Centre and the complete set of raw data passes this facility. The first reconstruction is also made at T0 producing Primary Datasets with specific data formats; Reconstructed (RECO) and Analysis Object Data (AOD). At Run II, the RECO datasets had 4 MB of disk usage per event which was too much for the disk capabilities. Furthermore, the RECO datasets contain mostly unnecessary information for the majority of the analyses; only a few of them need all the silicon tracker hits or ECAL clusters etc. Hence the RECO datasets are not stored unless for debugging purposes. Instead, the analysis objects are chosen among RECO datasets and stored in AOD format. The AOD datasets usually contains 500 kB per event, which could still fill large storage. Another skimmed dataset is thus created to be widely used by the physics analyses at Run II, called MiniAOD. This data format reduces the precision of measurements to an acceptable level and does not include the track hits and the p_T tracks that are too small. MiniAOD datasets occupy about 50 kB per event and are used by the 95 % of physics analyses. The remaining 5 % uses either AOD samples or further skimmed datasets called NanoAOD. The NanoAOD is the last chain in the process containing a ROOT flat ntuple where mostly the tracks are not included.

The Tier 0 distributes the raw and the RECO datasets to Tier 1 sites via the LHC Optical Private Network that consists of optical-fibre links operating at 10 Gbps, and reprocesses data outside the LHC runs. Tier 1, with 13 computing sites, stores the LHC data. It allows continuous access to the Grid as well reprocessing and storing the output. Tier 1 distributes data to Tier 2 sites, also stores a part of the simulated data coming from Tier 2 sites. T2s, with more than 150 sites, are typically the universities and institutes where a proportion of the LHC data is stored and some computing power is provided for the production and reconstruction of simulated event samples. Finally,

Tier 3 is formed by the individual researchers accessing to Grid from local computing resources with no formal engagement with WLCG.

2.3 The HL-LHC Programme and the Phase II Upgrade of the CMS Detector

The inaugural schedule of the LHC programme, detailed in Figure 2.1, is aimed to reach its final design expectations, so called the *High-Luminosity LHC*, at the end of the 2020s. The integrated luminosity delivered is expected to be 3000 - 4000 fb⁻¹ for pp collisions and, PbPb and pPb collisions with integrated luminosities of 13 nb⁻¹ and 50 nb⁻¹, respectively, in 2030s and onwards. If the HL-LHC upgrades were not to be realised, the total integrated luminosity for another 10 years of operation would be around 1000 fb⁻¹ [86]. In any case, the current LHC inner triplets will have to be replaced after an integrated luminosity of 400 fb⁻¹.

The average pile-up will reach 200 interactions per bunch crossing while providing a centre-of-mass energy of 14 TeV, which will surely impose a demanding environment for the detectors around the LHC ring. Especially, after the discovery of the SM Higgs boson, the case study for an HL-LHC is easier to define and quantify [87]. The SM expects the discovery of all final states of the Higgs boson given a high integrated luminosity, for example the decays $H \rightarrow \mu^+ \mu^-$ and $H \rightarrow Z\gamma$ have the need of an integrated luminosity on the order of 3000 fb⁻¹, which supports the HL-LHC programme.

The excess amount of data will also provide higher-precision measurements of the Higgs couplings, for instance in the $HH \rightarrow \tau\tau bb$ and $HH \rightarrow \gamma\gamma bb$ [88]. In addition to the SM Higgs sensitivity, the BSM models such as supersymmetric partners of quarks and gluons with mass greater than 1.5 TeV/c² will benefit from the high amount of data. Another example is the BSM Z' bosons which would show the existence of new weak interactions with the HL-LHC programme if its mass is above 2.5 TeV/c².

In this context, the demanding conditions of the HL-LHC such as the increased luminosity and the higher pile-up, entails a major upgrade to the current *CMS detector* in order to fully exploit the physics potential. The pixel and strip tracker detectors will be replaced in order to increase the granularity, enhance the radiation hardness, decrease the material budget and provide extended geometrical acceptance up to

$\eta = |4|$. The front-end electronics of the barrel ECAL will be upgraded to be able to access single crystal information at L1 trigger, to provide 160 MHz sampling which allows high precision timing capability for photons, and to accommodate bandwidth requirements and trigger latency. The muon system will undergo upgrades for its 3 types of gaseous detectors; new muon detectors with enhanced RPC and gas electron multiplier technologies will contribute to the redundancy of the detector, increase the pseudorapidity coverage up to $\eta = |2.8|$ and enhance the trigger and reconstruction performance in the forward region. The endcap ECAL and HCAL will be replaced with an advanced combined sampling calorimeter which will provide higher-precision timing information and higher-segmented spatial information both in the longitudinal direction and the transverse plane. A new timing detector will be added for minimum ionising particles in both barrel and endcap regions and aimed to better reconstruct the interaction vertices. This upgrade will also help mitigate the performance degradation due to high pile-up. Finally, the complete trigger system; L1, HLT and DAQ will be upgraded substantially [89].

3. DATA ANALYSIS

The first section of this chapter includes an overview of the simulation of the proton-proton collisions and the second section explains the details of the prospects for non-resonant Higgs boson pair production measurements at the HL-LHC in pp collisions with a centre-of-mass energy of $\sqrt{s} = 14 \text{ TeV}$.

3.1 Simulation and Reconstruction of Proton-Proton Collisions

Monte Carlo event generators are crucial for high energy collision simulations. These software are used by many scientists for many purposes including the anticipation of the design outcomes of colliders. Simulation of a pp collision starts by describing the hard interactions and moves on to hadronisation and decay. In this section, the working principles of MC simulations, detector simulations along with reconstruction techniques are explained.

3.1.1 Event generation

The interactions at the level of partons as well as other subatomic particles, are not deterministic contrary to the classical physics hence the transitions from one quantum state to another is computed as a probability. The transition rate $\Gamma_{i \rightarrow f}$ from an initial quantum state i to a final state f is given by the Fermi's Golden Rule as

$$\Gamma_{i \rightarrow f} = \frac{2\pi}{\hbar} |\mathcal{M}|^2 \rho, \quad (3.1)$$

where \mathcal{M} is the **Matrix Element** and is defined as $\langle f | V_i | i \rangle$ for an interaction potential V_i . The matrix element describes all the dynamics of the transition and is calculated by evaluating the corresponding Feynman diagrams using the Feynman rules convenient to the interaction of interest. The second ingredient in the recipe is the available *phase space* ρ , also called the *density of final states*, is exclusively kinematic and depends on the masses, energies and momenta of the incident particles. The term represent the phenomenon that the interaction in question is more likely to happen given more space in the final state.

The Golden Rule can be used to calculate cross sections in addition to decay rates. The lowest order term in the perturbation theory calculation for cross section has the highest contribution which is often referred as **Leading Order** (LO) term and contains only the LO Feynman diagrams. Today, the cross sections are computed using **next-to-leading order** (NLO) calculations in MC event generators which include one-loop diagrams, initial and final state radiations. However, the NLO calculations consume more processor power and time.

The event generation is the first step of the simulation of the high energy collisions. The Monte Carlo techniques are used to simulate the events which happen in the actual collider experiments. The collisions between the protons at the LHC, takes place in fact between the constituents of the protons, called partons to denote the quark and gluon content. In the pp collisions, not only two partons from different protons collide, which produces an interesting event to study, but also two partons can interact but not hardly. The different types of interactions that may occur in the collisions are shown in Figure 3.1. The **hard scattering** processes occur by the high exchange of momentum between the constituents of two colliding protons. The products of this type of interaction usually have high transverse momenta. The high energetic partons may also emit radiation in the initial and final states, creating parton showers (PS) which is a considerable effect and taken into account in the event generation process.

An **underlying event** is a soft interaction that occurs between the residue of the protons which have not taken place in the hard interaction. The partons may also radiate gauge bosons before or after interacting with each other, called initial state radiation and final state radiation, respectively. Additionally, partons can emit radiation via the strong interactions which create jets close to the direction of the initial particles.

At the LHC, bunches including 10^{11} protons are collided and pile-up processes, any type of interactions outside the hard scatterings, constitute a significant effect in the event generation. Another aspect of the pp collisions that needs to be considered in MC simulations is the **hadronisation** process. It is the merger process of quarks and gluons to form colourless, and seen in the final states. The hadronisation process occurs if the partons reach the energy scale of about 1 GeV. The creation of colourless primary hadrons from partons is described in two different models. The **cluster model** describes the forming of the colourless hadrons with the gluons that are split into quark-antiquark

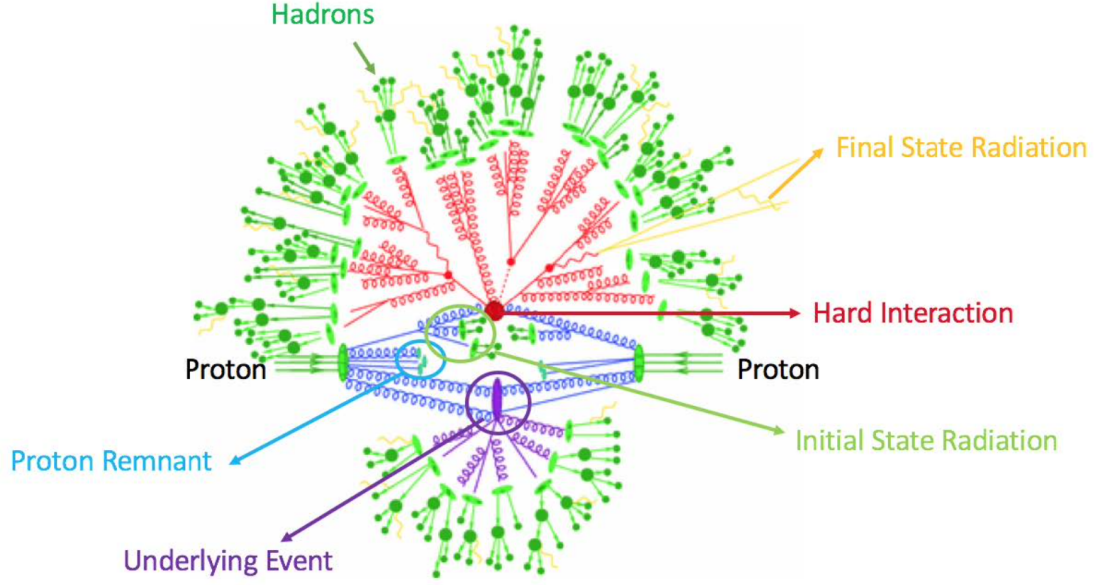


Figure 3.1 : Graphical representation of a pp collision event. The proton beams come from the either sides. The red diagram shows the hard interaction and the consequent decay of the products. A secondary interaction before the final state partons hadronise, is shown in purple. The hadronisation is indicated in green [90].

pairs which then combines as colourless groups. The states clustered in this manner usually possess a large invariant mass which consequently decay to lower mass states convenient to create hadrons [91]. In the *string model*, the gluons are thought to be split into quark pairs that move away from each other where a string-like configuration is formed between them. As the string is stretched, its potential energy increases lowering the kinetic energy. The string eventually breaks in two creating a quark-antiquark pair. The mechanism is repeated until there is no energy left for another quark pair to be created [92].

Various theoretical, phenomenological and experimental inputs are injected into MC simulations in order to produce a consistent description of the pp collisions. Different approaches are utilised, for example in order to describe the QCD induced processes which have varying phenomenology with the changing energy scale [93]. The hadronic cross section σ_{pp} is calculated using the QCD factorisation theorem, which describes it as a convolution integral of the partonic cross section $\hat{\sigma}_{ij}$ with the parton distribution functions $f_i(x)$:

$$\sigma_{pp} = \int_{x_{min}}^1 f_i(x_1) f_j(x_2) \hat{\sigma}_{ij}(x_1 p_1, x_2 p_2) dx_1 dx_2, \quad (3.2)$$

where $f_i(x)$ is the probability density that a parton of type i has a fraction x of the hadron's energy.

The overall chain of calculating a process includes;

- the PDF, which is basically the phenomenological interpretations calculated with the information coming from the experiments,
- the hard scattering, calculated in perturbative orders,
- the parton shower process, the radiations in perturbative QCD,
- the hadronisation, lays out the forming of colourless hadrons from coloured partons based on phenomenological models,
- the decay of unstable particles treated using experimental data.

The first two steps are usually calculated in Matrix Elements generators, while the rest is usually calculated in parton shower simulations. MC techniques are used in both sides and the transition is made in a manner to avoid double counting of the QCD radiation.

The types on MC generators include general-purpose MC generators which computes the hard processes and the parton shower in separate structures. PYTHIA8 and SHERPA are most common generators of this group. Other types include generators that compute ME+PS or NLO+PS e.g. MADGRAPH5_AMC@NLO which is a unified framework with LO and NLO calculations, and POWHEG BOX which is an ME event generator. The MC event generators used in this thesis is explained later in subsection 3.2.1.

3.1.2 Detector simulation

The detector simulation is another crucial step for the simulation of the high energy collisions. It computes the interactions between the particles and the detector material just as in the actual detector. The final state particles simulated in MC event generators are fed into detector simulations to perform the event reconstruction. There are several

detector simulations for high energy collisions. Geometry and Tracking (GEANT4) simulation software [94], includes a complete and accurate description of detectors. However, the full simulation of the detector is a long task with high CPU consumption, and in order to cope with the limited computing resources and still allow the use of large samples, the LHC experiments have developed fast simulation software with novel techniques [95, 96] with two to three orders of magnitude faster than GEANT software. In this thesis, all signal and background samples are simulated with the Phase II upgraded CMS detector geometry using DELPHES 3 fast simulation [97].

DELPHES 3 aims at simulating multipurpose detector response which includes a track propagation system in the presence of magnetic field, calorimeters and a muon system. It uses the input from the most common event generators and mimics the detector response. In order to achieve this, long-lived particles coming from the hard scattering are propagated through the electromagnetic and hadronic calorimeters withing a uniform magnetic field parallel to the beam line. If the particle is neutral, its trajectory is defined as a straight line from the interaction point to a calorimeter cell. The user may specify the calorimeter segmentation and the magnitude of the magnetic field. After the propagation step, long-lived particles reach the calorimeters, ECAL and HCAL depositing a fixed fraction of their energy. ECAL and HCAL are overlaid and assumed to have same granularity, hence a particle reaches one ECAL and one HCAL cell. These two cells are grouped in a calorimeter tower which is computed as the geometrical centre of the cells. The photons and electrons are defined to deposit all their energy in ECAL with the fraction $f_{ECAL} = 1$, and hadrons in HCAL with $f_{HCAL} = 1$ although in reality hadrons deposit some energy also in ECAL. Muons, neutrinos and neutralinos leave the calorimeters intact and some other particles such as Kaons are treated in a way that they leave their energy in calorimeters instead of decaying. The ECAL and HCAL energy deposits are independently smeared by a log-normal distribution and the energy of each particle is concentrated in one single tower, with the following formula,

$$E_{tower} = \sum_{particles} \ln\mathcal{N}(f_{ECAL} \times E, \sigma_{ECAL}) + \ln\mathcal{N}(f_{HCAL} \times E, \sigma_{HCAL}) , \quad (3.3)$$

where $\ln\mathcal{N}(m, s)$ is the log-normal distribution with mean m and variance s , and the parameters σ_{ECAL} and σ_{HCAL} are the ECAL and HCAL resolutions given in Equation 2.5, respectively. The energy of each particle is summed over for a given

single tower. High level objects such as jets and missing transverse energy is computed either from the calorimeter deposits or by the particle flow (PF).

3.1.3 Particle flow algorithm

The PF reconstruction, whose philosophy is to use a maximum amount of information obtained from the sub-detectors, is adopted for some experiments including CMS [98] and ALEPH [99], and is implemented in DELPHES with a basic approach based on the tracker and the calorimeter information. In the experiments, the tracker provides a higher resolution than the calorimeters up to some energy scale, therefore DELPHES uses the tracker information to calculate the charged particle momenta. The PF algorithm creates two set of 4-vectors; *particle flow tracks* and particle flow towers. For each calorimeter tower, the algorithm sums the energy deposited in ECAL and HCAL, denoted E_{ECAL} and E_{HCAL} respectively. The total energy deposited by the charged particles for which the tracks have been reconstructed, $E_{ECAL,trk}$ and $E_{HCAL,trk}$ are used in the algorithm to define,

$$\Delta_{ECAL} = E_{ECAL} - E_{ECAL,trk} , \Delta_{HCAL} = E_{HCAL} - E_{HCAL,trk} , \quad (3.4)$$

$$E_{tower} = \max(0, \Delta_{ECAL}) + \max(0, \Delta_{HCAL}) \quad (3.5)$$

then the algorithm creates a particle flow track for each reconstructed track, and if $E_{tower} > 0$, it creates a particle flow tower with E_{tower} . In order to illustrate the algorithm with few examples, we can think of a single photon with energy deposited in ECAL with E_{ECAL} . A particle flow tower is created for the photon with the deposited energy and no track is created. Another example can be a charged pion with a track energy $E_{HCAL,trk}$ and with an energy deposit of E_{HCAL} . If Equation 3.5 yields zero or a negative value that is $E_{HCAL} \leq E_{HCAL,trk}$, then only a particle flow track with energy $E_{HCAL,trk}$ is created. If $E_{HCAL} \geq E_{HCAL,trk}$, a particle flow track with energy $E_{HCAL,trk}$ along with a particle tower of energy E_{HCAL} is created.

In brief, the particle flow tracks define the charged particles with a good resolution, and particle flow towers define both charged and neutral particles with no relevant tracks characterised by a lower resolution. This approach can be useful for pile-up mitigation, and providing high-resolution input for jet and MET reconstructions. Despite being

very plain compared to actual experiments, the PF algorithm is shown to have a good performance at pp collisions [100].

3.1.4 Reconstruction of physics objects

DELPHES implements the reconstruction and identification of objects based on a set of approximations to reduce the computation time but still keeping a good efficiency.

The *photon* reconstruction depends only on the ECAL, and the final energy of the photons is calculated by using the ECAL resolution function presented in subsection 2.2.4. The photons and electrons that reach ECAL and do not have tracks, are reconstructed as photons in DELPHES.

The reconstruction of *electrons* is usually performed with the combined information from the tracking system and the ECAL. However, DELPHES parametrises the combined reconstruction efficiency as a function of the energy and pseudorapidity instead of dealing with the usual complex reconstruction. The electron reconstruction efficiency drops to zero outside the tracker's geometrical acceptance and below some energy threshold. The energy resolution for the electrons are computed with the combination of tracker resolution at low energies and with ECAL resolution at high energies.

The probability for a *muon* to be reconstructed in DELPHES depends on the user-specified efficiency parameter. The probability drops to zero outside the tracker acceptance and below some momenta threshold. The final momentum is computed by a Gaussian smearing of the initial 4 momentum vector of the muon. The resolution is specified by the user as a function of transverse momentum and pseudorapidity.

Tau leptons decay before being detected, and hereafter, the lepton term is used for electrons and muons. The reconstruction of the hadronically decaying taus will be explained along with jets later.

A photon, electron or muons is said to be isolated if there are small amount of particles around it. The definition of such feature is important since an isolated object is not likely to originate from a jet. Various definitions are available for the isolation of an object. DELPHES implements a simple definition that fits well to hadron collisions. An *isolation* variable is defined for electrons, muons and photons, as the ratio of the

sum of p_T above a threshold in a cone of radius R around a given particle, to the p_T of the particle of interest. The values close to zero signifies that the particle is *isolated*. The user is set free to define parameters of the isolation; p_T , R or a minimum isolation, where the default values are 0.1 GeV , 0.5 and 0.1, respectively.

Jets are produced from three different input collections in DELPHES. The *generated jets* are the clusters of generator level long-lived particles after parton-shower and hadronisation. The *calorimeter jets* are produced using calorimeter towers explained in subsection 3.1.2. Finally the *particle flow jets* are the clusters produced from the particle flow tracks and towers defined in subsection 3.1.3. Additionally, the user is set free to specify a minimum jet p_T for the final jet collection, and a jet clustering algorithm with FASTJET package [101] implemented in DELPHES, supporting the most common jet clustering algorithms. A default reconstruction step is defined in DELPHES that removes the jet from the event if it has been already reconstructed as a lepton or a photon assuring no double counting of objects.

The jets originating from τ decays or from the hadronisation of heavy quarks e.g. b or c quarks, should be identified in pp collisions. DELPHES tags a jet as b or τ jet, if the b or τ is found within some close distance ΔR of the jet axis, respectively. The user is set free to specify the tagging efficiency which affects the probability of being identified as b or τ . Additionally, a mis-tagging efficiency, that is the probability that a particle other than b or τ is identified falsely as b or τ .

DELPHES is also capable of computing the missing transverse energy E_T^{miss} and scalar transverse energy sum H_T .

3.1.5 High level corrections and validation

The reconstructed objects need to be corrected for residual effects before being used in the final analysis. **Jet energy scale correction** (JEC) is applied in order to meet the deficit between the average momenta of the reconstructed objects and that of their generator-level equivalent. This deficiency is common for complex objects such as jets, because the total smearing is ambiguous as a result of the loss of generator-level objects such as neutrinos and muons. The JEC is therefore applied only on jets, where the user can specify it as a function of η and p_T of the reconstructed jet.

The pile-up effects are simulated using a pre-generated QCD, and randomly placed on the beamline with a number obtained from a Poisson distribution. Pile-up has direct effect on the performance of jets, E_T^{miss} and isolation however pile-up mitigation on E_T^{miss} is a demanding task and is not implemented in DELPHES. Instead, ***pile-up subtraction*** is applied on jets and the isolation with two methods. *Charged pile-up subtraction* assumes that the track reconstruction efficiency does not depend on the vertex position, and if the PF algorithm is used, the PF tracks that originate from pile-up are cleared and do not enter the jet clustering and isolation algorithms. Some other residual effects, e.g. the particles that are in close proximity of the hard interaction vertex, the charged particles that are outside the geometrical acceptance of the tracker, or the neutral particles, need to be removed. In order to do this, an average contamination is defined in the *residual pile-up subtraction* which is then removed from the jet energies and from the isolation variable.

All in all, DELPHES has been validated in many use cases, and is a good simulation to be used in pp collisions. Figure 3.2 shows a comparison of DELPHES and CMS detector performance using photons and electrons.

3.2 Higgs Boson Pair Analysis

This section describes the analysis steps of the prospects for non-resonant Higgs boson pair production measurements at the HL-LHC in pp collisions with a centre-of-mass energy of $\sqrt{s} = 14 \text{ TeV}$ with the Phase II upgraded CMS detector. In this analysis, a Higgs boson pair search in the $WW\gamma\gamma$ channel is performed benefiting from the $H \rightarrow \gamma\gamma$ process which provides a clean and distinguishable signature. The final states of Higgs boson pair in $WW\gamma\gamma$ channel are included in the study due to overlap, which in turn increases the overall analysis sensitivity. A schematic flow chart describing the analysis is shown in Figure 3.3.

3.2.1 Signal and background samples

In this thesis, signal ($gg \rightarrow HH$) samples are generated using POWHEG v2 [103–106] at (NLO) calculations in QCD including the full top mass dependence with SM parameters, and subsequent decays of the Higgs boson pairs into WW and $\tau\tau$ each with a photon pair is implemented using PYTHIA 8.212 [107]. The signal samples

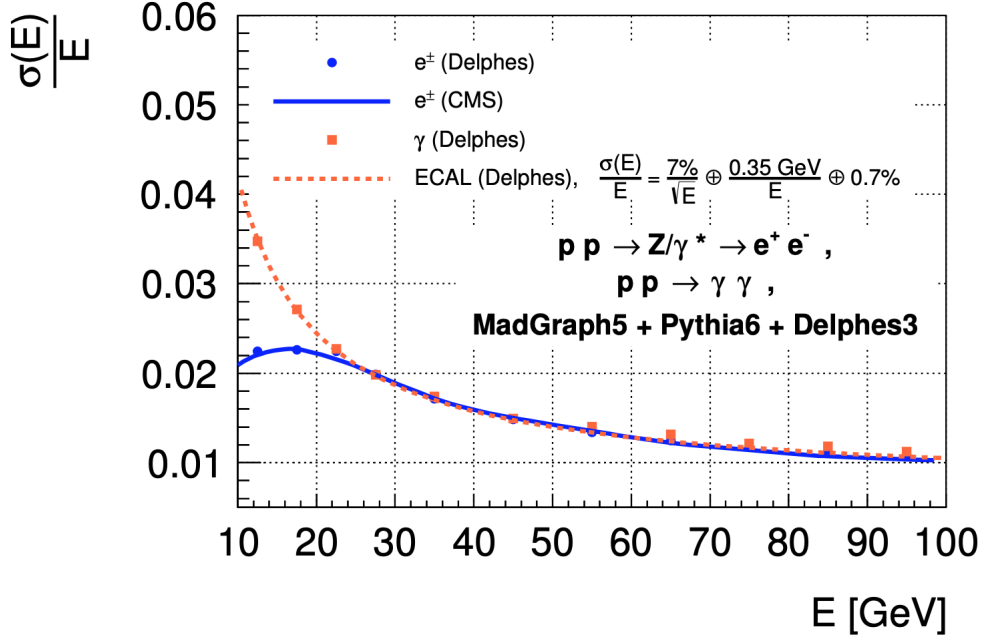


Figure 3.2 : Energy resolution of photons and electrons as a function of energy for a CMS detector configuration. The energy resolution formula of the CMS ECAL, and the samples used for the study along with their generator are given. The CMS electron distribution is taken from [102]. The electron and photon resolutions are identical at high energies since both objects are dominated by ECAL, however at low energies, the electron resolution differs widely because of the finer tracker resolution [100].

for three separate final states of $WW\gamma\gamma$ channel are produced while a single sample including only hadronic taus are used for $\tau\tau\gamma\gamma$ channel.

The analysis is overwhelmed by non-resonant backgrounds with continuum di-photon invariant mass $m_{\gamma\gamma}$ spectra and by single Higgs boson productions. The sample generation for single Higgs productions is performed via MADGRAPH5_AMC@NLO [108, 109] with the FxFx merging scheme [110] for the gluon fusion (ggFH), vector boson fusion (VBFH), associated production with a vector boson (VH) and associated production with a top quark pair (ttH), while the top quark associated production with a Higgs boson (tHq) was performed using MADGRAPH version-2.7 at LO.

Several SM processes contribute to the continuum background. A large portion of the dominant backgrounds across all final states comes from the $\gamma\gamma$ +jets processes that are modelled with SHERPA v.2.2.1 generator [111] while QCD-induced processes, γ +jets and WW processes are modelled with PYTHIA 8 generator [107]. W production in association with photons and jets, and Drell-Yan processes are modelled with MADGRAPH5 version-2.7 at LO, $t\bar{t}$ with POWHEG v2, and $t\bar{t}W$, $t\bar{t}\gamma$, $t\bar{t}\gamma\gamma$, $Z\gamma$ processes

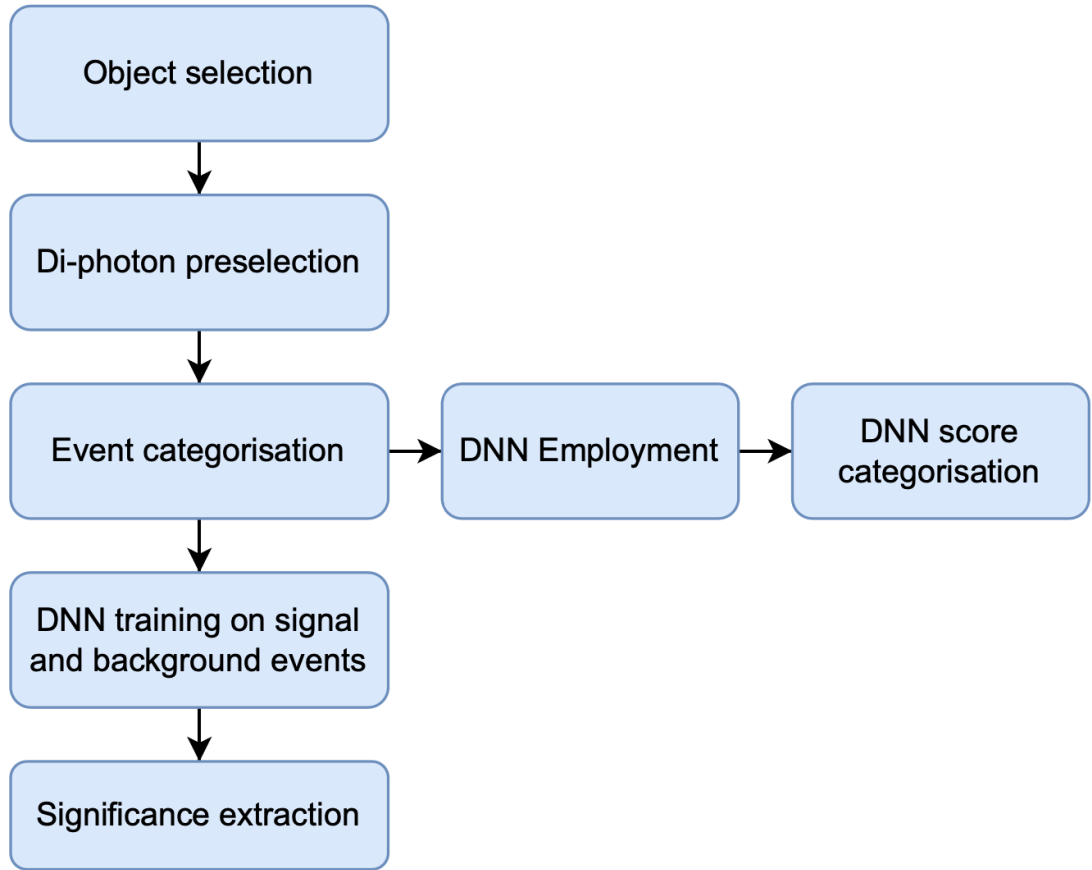


Figure 3.3 : Analysis flow chart.

with MADGRAPH5_AMC@NLO. A complete list of samples used in the analysis is shown in Table A.1.

In this thesis, all signal and background samples are simulated with the Phase II upgraded CMS detector geometry using DELPHES 3 fast simulation [97] with an average pile-up of 200 interactions at $\sqrt{s} = 14$ TeV. An ntupliser is used to produce flat ntuples using the outputs from DELPHES, whose form is shown in Figure 3.4.

For the Phase II upgraded CMS detector configuration in DELPHES input card [114], the isolation (ISO) and identification (ID) variables for photons and electrons are defined in different bit-wise working points (WP); loose, medium and tight (LMT) based on the MVA trained ID and ISO variables in full simulation of CMS (FullSim). Similar ID and ISO variables are defined for muons and taus, and a *btag* value in bit-wise LMT WPs is also defined for jets.

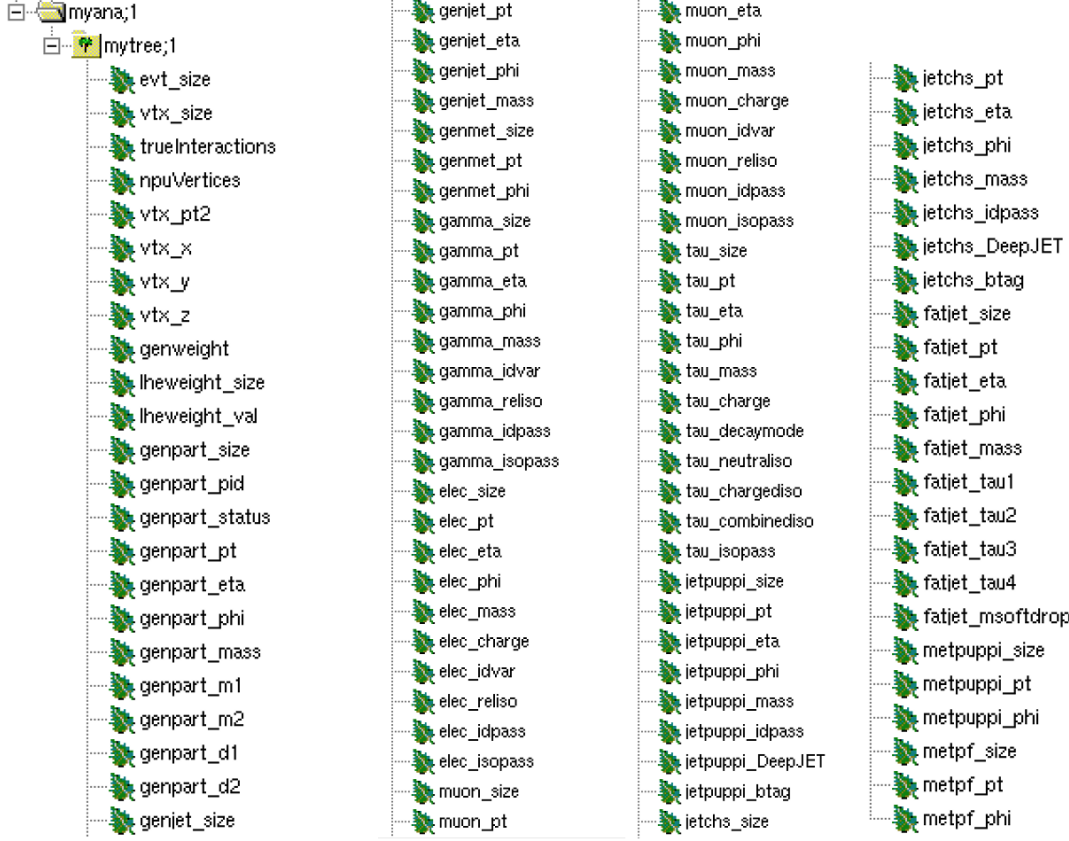


Figure 3.4 : The flat tree format of the DELPHES simulated samples. The reconstructed objects include photons, electrons, muons, taus, jets and E_T^{miss} . Each object has the multiplicity, mass, p_T , η , ϕ values except E_T^{miss} which do not have mass and η values. Identification and isolation variables are available for most of the objects and b-originated jets information is available for jets. There is also the possibility to work on the generator level with generated particles (genpart), jets (genjet) and E_T^{miss} (genmet) collections. Genparts collection possess mother (m1, m2) and daughter (d1, d2) particles information along with particle status codes (PID). Jets and E_T^{miss} have different versions produced with different algorithm, i.e. pile-up per particle identification (PUPPI) [112], charged hadrons subtracted (CHS) [113] or particle flow (PF). Additionally, number of events, vertex information, generator level weight are among the accessible information.

3.2.2 Object selection

Photons used in this analysis are required to have a p_T above 25 GeV and be in the geometrical acceptance $|\eta| < 2.5$. The photon with the highest p_T in the events, namely the leading photon is required to have a p_T of at least 35 GeV. The relative isolation ISO, defined in subsection 3.1.4, is required to be less than 0.3 for photon

candidates. A bit-wise criteria; 1, 3, 7 for loose, medium, tight, respectively, is applied in order to select the loose-ID photons. The photons that pass all these selections are called *good photons*.

Electrons are required to have transverse momenta above 10 GeV within $|\eta| < 2.5$ region. The electrons and good photons are isolated from each other with an angular separation of at least 0.4. The muons that have p_T above 10 GeV and in $|\eta| < 2.5$ have been selected and they are isolated from the good electrons and good muons with an angular separation greater than 0.2. Tau leptons with $p_T > 20$ GeV are selected in $|\eta| < 2.5$ region and are cleaned with respect to good photons/electrons/muons. PUPPI-jets that have $p_T > 30$ GeV with $|\eta| < 5$ are used with an angular separation of 0.4 with respect to good photons/electrons/muons/taus. A list of all selections is shown in Table 3.1.

Table 3.1 : Object selections.

Object	p_T [GeV]	$ \eta $	(ΔR) w.r.t.	Working Point
Photon (γ)	25	3	-	loose ISO+ID
Electron (e)	10	3	γ (0.4)	loose ISO+ID
Muon (μ)	10	2.8	e, γ (0.2)	tight ISO+ID
Tau (τ)	20	3	e, γ, μ (0.2)	tight ISO
Jet	30	5	e, γ, μ, τ (0.4)	tight ID, medium btag

3.2.3 Event selection and categorisation

All events in the analysis are required to have an invariant mass of two good photons in the $100 < m_{\gamma\gamma} < 180$ GeV signal region as a preselection for all final states. The analysis is performed in four mutually orthogonal final states; semi-leptonic and fully-leptonic final states targeting the decays of the W bosons, and single and double τ lepton final states each with a photon pair. Fully hadronic final state of $WW\gamma\gamma$ is not included in this thesis due to modelling difficulties in QCD induced background samples.

Semi-leptonic Final State

In the semi-leptonic final state, the events that contain at least one good photon pair (the preselection) and exactly one good electron or good muon are selected. This final state is expected to be the most sensitive of the two $WW\gamma\gamma$ final states due to the mix of

a high energy lepton from the $W \rightarrow l\nu$ decay with the relatively large branching fraction of the $W \rightarrow qq$ leg of the decay.

In order to separate the Higgs pair signal from the resonant background (single Higgs boson production) and the continuum background, two binary *Deep Neural Networks* (DNN) are employed for signal. Due to the imbalance between the two classes caused by the dataset, event weights are first scaled to the expected cross sections, then re-weighted so that the learning weight in the classes are normalised. This procedure assured the neural networks to focus on categorising the two classes with equal importance. In order to avoid over-fitting in the DNN, the data sets are divided into two almost equal parts based on, by choice, the fifth decimal place of the leading good photon's ϕ value. The output sets are labelled as *even* and *odd* according to the chosen decimal, and the DNN trained with the even set is evaluated on the odd set, and vice versa.

The framework choice for the DNN is the TENSORFLOW [115], which is an open source software library for machine learning that can be used with a wide variety of programming languages; Python, Javascript, C++, or Java. KERAS [116] is the interface chosen to be used with TENSORFLOW. The reason to choose these software is that they are compatible with PYTHON programming language. Additionally, both systems can run on both CPU and GPU which makes them very fast, and there is a large community working on them which makes it easy to find solutions to the problems encountered in the DNN setup. Another reason to use these software is the compatibility between them and the analysis framework used, called BAMBOO [117]. It is a high-level HEP analysis library built on top of ROOT's *RDataFrame* class. *RDataFrame* offers a modern, high-level interface for analysis of data stored in TTree and other data formats, in Python or C++. TTree is the data format of the samples used in this thesis, and since BAMBOO is built using PYTHON language, the whole analysis and DNN frameworks work in good harmony. The analysis script that is used to define objects and for the event selection and categorisation is given in Appendix Listing A.1. The DNN setup script is also provided in Appendix Listing A.2.

The DNNs are trained using skimmed signal and background samples and are evaluated on the full samples in order to perform DNN categorisation. The feature variables used as input to train the DNNs include the kinematic variables such as p_T ,

η , ϕ and energy values of objects along with the constructed variables. The full list of input features is given in Table A.5. During the training, events in the signal MC samples are given a target value of 1 and events in the background MC samples are given a target value of 0. The resulting DNN performance plots along with the input distribution and ROC curves are shown in Figure A.1. The ROC curves, *receiver operating characteristic curves* show the performance of the DNN classification model and allows to decide on categorisation bases on the significance level plotted on the same plot which is simply the approximation $\sigma \approx \frac{S}{\sqrt{B}}$ [118] where signal/background is too low ($S \ll B$). The events are then categorised into four making used of the DNN score distribution such that the expected significance is maximised, as shown in Figure 3.5 and Table 3.2. The DNN hyper-parameter settings in this final state is given in Table A.6. A cut-flow table, showing the number of events in the semi-leptonic channel and its DNN score categories is given in Table A.3 along with the number of events before any selection.

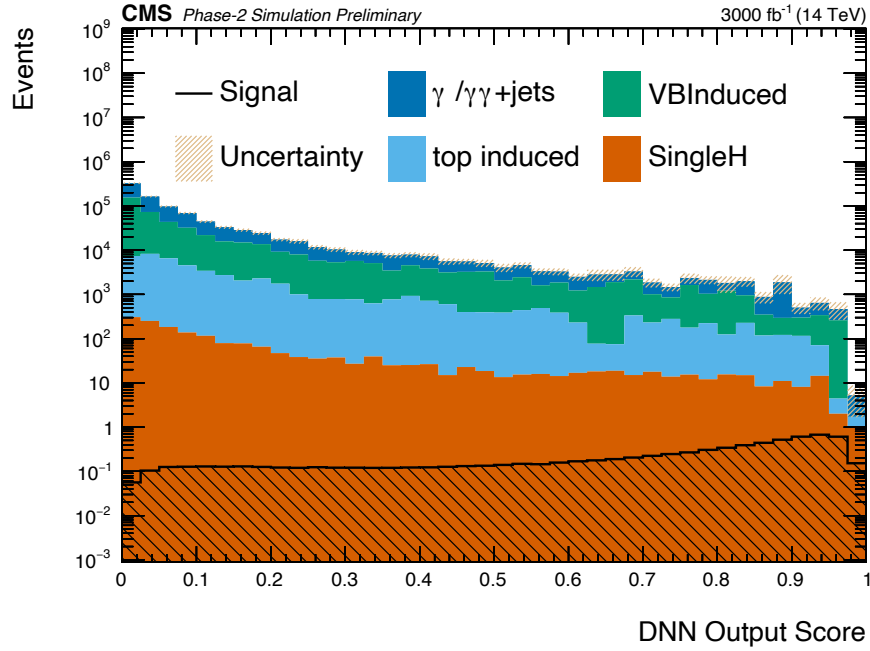


Figure 3.5 : DNN score distribution in the semi-leptonic final state.

Table 3.2 : Semi-leptonic final state DNN score categories.

Category	Definition
Category 1	$0.1 < DNNscore < 0.6$
Category 2	$0.6 < DNNscore < 0.8$
Category 3	$0.8 < DNNscore < 0.92$
Category 4	$0.92 < DNNscore < 1$

The categorisation allows to deal with different signal-to-background scenarios, and the Category 4 provides the best signal purity and significance. Di-photon mass distributions in the semi-leptonic final state and in its DNN categories is shown in Figure 3.6.

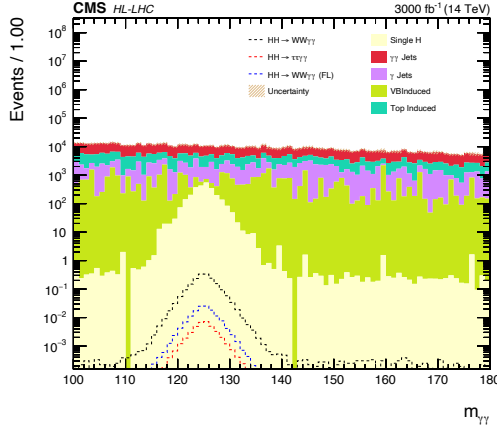


Figure 3.6a : Semi-leptonic final state

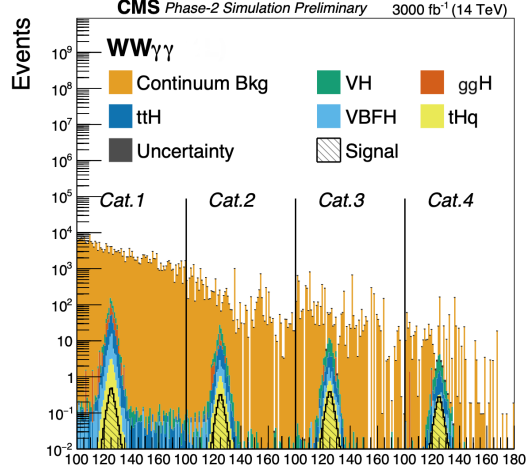


Figure 3.6b : DNN categories

Figure 3.6 : $m_{\gamma\gamma}$ distribution in the semi-leptonic final state (left) and in its DNN categories (right).

Fully-leptonic Final State

A cut-based analysis is performed in this final state. The events are required to have at least two oppositely charged good leptons (e^+e^- , $\mu^+\mu^-$, $e^\pm\mu^\mp$). These two leptons are separated from each other by an angular distance of at least 0.4. Since W boson decay to neutrinos along with leptons in this final state, a relatively larger missing transverse energy is required in the events. The lepton pair that have invariant mass value between 80 – 100 GeV, and the invariant mass of the leading electron and the leading photon in 5 GeV vicinity of Z boson mass in the events that have at least one electron are removed in order to suppress the Z boson contribution (Z veto). The selections are given as a list in Table 3.3. The di-photon invariant mass distribution in this final state is shown in Figure 3.7. A cut-flow report is also provided in Table A.7.

Single τ Final State

This final state contains only the events where one τ lepton of the pair is lost and has not been reconstructed at all. Two binary DNNs are trained in this final state following the same strategy as in semi-leptonic final state of the $WW\gamma\gamma$ channel of the double

Table 3.3 : Post-selections of the Fully-leptonic channel of $HH \rightarrow WW\gamma\gamma$.

Variable	Selection
$\Delta R(l, l)$	> 0.4
$p_T^{\gamma\gamma}$	$> 91 \text{ GeV}$
E_T^{miss}	$> 20 \text{ GeV}$
m_{ll}	$< 80 \text{ GeV}$ or $> 100 \text{ GeV}$
number of medium btagged jets	$= 0$
$ m_{e\gamma} - m_z $	$> 5 \text{ GeV}$

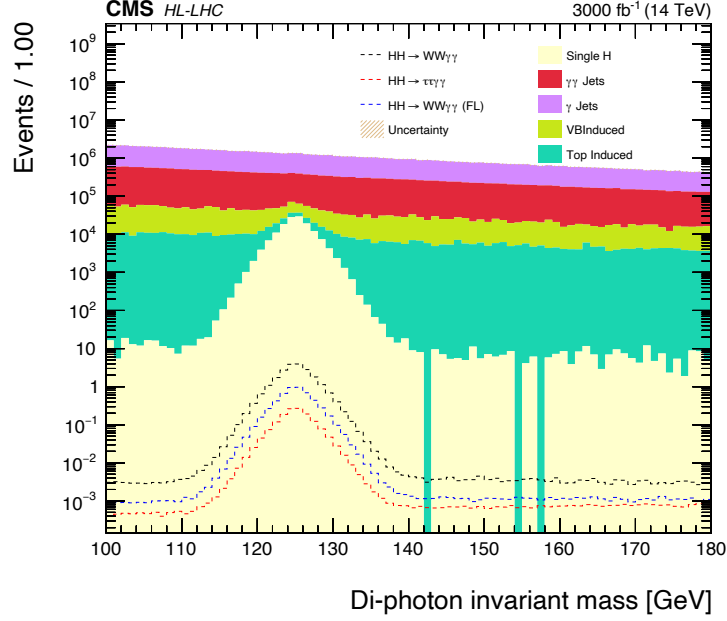


Figure 3.7 : $m_{\gamma\gamma}$ distribution in the fully-leptonic final state.

Higgs decay. The input variables to the DNN training are shown in Table A.8. The information of τ leptons are added to the training while removing the electron, muon and some jet information in this final state. Full distributions are given in Figure A.7 along with the training weights and the ROC curves. A cut-flow report showing the number of events in each category per sample is given in Table A.9. The resulting categorisation is shown in Table 3.4 and the DNN score in Figure 3.8. The di-photon mass distributions in each category is given in Figure 3.9.

Categories	Definition
Category 1	$0.1 < \text{DNN score} < 0.65$
Category 2	$0.65 \text{ DNN score} < 1$

Table 3.4 : Single τ final state DNN categories.

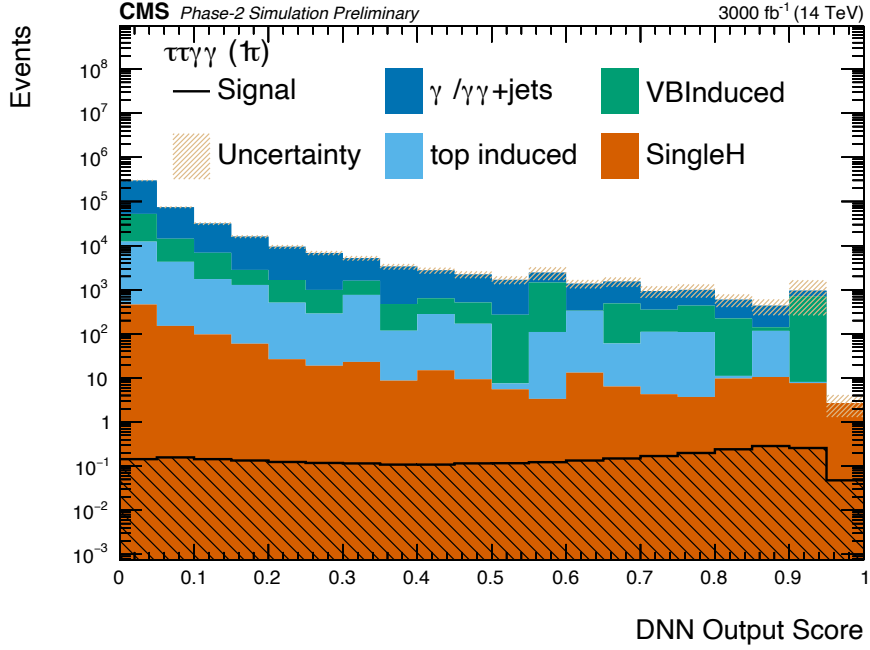


Figure 3.8 : DNN score distribution in the single τ final state.

Double τ Final State

A cut-based analysis is performed in this final state. The events that pass the di-photon preselection and have at least one oppositely charged τ lepton pair are included. The Z veto is applied such that the events with di-photon invariant mass in the 80-100 GeV range are rejected. The di-photon invariant mass distribution is shown in Figure 3.9 and a cut-flow report is provided in Table A.11.

3.2.4 Systematic uncertainties

The uncertainty on the measurements has two components; **statistical** uncertainty and **systematic** uncertainties. The former depends on the amount of data collected while the latter arises from the inaccuracy of the experimental setup due to detector or from theoretical calculations. The experiment-induced systematic uncertainties are caused by the insufficient knowledge of the detector response, and the theoretical systematic uncertainties affect the modelling of signal and background processes. These systematic uncertainties are taken as nuisance parameters in the statistical analysis such as normalisation uncertainty which affect the signal/background yield, and shape uncertainties which affect the distribution of the observables. The experimental systematic uncertainties have several sources;

- the uncertainty on the integrated luminosity which has a log-normal contribution

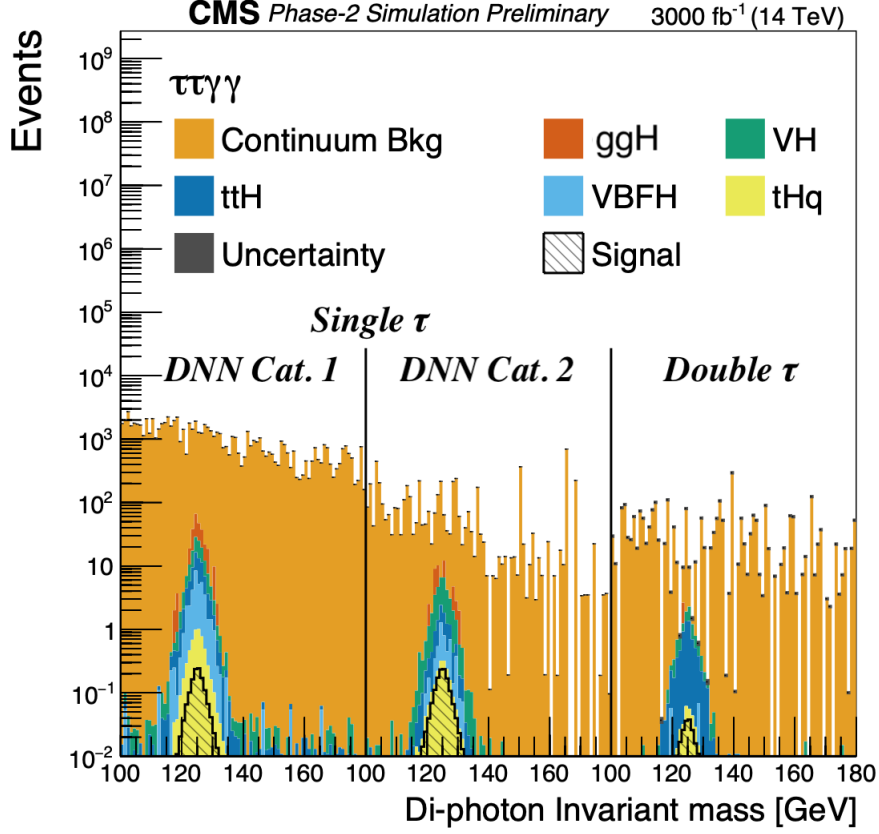


Figure 3.9 : $m_{\gamma\gamma}$ distributions in the DNN categories of single τ final state and double τ final state.

- uncertainty on the di-photon trigger and its mass resolution
- the lepton identification and reconstruction efficiency, with a log-normal contribution
- Jet energy scale which is computed by propagating the variation of the jet energy through the event reconstruction

The summary of the experimental systematic uncertainties is given in Table 3.5.

The theoretical uncertainties come from the choice of the PDF set, the uncertainty on the α_s and the renormalisation and factorisation QCD scale [119]. These uncertainties affect both the signal and background processes. Additionally, an uncertainty related to missing top quark mass effects has a contribution for the double Higgs signal process [120]. A summary of the theoretical uncertainties is given in Table 3.6.

Table 3.5 : Experimental uncertainties considered in this study. The values are recommended in the Yellow Report for HL-LHC studies [1].

Uncertainty Source	Input (%)
Luminosity	1
Diphoton trigger	2
$m_{\gamma\gamma}$ resolution	5
PhotonID	0.5/photon
electronID	0.5/electron
muonID	0.5/muon
tauID	2.5/tau
Jet energy Scale	1

Table 3.6 : Theoretical uncertainties considered on ggHH signal and single Higgs processes.

Process	Uncertainty Source		
	PDF + α_s (%)	QCD Scale (%)	m_{top} (%)
ggHH	± 3	+2.1/-4.9	+4.0/-18
ggH	+4.6/-6.7	± 3.2	-
VBFH	+0.5/-0.3	± 2.1	-
VH	+0.4/-0.7	± 1.8	-
ttH	+6/-9.2	± 3.5	-
tHq	+6.4/-14.7	± 3.6	-

4. RESULTS & CONCLUSION

Results are performed by fitting the invariant mass distributions of the leading and sub-leading photons using a binned maximum likelihood approach. The full systematic uncertainties are also applied in the form of nuisance parameters with log-normal distributions. These steps are performed with HIGGS COMBINE TOOL [121]. The correlations among different sources of systematic uncertainties are taken into account while different final states are considered as independent in the fit. The significance values obtained are shown in Table 4.1 for $WW\gamma\gamma$ final states and in Table 4.2 for $\tau\tau\gamma\gamma$ final states. A combination of all the two channels are shown in Table 4.3. An expected significance 0.22σ for signal is reported in the combination of $WW\gamma\gamma$ and $\tau\tau\gamma\gamma$ final states of the double Higgs production. Fitted distributions for the best DNN categories of semi-leptonic $WW\gamma\gamma$ channel and for the single τ DNN categories are shown in Figure 4.1 while all final state fit distributions are given in Figure A.11.

Table 4.1 : Significance numbers extracted in each category of one lepton final state, two leptonic final state and their combination.

Categories	Significance (stat)	Significance (stat+exp)	Significance (stat+exp+theory)
Category 1	0.0183	0.0138	0.0138
Category 2	0.0838	0.0416	0.0413
Category 3	0.0982	0.0908	0.0888
Category 4	0.1629	0.1246	0.1268
Semi-leptonic combined	0.2038	0.1722	0.1615
Fully-leptonic	0.0324	0.0294	0.0290
$WW\gamma\gamma$ combined	0.2180	0.2059	0.2056

Table 4.2 : Significance numbers extracted in each category of one tau final state, two tau's final state and their combination.

Categories	Significance (stat)	Significance (stat+exp)	Significance (stat+exp+theory)
Category 1	0.0143	0.0117	0.0110
Category 2	0.0551	0.0337	0.0336
1 τ	0.0782	0.0453	0.0452
2 τ s	0.0141	0.0129	0.0122
Combination	0.0792	0.0494	0.0493

Table 4.3 : Full Phase-II results of $WW\gamma\gamma$ and $\tau\tau\gamma\gamma$ processes with combination.

Categories	Significance (stat)	Significance (stat+exp)	Significance (stat+exp+theory)
$WW\gamma\gamma$	0.2180	0.2059	0.2056
$\tau\tau\gamma\gamma$	0.0792	0.0494	0.0493
Combination	0.2223	0.2182	0.2179

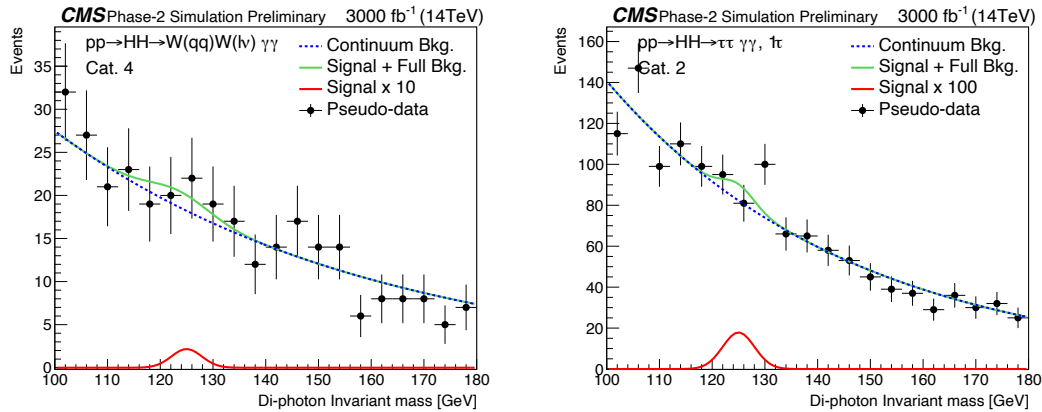


Figure 4.1 : $m_{\gamma\gamma}$ fit distributions in the semi-leptonic final state's DNN category 4 (left) and in the single τ final state's DNN category 2 (right).

REFERENCES

- [1] **Dainese, A., Mangano, M., Meyer, A.B., Nisati, A., Salam, G. and Vesterinen, M.A.** (2019). Report on the Physics at the HL-LHC, and Perspectives for the HE-LHC, **Technical Report**, CERN, Geneva, Switzerland, <https://cds.cern.ch/record/2703572>.
- [2] **Bagnaia, P. et al.** (1983). Evidence for $Z^0 \rightarrow e^+e^-$ at the CERN pp collider, *Physics Letters B*, 129(1), 130–140, [https://doi.org/10.1016/0370-2693\(83\)90744-X](https://doi.org/10.1016/0370-2693(83)90744-X).
- [3] **Arnison, G. et al.** (1983). Experimental observation of isolated large transverse energy electrons with associated missing energy at $\sqrt{s} = 540\text{GeV}$, *Phys. Lett. B*, 122, 103–116. 31 p, <https://doi.org/10.5170/CERN-1983-004.123>.
- [4] **Banner, M. et al.** (1983). Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN pp collider, *Physics Letters B*, 122(5), 476–485, [https://doi.org/10.1016/0370-2693\(83\)91605-2](https://doi.org/10.1016/0370-2693(83)91605-2).
- [5] **Arnison, G. et al.** (1983). Further evidence for charged intermediate vector bosons at the SPS collider, *Physics Letters B*, 129(3), 273–282, [https://doi.org/10.1016/0370-2693\(83\)90860-2](https://doi.org/10.1016/0370-2693(83)90860-2).
- [6] **Arnison, G. et al.** (1983). Experimental observation of lepton pairs of invariant mass around 95 GeV/c² at the CERN SPS collider, *Physics Letters B*, 126(5), 398–410, [https://doi.org/10.1016/0370-2693\(83\)90188-0](https://doi.org/10.1016/0370-2693(83)90188-0).
- [7] **CDF Collaboration** (1995). Observation of Top Quark Production in $\bar{p}p$ Collisions with the Collider Detector at Fermilab, *Phys. Rev. Lett.*, 74, 2626–2631, <https://doi.org/10.1103/PhysRevLett.74.2626>.
- [8] **D0 Collaboration** (1995). Observation of the Top Quark, *Phys. Rev. Lett.*, 74, 2632–2637, <https://doi.org/10.1103/PhysRevLett.74.2632>.
- [9] **CMS Collaboration** (2012). Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC, *Physics Letters B*, 716(1), 30–61, <https://doi.org/10.1016/j.physletb.2012.08.021>.
- [10] **ATLAS Collaboration** (2012). Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC,

- Physics Letters B*, 716(1), 1–29, <https://doi.org/10.1016/j.physletb.2012.08.020>.
- [11] **Higgs, P.W.** (1964). Broken Symmetries and the Masses of Gauge Bosons, *Phys. Rev. Lett.*, 13, 508–509, <https://doi.org/10.1103/PhysRevLett.13.508>.
 - [12] **Englert, F. and Brout, R.** (1964). Broken Symmetry and the Mass of Gauge Vector Mesons, *Phys. Rev. Lett.*, 13, 321–323, <https://doi.org/10.1103/PhysRevLett.13.321>.
 - [13] **Guralnik, G.S., Hagen, C.R. and Kibble, T.W.B.** (1964). Global Conservation Laws and Massless Particles, *Phys. Rev. Lett.*, 13, 585–587, <https://doi.org/10.1103/PhysRevLett.13.585>.
 - [14] **Salam, A. and Ward, J.** (1964). Electromagnetic and weak interactions, *Physics Letters*, 13(2), 168–171, [https://doi.org/10.1016/0031-9163\(64\)90711-5](https://doi.org/10.1016/0031-9163(64)90711-5).
 - [15] **'t Hooft, G. and Veltman, M.** (1972). Regularization and renormalization of gauge fields, *Nuclear Physics B*, 44(1), 189–213, [https://doi.org/10.1016/0550-3213\(72\)90279-9](https://doi.org/10.1016/0550-3213(72)90279-9).
 - [16] **Pettersson, T.S. and Lefèvre, P.** (1995). The Large Hadron Collider: conceptual design, **Technical Report**, CERN, <https://cds.cern.ch/record/291782>.
 - [17] **URL**, https://en.wikipedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg#metadata, accessed 26 Jan. 2022.
 - [18] **Weinberg, S.** (1967). A Model of Leptons, *Phys. Rev. Lett.*, 19, 1264–1266, <https://doi.org/10.1103/PhysRevLett.19.1264>.
 - [19] **Hagedorn, R.** (1965). Statistical thermodynamics of strong interactions at high energies, *Nuovo Cimento, Suppl.*, 3, 147–186, <https://cds.cern.ch/record/346206>.
 - [20] **Fukuda, Y. et al.** (1998). Measurements of the Solar Neutrino Flux from Super-Kamiokande's First 300 Days, *Phys. Rev. Lett.*, 81, 1158–1162, <https://doi.org/10.1103/PhysRevLett.81.1158>.
 - [21] **Planck Collaboration** (2014). Planck 2013 results. XVI. Cosmological parameters, *A&A*, 571, A16, <https://doi.org/10.1051/0004-6361/201321591>.
 - [22] **Solà, J.** (2013). Cosmological constant and vacuum energy: old and new ideas, *Journal of Physics: Conference Series*, 453, 012015, <https://doi.org/10.1088/1742-6596/453/1/012015>.
 - [23] **Griffiths, D.** (2008). *Introduction to Elementary Particles*, Physics textbook, Wiley.

- [24] **Khachatryan, V. et al.** (2015). Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV, *The European Physical Journal C*, 75(5), <http://doi.org/10.1140/epjc/s10052-015-3351-7>.
- [25] **Aad, G. et al.** (2016). Measurements of the Higgs boson production and decay rates and coupling strengths using pp collision data at $\sqrt{s} = 7$ and 8 TeV in the ATLAS experiment, *The European Physical Journal C*, 76(1), <http://doi.org/10.1140/epjc/s10052-015-3769-y>.
- [26] **Zyla, P.A. et al.** (2020). Review of Particle Physics, *PTEP*, 2020(8), 083C01, <https://doi.org/10.1093/ptep/ptaa104>.
- [27] **CMS Collaboration** (2013). Study of the Mass and Spin-Parity of the Higgs Boson Candidate via Its Decays to Z Boson Pairs, *Phys. Rev. Lett.*, 110, 081803, <https://doi.org/10.1103/PhysRevLett.110.081803>.
- [28] **CMS Collaboration** (2021). Measurements of production cross sections of the Higgs boson in the four-lepton final state in proton–proton collisions at $\sqrt{s} = 13$ TeV, *The European Physical Journal C*, 81(6), 488, <https://doi.org/10.1140/epjc/s10052-021-09200-x>.
- [29] **ATLAS Collaboration** (2020). Measurement of the properties of Higgs boson production at $\sqrt{s} = 13$ TeV in the $H \rightarrow \gamma\gamma$ channel using 139 fb^{-1} of pp collision data with the ATLAS experiment, *ATLAS-CONF-2020-026*, <https://cds.cern.ch/record/2725727>.
- [30] **Dittmaier, S. et al.** (2011). *Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables*, CERN Yellow Reports: Monographs, CERN, Geneva, <https://doi.org/10.5170/CERN-2011-002>.
- [31] **Dittmaier, S. et al.** (2012). *Handbook of LHC Higgs Cross Sections: 2. Differential Distributions*, CERN Yellow Reports: Monographs, CERN, Geneva, <https://doi.org/10.5170/CERN-2012-002>.
- [32] **Heinemeyer, S et al.** (2013). *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties: Report of the LHC Higgs Cross Section Working Group*, CERN Yellow Reports: Monographs, CERN, <https://doi.org/10.5170/CERN-2013-004>.
- [33] **CMS Collaboration** (2018). Observation of $t\bar{t}H$ Production, *Phys. Rev. Lett.*, 120, 231801, <https://doi.org/10.1103/PhysRevLett.120.231801>.
- [34] **CERN**, (2017), CERN Yellow Reports: Monographs, Vol 2 (2017): Handbook of LHC Higgs cross sections: 4. Deciphering the nature of the Higgs sector, <https://doi.org/10.23731/CYRM-2017-002>.
- [35] **CMS Collaboration** (2017). Observation of the SM scalar boson decaying to a pair of τ leptons with the CMS experiment at the LHC,

Technical Report, CERN, Geneva, <https://cds.cern.ch/record/2264522>.

- [36] **ATLAS and CMS Collaborations** (2016). Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s}=7$ and 8 TeV, *Journal of High Energy Physics*, 2016(8), [https://doi.org/10.1007/jhep08\(2016\)045](https://doi.org/10.1007/jhep08(2016)045).
- [37] **CMS Collaboration** (2020). Measurement of Higgs boson decay to a pair of muons in proton-proton collisions at $\sqrt{s}=13$ TeV, **Technical Report**, CERN, Geneva, <https://cds.cern.ch/record/2725423>.
- [38] **Glover, E. and van der Bij, J.** (1988). Higgs boson pair production via gluon fusion, *Nuclear Physics B*, 309(2), 282–294.
- [39] **de Florian, D. et al.** (2016). Differential Higgs boson pair production at next-to-next-to-leading order in QCD, *Journal of High Energy Physics*, 2016(9), [https://doi.org/10.1007/jhep09\(2016\)151](https://doi.org/10.1007/jhep09(2016)151).
- [40] **Di Micco, B. et al.** (2019). Higgs boson potential at colliders: status and perspectives, **Technical Report**, CERN, Geneva, <https://cds.cern.ch/record/2690841>.
- [41] **ATLAS Collaboration** (2020). Combination of searches for Higgs boson pairs in pp collisions at $\sqrt{s}=13$ TeV with the ATLAS detector, *Physics Letters B*, 800, 135103, <https://doi.org/10.1016/j.physletb.2019.135103>.
- [42] **CMS Collaboration** (2019). Combination of Searches for Higgs Boson Pair Production in Proton-Proton Collisions at $\sqrt{s}=13$ TeV, *Physical Review Letters*, 122(12), <https://doi.org/10.1103/physrevlett.122.121803>.
- [43] **Gouzevitch, M. and Carvalho, A.** (2020). A review of Higgs boson pair production, *Reviews in Physics*, 5, 100039, <https://doi.org/10.1016/j.revip.2020.100039>.
- [44] **Bezrukov, F. and Shaposhnikov, M.** (2008). The Standard Model Higgs boson as the inflaton, *Physics Letters B*, 659(3), 703–706, <https://doi.org/10.1016/j.physletb.2007.11.072>.
- [45] **Dawson, S. and Murphy, C.W.** (2017). Standard model EFT and extended scalar sectors, *Physical Review D*, 96(1), <https://doi.org/10.1103/physrevd.96.015041>.
- [46] **Branco, G., Ferreira, P., Lavoura, L., Rebelo, M., Sher, M. and Silva, J.P.** (2012). Theory and phenomenology of two-Higgs-doublet models, *Physics Reports*, 516(1-2), 1–102, <https://doi.org/10.1016/j.physrep.2012.02.002>.

- [47] **Bélusca-Maïto, H., Falkowski, A., Fontes, D., Romão, J.C. and Silva, J.P.** (2017). Higgs EFT for 2HDM and beyond, *The European Physical Journal C*, 77(3), <https://doi.org/10.1140/epjc/s10052-017-4745-5>.
- [48] **Gröber, R. and Mühlleitner, M.** (2011). Composite Higgs boson pair production at the LHC, *Journal of High Energy Physics*, 2011(6), [https://doi.org/10.1007/jhep06\(2011\)020](https://doi.org/10.1007/jhep06(2011)020).
- [49] **Contino, R.**, (2010), Tasi 2009 lectures: The Higgs as a Composite Nambu-Goldstone Boson, 1005.4269.
- [50] **Cheng, H.C., Dobrescu, B.A. and Gu, J.** (2014). Higgs mass from compositeness at a multi-TeV scale, *Journal of High Energy Physics*, 2014(8), [http://doi.org/10.1007/JHEP08\(2014\)095](http://doi.org/10.1007/JHEP08(2014)095).
- [51] **CERN** (2021). CERN Annual report 2020, **Technical Report**, CERN, Geneva, <https://doi.org/10.17181/AnnualReport2020>.
- [52] **Baconnier, Y., Brianti, G., Lebrun, P., Mathewson, A.G., Perin, R. and Baconnier, Y.** (1993). *LHC: the Large Hadron Collider accelerator project*, CERN, Geneva, <https://cds.cern.ch/record/257706>.
- [53] **URL**, <https://home.cern/news/press-release/cern/two-circulating-beams-bring-first-collisions-lhc>, accessed 8 Feb. 2022.
- [54] **Apollinari, G. et al.** (2017). *High-Luminosity Large Hadron Collider (HL-LHC): Technical Design Report V. 0.1*, CERN Yellow Reports: Monographs, CERN, Geneva, <https://doi.org/10.23731/CYRM-2017-004>.
- [55] **Mobs, E.** (2016). The CERN accelerator complex., *OPEN-PHO-ACCEL-2016-009-2*, <https://cds.cern.ch/record/2197559>.
- [56] **ATLAS Collaboration** (2008). The ATLAS Experiment at the CERN Large Hadron Collider, *Journal of Instrumentation*, 3(08), <https://doi.org/10.1088/1748-0221/3/08/s08003>.
- [57] **ALICE Collaboration** (2008). The ALICE experiment at the CERN LHC, *Journal of Instrumentation*, 3(08), <https://doi.org/10.1088/1748-0221/3/08/s08002>.
- [58] **LHCb Collaboration** (2008). The LHCb Detector at the LHC, *Journal of Instrumentation*, 3(08), <https://doi.org/10.1088/1748-0221/3/08/s08005>.
- [59] **LHCf Collaboration** (2006). *LHCf experiment: Technical Design Report*, Technical design report. LHCf, CERN, Geneva, <https://cds.cern.ch/record/926196>.

- [60] **TOTEM Collaboration** (2004). *Total cross-section, elastic scattering and diffraction dissociation at the Large Hadron Collider at CERN: TOTEM Technical Design Report*, CERN, Geneva, <https://cds.cern.ch/record/704349>.
- [61] **Mavromatos, N.E. and Mitsou, V.A.** (2016). Physics reach of MoEDAL at LHC: magnetic monopoles, supersymmetry and beyond., *EPJ Web Conf.*, 164, 04001. 17 p, <https://doi.org/10.1051/epjconf/201716404001>.
- [62] **URL**, <http://www.hep.ph.ic.ac.uk/~wstirlin/plots/plots.html>, accessed 9 Feb. 2022.
- [63] **URL**, <https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults>, accessed 9 Feb. 2022.
- [64] **URL**, <https://cms.cern/detector>, accessed 9 Feb. 2022.
- [65] **URL**, https://tikz.net/axis3d_cms/, accessed 9 Feb. 2022.
- [66] **Barney, D.** (2016). CMS Detector Slice, <https://cds.cern.ch/record/2120661>, CMS Collection.
- [67] **CMS Collaboration** (2014). Description and performance of track and primary-vertex reconstruction with the CMS tracker, *Journal of Instrumentation*, 9(10).
- [68] **CMS Collaboration** (2010). CMS tracking performance results from early LHC operation, *The European Physical Journal C*, 70(4), 1165–1192.
- [69] **The Tracker Group of the CMS Collaboration**, (2020), The CMS Phase-1 Pixel Detector Upgrade, 2012.14304.
- [70] **Collaboration, T.C.** (2014). Description and performance of track and primary-vertex reconstruction with the CMS tracker, *Journal of Instrumentation*, 9(10), P10009–P10009.
- [71] **CMS Collaboration** (1997). *The CMS electromagnetic calorimeter project: Technical Design Report*, Technical design report. CMS, CERN, Geneva, <https://cds.cern.ch/record/349375>.
- [72] **CMS Collaboration** (2013). Energy calibration and resolution of the CMS electromagnetic calorimeter in pp collisions at $\sqrt{s} = 7$ TeV, *Journal of Instrumentation*, 8(09), P09009–P09009, <https://doi.org/10.1088/1748-0221/8/09/p09009>.
- [73] **Adzic, P. et al.** (2007). Energy resolution of the barrel of the CMS Electromagnetic Calorimeter, *Journal of Instrumentation*, 2(04), P04004–P04004, <https://doi.org/10.1088/1748-0221/2/04/p04004>.
- [74] **CMS Collaboration** (1997). *The CMS hadron calorimeter project: Technical Design Report*, Technical design report. CMS, CERN, Geneva, <https://cds.cern.ch/record/357153>.

- [75] **URL**, (2012), Detector Drawings, <https://cds.cern.ch/record/1433717>.
- [76] **Layter, J.G.** (1997). *The CMS muon project: Technical Design Report*, Technical design report. CMS, CERN, Geneva, <https://cds.cern.ch/record/343814>.
- [77] **CMS Collaboration** (2018). Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s}=13$ TeV, *Journal of Instrumentation*, 13(06), P06015–P06015.
- [78] **CMS Collaboration** (2018). Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s}=13$ TeV, *Journal of Instrumentation*, 13(06), P06015–P06015.
- [79] **Isidro Gonzalez Caballero et al.** (2019). Study of the effects of radiation on the CMS Drift Tubes Muon Detector for the HL-LHC, **Technical Report**, CERN, Geneva, <https://doi.org/10.1088/1748-0221/14/12/C12010,1912.06178>.
- [80] **Acosta, D. et al.** (2008). Efficiency of finding muon track trigger primitives in CMS cathode strip chambers, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 592(1-2), 26–37, <https://doi.org/10.1016/j.nima.2008.03.118>.
- [81] **Kumari, P et al.** (2020). Improved RPC for CMS muon system upgrade for HL-LHC. Improved-RPC for the CMS muon system upgrade for the HL-LHC, **Technical Report**, CERN, Geneva, <https://doi.org/10.1088/1748-0221/15/11/C11012,2005.11396>.
- [82] **The TOTEM Collaboration** (2011). First measurement of the total proton-proton cross-section at the LHC energy of $\sqrt{s} = 7$ TeV, *EPL (Europhysics Letters)*, 96(2), 21002.
- [83] **Khachatryan, V. et al.** (2017). The CMS trigger system, *Journal of Instrumentation*, 12(01), P01020–P01020.
- [84] **Cittolin, S., Rácz, A. and Sphicas, P.** (2002). *CMS The TriDAS Project: Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger. CMS trigger and data-acquisition project*, Technical design report. CMS, CERN, Geneva, <https://cds.cern.ch/record/578006>.
- [85] **Bayatyan, G.L., Della Negra, M., Foà, L., Hervé, A. and Petrilli, A.** (2005). *CMS computing: Technical Design Report*, Technical design report. CMS, CERN, Geneva, <https://cds.cern.ch/record/838359>.
- [86] **Aleksan, R et al.** (2013). Physics Briefing Book: Input for the Strategy Group to draft the update of the European Strategy for Particle Physics, **Technical Report**, CERN, Geneva, <https://cds.cern.ch/record/1433717>.

ch/record/1628377, open Symposium held in Cracow from 10th to 12th of September 2012.

- [87] **CMS Collaboration** (2020). The CMS Collaboration contribution to Snowmass 2021, *Letter of Interest*, https://www.snowmass21.org/docs/files/summaries/EF/SNOWMASS21-EF1_EF10-RF5_RF7_CMSCollaboration-109.pdf, accessed 4 Mar. 2022.
- [88] **ATLAS, C.** (2012). Letter of Intent for the Phase-II Upgrade of the ATLAS Experiment, **Technical Report**, CERN, Geneva, <https://cds.cern.ch/record/1502664>, draft version for comments.
- [89] **Contardo, D., Klute, M., Mans, J., Silvestris, L. and Butler, J.** (2015). Technical Proposal for the Phase-II Upgrade of the CMS Detector, **Technical Report**, CERN, Geneva, <https://cds.cern.ch/record/2020886>.
- [90] **Pöttgen, R.**, (2016). Proton-Proton Collisions, Search for Dark Matter with ATLAS, Springer International Publishing, pp.45–60, https://doi.org/10.1007/978-3-319-41045-6_4.
- [91] **Webber, B.** (1984). A QCD model for jet fragmentation including soft gluon interference, *Nuclear Physics B*, 238(3), 492–528, [https://doi.org/10.1016/0550-3213\(84\)90333-x](https://doi.org/10.1016/0550-3213(84)90333-x).
- [92] **Andersson, B., Gustafson, G., Ingelman, G. and Sjöstrand, T.** (1983). Parton fragmentation and string dynamics, *Physics Reports*, 97(2-3), 31–145, [https://doi.org/10.1016/0370-1573\(83\)90080-7](https://doi.org/10.1016/0370-1573(83)90080-7).
- [93] **Skands, P.**, (2012), QCD for Collider Physics, 1104.2863.
- [94] **Agostinelli, S. et al.** (2003). Geant4—a simulation toolkit, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3), 250–303, [https://doi.org/10.1016/S0168-9002\(03\)01368-8](https://doi.org/10.1016/S0168-9002(03)01368-8).
- [95] **Sekmen, S.** (2017). Recent Developments in CMS Fast Simulation, *PoS, ICHEP2016*, 181. 6p, <https://doi.org/10.22323/1.282.0181>, 5 pages, 5 figures. to appear in the proceedings of the 38th International Conference of High Energy Physics (ICHEP), Chicago, Aug. 3-10, 2016, 1701.03850.
- [96] **Lukas, W.** (2012). Fast Simulation for ATLAS: Atlfast-II and ISF, *J. Phys. Conf. Ser.*, 396, 022031.
- [97] **Selvaggi, M.** (2014). DELPHES 3: A modular framework for fast-simulation of generic collider experiments, *J. Phys. Conf. Ser.*, 523, 012033.
- [98] **CMS Collaboration** (2009). Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET, **Technical Report**, CERN, Geneva, <https://cds.cern.ch/record/1194487>.

- [99] **Buskulic, D. et al.** (1995). Performance of the ALEPH detector at LEP, *Nucl. Instrum. Meth. A*, 360, 481–506, <https://cds.cern.ch/record/272484>.
- [100] **de Favereau, J., Delaere, C., Demin, P., Giammanco, A., Lemaître, V., Mertens, A. and Selvaggi, M.** (2014). DELPHES 3: a modular framework for fast simulation of a generic collider experiment, *Journal of High Energy Physics*, 2014(2), [https://doi.org/10.1007/jhep02\(2014\)057](https://doi.org/10.1007/jhep02(2014)057).
- [101] **Cacciari, M., Salam, G.P. and Soyez, G.** (2012). FastJet user manual, *The European Physical Journal C*, 72(3), <https://doi.org/10.1140/epjc/s10052-012-1896-2>.
- [102] **CMS Collaboration** (2013). Electron performance with 19.6 fb^{-1} of data collected at $\sqrt{s} = 8 \text{ TeV}$ with the CMS detector., <https://cds.cern.ch/record/1523273>.
- [103] **Nason, P.** (2004). A New Method for Combining NLO QCD with Shower Monte Carlo Algorithms, *Journal of High Energy Physics*, 2004(11), 040–040, <https://doi.org/10.1088/1126-6708/2004/11/040>.
- [104] **Frixione, S., Nason, P. and Oleari, C.** (2007). Matching NLO QCD computations with parton shower simulations: the POWHEG method, *Journal of High Energy Physics*, 2007(11), 070–070, <https://doi.org/10.1088/1126-6708/2007/11/070>.
- [105] **Alioli, S., Nason, P., Oleari, C. and Re, E.** (2010). A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX, *Journal of High Energy Physics*, 2010(6), [https://doi.org/10.1007/jhep06\(2010\)043](https://doi.org/10.1007/jhep06(2010)043).
- [106] **Heinrich, G., Jones, S.P., Kerner, M., Luisoni, G. and Scyboz, L.** (2019). Probing the trilinear Higgs boson coupling in di-Higgs production at NLO QCD including parton shower effects, *Journal of High Energy Physics*, 2019(6), [https://doi.org/10.1007/jhep06\(2019\)066](https://doi.org/10.1007/jhep06(2019)066).
- [107] **Sjöstrand, T. et al.** (2015). An introduction to PYTHIA 8.2, *Computer Physics Communications*, 191, 159–177, <https://doi.org/10.1016/j.cpc.2015.01.024>.
- [108] **Alwall, J.** (2014). The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations, *Journal of High Energy Physics*, 2014(7), [https://doi.org/10.1007/jhep07\(2014\)079](https://doi.org/10.1007/jhep07(2014)079).
- [109] **Artoisenet, P., Frederix, R., Mattelaer, O. and Rietkerk, R.** (2013). Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations, *Journal of High Energy Physics*, 2013(3), [https://doi.org/10.1007/jhep03\(2013\)015](https://doi.org/10.1007/jhep03(2013)015).

- [110] **Frederix, R. and Frixione, S.** (2012). Merging meets matching in MC@NLO, *Journal of High Energy Physics*, 2012(12), [https://doi.org/10.1007/jhep12\(2012\)061](https://doi.org/10.1007/jhep12(2012)061).
- [111] **Bothmann, E. et al.** (2019). Event Generation with Sherpa 2.2, *SciPost Phys.*, 7, 34, <https://doi.org/10.21468/SciPostPhys.7.3.034>.
- [112] **Bertolini, D., Harris, P., Low, M. and Tran, N.** (2014). Pileup per particle identification, *Journal of High Energy Physics*, 2014(10), [http://dx.doi.org/10.1007/JHEP10\(2014\)059](http://dx.doi.org/10.1007/JHEP10(2014)059).
- [113] **Kirschenmann, H.** (2013). Jet performance in CMS, **Technical Report**, CERN, Geneva, <https://cds.cern.ch/record/1627818>.
- [114] **URL,** https://github.com/recotoolsbenchmarks/DelphesNtuplizer/blob/master/cards/CMS_PhaseII_200PU_Snowmass2021_v0.tcl, accessed 7 Mar. 2022.
- [115] **Abadi, M et al.,** (2015), TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, <https://www.tensorflow.org/>, software available from tensorflow.org.
- [116] **Chollet, François et al.,** (2015), Keras, <https://keras.io>.
- [117] **David, P.** (2021). Readable and efficient HEP data analysis with bamboo, *EPJ Web Conf.*, 251, 03052, 2103.01889.
- [118] **Cowen, G. and Gross, E.** (2008). Discovery significance with statistical uncertainty in the background estimate, *ATLAS Statistics Forum*, <https://www.pp.rhul.ac.uk/~cowan/stat/notes/SigCalcNote.pdf>.
- [119] **Cepeda, M. et al.,** (2019), Higgs Physics at the HL-LHC and HE-LHC, [arXiv:1902.00134](https://arxiv.org/abs/1902.00134).
- [120] **Grazzini, M., Heinrich, G., Jones, S., Kallweit, S., Kerner, M., Lindert, J.M. and Mazzitelli, J.** (2018). Higgs boson pair production at NNLO with top quark mass effects, *Journal of High Energy Physics*, 2018(5), [http://dx.doi.org/10.1007/JHEP05\(2018\)059](http://dx.doi.org/10.1007/JHEP05(2018)059).
- [121] **The LHC Higgs Combination Group** (2011). Procedure for the LHC Higgs boson search combination in Summer 2011, **Technical Report**, CERN, Geneva, <https://cds.cern.ch/record/1379837>.

APPENDICES

APPENDIX A.1 : Python module of the analysis used in Bamboo framework and DNN setup for the $WW\gamma\gamma$ semi-leptonic final state

APPENDIX A.2 : Monte Carlo samples used in the analysis and cut-flow report in the semi-leptonic final state of $HH \rightarrow WW\gamma\gamma$

APPENDIX A.3 : DNN input variables, their distributions, DNN performance plots and parameters for the semi-leptonic channel of $HH \rightarrow WW\gamma\gamma$

APPENDIX A.4 : Cut-flow report in the fully-leptonic final state of $HH \rightarrow WW\gamma\gamma$

APPENDIX A.5 : DNN input variables, their distributions, DNN performance plots and the cut-flow table for the single τ channel of $HH \rightarrow \tau\tau\gamma\gamma$

APPENDIX A.6 : Cut-flow report for the double τ channel of $HH \rightarrow \tau\tau\gamma\gamma$

APPENDIX A.7 : Fit plots for each final state

APPENDIX A.1

Listing A.1 : Python module of the analysis used in Bamboo framework

```
1 import logging
2 from bamboo.analysisutils import loadPlotIt
3 import os.path
4 import copy
5 from bamboo.analysismodules import AnalysisModule, /
6 HistogramsModule
7
8
9 class SnowmassExample(CMSPhase2SimRTBHistoModule):
10     def addArgs(self, parser):
11         super().addArgs(parser)
12         parser.add_argument("--mvaSkim",
13                             action="store_true",
14                             help="Produce_skims")
15         parser.add_argument("--datacards", action="store_true",
16                             help="Produce_histograms_for_datacards")
17         parser.add_argument("--mvaEval", action="store_true",
18                             help="Import_MVA_model_and_evaluate_it")
19
20     def definePlots(self, t, noSel, sample=None, sampleCfg=None):
21         from bamboo.plots import Plot, CutFlowReport, SummedPlot
22         from bamboo.plots import EquidistantBinning as EqB
23         from bamboo import treefunctions as op
24
25         plots = []
26
27         noSel = noSel.refine("genweight", weight=t.genweight)
28
29         # yields
30         yields_OneL = CutFlowReport(
31             "yields_OneL", recursive=True, printInLog=True)
32         yields_TwoL = CutFlowReport(
33             "yields_TwoL", recursive=True, printInLog=True)
34         yields_ZeroL = CutFlowReport(
35             "yields_ZeroL", recursive=True, printInLog=True)
36         yields_OneTau = CutFlowReport(
37             "yields_OneTau", recursive=True, printInLog=True)
38         yields_TwoTaus = CutFlowReport(
39             "yields_TwoTaus", recursive=True, printInLog=True)
40
41         yields_OneL.add(noSel, title='noSel')
42         yields_TwoL.add(noSel, title='noSel')
43         yields_ZeroL.add(noSel, title='noSel')
44         yields_OneTau.add(noSel, title='noSel')
45         yields_TwoTaus.add(noSel, title='noSel')
46
47         plots.append(yields_OneL)
48         plots.append(yields_TwoL)
49         plots.append(yields_ZeroL)
```

```

50     plots.append(yields_OneTau)
51     plots.append(yields_TwoTaus)
52
53     # select photons in the detector acceptance
54     photons = op.select(t.gamma, lambda ph: op.AND(
55         op.abs(ph.eta) < 2.5, ph.pt > 25.))
56
57     # sort photons by pT
58     sort_ph = op.sort(photons, lambda ph: -ph.pt)
59
60     # select photons with loose ISO & ID
61     isoPhotons = op.select(
62         sort_ph, lambda ph: ph.isopass & (
63             1 << 0))
64     idPhotons = op.select(
65         isoPhotons, lambda ph: ph.idpass & (1 << 0))
66
67     # select electrons w loose ISO&ID and clean them
68     # w.r.t good photons
69     electrons = op.select(t.elec, lambda el: op.AND(
70         el.pt > 10., op.abs(el.eta) < 2.5))
71     sort_el = op.sort(electrons, lambda el: -el.pt)
72
73     isoElectrons = op.select(
74         clElectrons, lambda el: el.isopass & (1 << 0))
75
76     idElectrons = op.select(
77         isoElectrons, lambda el: el.idpass & (1 << 0))
78
79     clElectrons = op.select(
80         idElectrons, lambda el: op.AND(
81             op.NOT(op.rng_any(
82                 idPhotons, lambda ph: op.deltaR(el.p4,ph.p4)<0.4)),
83             ))
84
85     # select muons with tight ISO & ID and clean them
86     # w.r.t good photons and electrons
87     muons = op.select(t.muon, lambda mu: op.AND(
88         mu.pt > 10., op.abs(mu.eta) < 2.5))
89
90     sort_mu = op.sort(clMuons, lambda mu: -mu.pt)
91
92     isoMuons = op.select(idMuons, lambda mu: mu.isopass&(1<<2))
93
94     idMuons = op.select(sort_mu, lambda mu: mu.idpass & (1 << 2))
95
96     clMuons = op.select(muons, lambda mu: op.AND(
97         op.NOT(op.rng_any(
98             idPhotons, lambda ph: op.deltaR(mu.p4,ph.p4)<0.4)),
99         op.NOT(op.rng_any(
100             clElectrons, lambda j: op.deltaR(mu.p4,j.p4)<0.4))))
101
102     # select taus with loose ISO and clean them
103     # w.r.t good photons & electrons & muons
104     taus = op.sort(op.select(t.tau, lambda tau: op.AND(
105         tau.pt > 20., op.abs(tau.eta)<2.5)), lambda tau:-tau.pt)
106
107     isoTaus = op.select(clTaus, lambda tau: tau.isopass&(1<<2))

```

```

108
109 clTaus = op.select(taus, lambda tau: op.AND(
110     op.NOT(op.rng_any(
111         idPhotons, lambda ph: op.deltaR(tau.p4, ph.p4) < 0.2)),
112     op.NOT(op.rng_any(
113         clElectrons, lambda el: op.deltaR(tau.p4, el.p4) < 0.2)),
114     op.NOT(op.rng_any(
115         clMuons, lambda mu: op.deltaR(tau.p4, mu.p4) < 0.2))
116     ))
117
118 # select jets with tight ID and clean
119 # them w.r.t good photons & electrons & muons & taus
120 jets = op.select(t.jetpuppi, lambda jet: op.AND(
121     jet.pt > 30., op.abs(jet.eta) < 5))
122
123 sort_jets = op.sort(jets, lambda jet: -jet.pt)
124
125 clJets = op.select(sort_jets, lambda j: op.AND(
126     op.NOT(op.rng_any(
127         idPhotons, lambda ph: op.deltaR(ph.p4, j.p4) < 0.4)),
128     op.NOT(op.rng_any(
129         clElectrons, lambda el: op.deltaR(el.p4, j.p4) < 0.4)),
130     op.NOT(op.rng_any(
131         clMuons, lambda mu: op.deltaR(mu.p4, j.p4) < 0.4)),
132     op.NOT(op.rng_any(
133         clTaus,
134         lambda tau: op.deltaR(j.p4, tau.p4) < 0.4))
135     ))
136
137 idJets = op.select(clJets, lambda j: j.idpass & (1 << 2))
138
139 # select b-jets with tight ID
140 bJets = op.select(idJets, lambda j: j.btag & (1 << 1))
141
142 # define variables for ease of use
143 # di-photon invariant mass
144 mGG = op.invariant_mass(idPhotons[0].p4, idPhotons[1].p4)
145 mTauTau = op.invariant_mass(
146     isoTaus[0].p4, isoTaus[1].p4) # di-tau invariant mass
147 pTGG = (op.sum(idPhotons[0].p4, idPhotons[1].p4)).pt()
148 mJets = op.invariant_mass(
149     idJets[0].p4, idJets[1].p4) # di-jet invariant mass
150 # sub-di-jet invariant mass
151 mJets_SL = op.invariant_mass(idJets[1].p4, idJets[2].p4)
152
153 # Fully leptonic FL invmasses
154 # di-electron invariant mass
155 mE = op.invariant_mass(idElectrons[0].p4, idElectrons[1].p4)
156 # di-muon invariant mass
157 mMu = op.invariant_mass(isoMuons[0].p4, isoMuons[1].p4)
158 # e-mu system invariant mass
159 mEMu = op.invariant_mass(idElectrons[0].p4, isoMuons[0].p4)
160
161 # another set of variables
162 nElec = op.rng_len(idElectrons) # number of electrons
163 nMuon = op.rng_len(isoMuons) # number of muons
164 nJet = op.rng_len(idJets) # number of jets
165 nTau = op.rng_len(isoTaus) # number of taus

```

```

166
167 pT_mGGL = op.product(idPhotons[0].pt, op.pow(mGG, -1))
168 pT_mGGSL = op.product(idPhotons[1].pt, op.pow(mGG, -1))
169 E_mGGL=op.product(idPhotons[0].p4.energy(), op.pow(mGG, -1))
170 E_mGGSL=op.product(idPhotons[1].p4.energy(), op.pow(mGG, -1))
171
172 # selections for efficiency check
173 sel1_p = noSel.refine("2Photon", cut=op.AND(
174     (op.rng_len(sort_ph) >= 2), (sort_ph[0].pt > 35.)))
175 sel2_p = sel1_p.refine("idPhoton", cut=op.AND(
176     (op.rng_len(idPhotons) >= 2), (idPhotons[0].pt > 35.)))
177
178 # selections of the events with inv mass of photons
179 # in the 100-180 window
180 hasInvM = sel2_p.refine("hasInvM", cut=op.AND(
181     (op.in_range(100, op.invariant_mass(
182         idPhotons[0].p4, idPhotons[1].p4), 180))
183     ))
184
185 ## Categories ##
186
187 # selections for semileptonic final state
188 semiLeptonic = hasInvM.refine("semiLeptonic",
189     cut=op.OR(
190         op.AND(nElec == 1, nMuon == 0),
191         op.AND(nElec == 0, nMuon == 1)))
192 yields_OneL.add(semiLeptonic, title='semiLeptonic')
193
194 hasOneEl = hasInvM.refine(
195     "hasOneEl", cut=op.AND(nElec == 1, nMuon == 0))
196 hasOneMu = hasInvM.refine(
197     "hasOneMu", cut=op.AND(nElec == 0, nMuon == 1))
198
199 # selections for fully-leptonic final state
200 fullyLeptonic = hasInvM.refine('fullyLeptonic',
201     cut=op.AND(
202         op.OR(
203             op.AND(
204                 op.AND(nElec >= 2, nMuon == 0),
205                 idElectrons[0].charge != idElectrons[1].charge,
206                 op.NOT(
207                     op.deltaR(idElectrons[0].p4, idElectrons[1].p4) < 0.4),
208                 op.OR(mE < 80, mE > 100), op.abs(m_eg - m_Z) > 5),
209             op.AND(
210                 op.AND(nElec >= 1, nMuon == 1),
211                 idElectrons[0].charge != isoMuons[0].charge,
212                 op.NOT(op.deltaR(
213                     idElectrons[0].p4, isoMuons[0].p4) < 0.4),
214                 op.OR(mEMu < 80, mEMu > 100), op.abs(m_eg - m_Z) > 5),
215             op.AND(
216                 op.AND(nElec == 1, nMuon >= 1),
217                 idElectrons[0].charge != isoMuons[0].charge,
218                 op.NOT(op.deltaR(
219                     idElectrons[0].p4, isoMuons[0].p4) < 0.4),
220                 op.OR(mEMu < 80, mEMu > 100), op.abs(m_eg - m_Z) > 5),
221             op.AND(
222                 op.AND(nMuon >= 2, nElec == 0),
223                 isoMuons[0].charge != isoMuons[1].charge,

```

```

224         op.NOT(op.deltaR(isoMuons[0].p4, isoMuons[1].p4) < 0.4),
225         op.OR(mMu < 80, mMu > 100))),
226     pTGG > 91,
227     op.AND(thirdEl < 10, thirdMu < 10),
228     op.rng_len(bJets) < 1,
229     met[0].pt > 20
230 ))
231
232 ##### fully-leptonic final state variables #####
233 m_eg = op.invariant_mass(idElectrons[0].p4, idPhotons[0].p4)
234 m_Z = 91.18
235
236 thirdEl = op.switch(op.rng_len(idElectrons) < 3,
237                     op.c_float(0.), idElectrons[2].pt)
238 thirdMu = op.switch(op.rng_len(isoMuons) < 3,
239                     op.c_float(0.), isoMuons[2].pt)
240 #####
241
242 yields_TwoL.add(fullyLeptonic, title='fullyLeptonic')
243
244 # selections for tautau final states
245 c3 = hasInvM.refine("hasOneTauNoLept", cut=op.AND(
246     nTau == 1,
247     op.rng_len(idElectrons) == 0,
248     op.rng_len(isoMuons) == 0))
249 yields_OneTau.add(c3, "One_Tau_No_Lept")
250
251 c4 = hasInvM.refine("hasTwoTaus", cut=op.AND(
252     nTau >= 2,
253     op.rng_len(idElectrons) == 0,
254     op.rng_len(isoMuons) == 0))
255 yields_TwoTaus.add(c4, "Two_Taus")
256
257 ##### Z veto #####
258 c4_Zveto = c4.refine("hasTwoTaus_Zveto", cut=op.NOT(
259     op.in_range(80, mTauTau, 100)))
260
261 ## End of Categories ##
262
263 # plots
264 # semiLeptonic
265 plots.append(Plot.make1D(
266     "LeadingPhotonPtOneL",
267     idPhotons[0].pt,
268     semiLeptonic,
269     EqB(30, 0., 300.),
270     title="Leading_Photon_pT")
271 )
272 plots.append(
273     Plot.make1D(
274         "SubLeadingPhotonPtOneL",
275         idPhotons[1].pt, semiLeptonic, EqB(
276             30, 0., 300.), title="SubLeading_Photon_pT"))
277 plots.append(
278     Plot.make1D(
279         "LeadingPhotonEtaOneL",
280         idPhotons[0].eta, semiLeptonic, EqB(
281             80, -4., 4.), title="Leading_Photon_eta"))

```

```

282 plots.append(
283     Plot.make1D(
284         "SubLeadingPhotonEtaOneL",
285         idPhotons[1].eta, semiLeptonic, EqB(
286             80, -4., 4.), title="SubLeading_Photon_eta"))
287 plots.append(
288     Plot.make1D(
289         "LeadingPhotonPhiOneL",
290         idPhotons[0].phi, semiLeptonic, EqB(
291             100, -3.5, 3.5), title="Leading_Photon_phi"))
292 plots.append(
293     Plot.make1D(
294         "SubLeadingPhotonPhiOneL",
295         idPhotons[1].phi, semiLeptonic, EqB(
296             100, -3.5, 3.5), title="SubLeading_Photon_phi"))
297 plots.append(
298     Plot.make1D(
299         "nElectronsOneL",
300         nElec,
301         semiLeptonic,
302         EqB(10, 0., 10.), title="Number_of_electrons"))
303 plots.append(
304     Plot.make1D(
305         "nMuonsOneL",
306         nMuon, semiLeptonic,
307         EqB(10, 0., 10.), title="Number_of_Muons"))
308 plots.append(
309     Plot.make1D(
310         "nJetsOneL",
311         nJet, semiLeptonic,
312         EqB(10, 0., 10.), title="Number_of_Jets"))
313 plots.append(
314     Plot.make1D(
315         "LeadingPhotonpT_mGGLsemiLeptonic",
316         pT_mGGL, semiLeptonic, EqB(100, 0., 5.),
317         title="Leading_Photon_p_{T}/m_{\gamma\gamma}"))
318 plots.append(
319     Plot.make1D(
320         "SubLeadingPhotonpT_mGGLsemiLeptonic",
321         pT_mGGL, semiLeptonic, EqB(100, 0., 5.),
322         title="SubLeading_Photon_p_{T}/m_{\gamma\gamma}"))
323 plots.append(
324     Plot.make1D(
325         "LeadingPhotonE_mGGLsemiLeptonic",
326         E_mGGL, semiLeptonic, EqB(100, 0., 5.),
327         title="Leading_Photon_E/m_{\gamma\gamma}"))
328 plots.append(
329     Plot.make1D(
330         "SubLeadingPhotonE_mGGLsemiLeptonic",
331         E_mGGL, semiLeptonic, EqB(100, 0., 5.),
332         title="SubLeading_Photon_E/m_{\gamma\gamma}"))
333 plots.append(
334     Plot.make1D(
335         "MET",
336         metPt, semiLeptonic,
337         EqB(80, 0., 800.), title="MET"))
338 plots.append(
339     Plot.make1D(

```

```

340         "Inv_mass_ggsemiLeptonic",
341         mGG, semiLeptonic,
342         EqB(80, 100., 180.), title="m_{\gamma\gamma}")
343     plots.append(
344         Plot.make1D(
345             "Inv_mass_ggsemiLeptonic_135",
346             mGG,
347             semiLeptonic, EqB(20, 115., 135.),
348             title="m_{\gamma\gamma}")
349
350     # Leading electron Plots
351     plots.append(Plot.make1D(
352         "ElectronpT", idElectrons[0].pt, hasOneEl,
353         EqB(30, 0., 300.), title='Leading_Electron_pT'))
354     plots.append(Plot.make1D(
355         "ElectronE", idElectrons[0].p4.E(), hasOneEl,
356         EqB(50, 0., 500.), title='Leading_Electron_E'))
357     plots.append(Plot.make1D(
358         "ElectronEta", idElectrons[0].eta, hasOneEl,
359         EqB(80, -4., 4.), title='Leading_Electron_eta'))
360     plots.append(Plot.make1D(
361         "ElectronPhi", idElectrons[0].phi, hasOneEl,
362         EqB(100, -3.5, 3.5), title='Leading_Electron_phi'))
363
364     # Leading muon Plots
365     plots.append(Plot.make1D(
366         "MuonpT", isoMuons[0].pt, hasOneMu, EqB(
367         30, 0., 100.), title='Leading_Muon_pT'))
368     plots.append(Plot.make1D(
369         "MuonE", isoMuons[0].p4.E(), hasOneMu, EqB(
370         50, 0., 500.), title='Leading_Muon_E'))
371     plots.append(Plot.make1D(
372         "MuonEta", isoMuons[0].eta, hasOneMu, EqB(
373         80, -4., 4.), title='Leading_Muon_eta'))
374     plots.append(Plot.make1D(
375         "MuonPhi", isoMuons[0].phi, hasOneMu, EqB(
376         100, -3.5, 3.5), title='Leading_Muon_phi'))
377
378     ### DNN variables ###
379
380     semiLeptonic = hasInvM.refine("semiLeptonic", cut=op.OR(
381         op.AND(nElec == 1, nMuon == 0),
382         op.AND(nElec == 0, nMuon == 1)
383     ))
384
385
386     hasOneJ = semiLeptonic.refine(
387         "semiLeptonic_hasOneJ", cut=op.rng_len(idJets) == 1)
388     hasTwoJ = semiLeptonic.refine(
389         "semiLeptonic_hasTwoJ", cut=op.rng_len(idJets) == 2)
390     hasthreeJ = semiLeptonic.refine(
391         "semiLeptonic_hasthreeJ", cut=op.rng_len(idJets) == 3)
392
393     plots.append(Plot.make1D(
394         "semiLeptonic_hasOneJ_jetpt",
395         idJets[0].pt, hasOneJ, EqB(
396         30, 0., 300.), title="Leading_Jet_p_T"))
397     plots.append(Plot.make1D(

```

```

398     "semiLeptonic_hasOneJ_jeteta",
399     idJets[0].eta, hasOneJ, EqB(
400     80, -5.5, 5.5), title="Leading_Jet_#eta"))
401 plots.append(Plot.make1D(
402     "semiLeptonic_hasOneJ_jetphi",
403     idJets[0].phi, hasOneJ, EqB(
404     100, -3.5, 3.5), title="Leading_Jet_#phi"))
405 plots.append(Plot.make1D(
406     "semiLeptonic_hasOneJ_jetE",
407     idJets[0].p4.E(
408     ), hasOneJ, EqB(50, 0., 500.), title="Leading_Jet_Energy"))
409
410 plots.append(Plot.make1D(
411     "semiLeptonic_hasTwoJ_jetpt",
412     idJets[1].pt, hasTwoJ, EqB(
413     30, 0., 300.), title="Sub-leading_Jet_p_T"))
414 plots.append(Plot.make1D(
415     "semiLeptonic_hasTwoJ_jeteta",
416     idJets[1].eta, hasTwoJ, EqB(
417     80, -5.5, 5.5), title="Sub-leading_Jet_#eta"))
418 plots.append(Plot.make1D(
419     "semiLeptonic_hasTwoJ_jetphi",
420     idJets[1].phi, hasTwoJ, EqB(
421     100, -3.5, 3.5), title="Sub-leading_Jet_#phi"))
422 plots.append(Plot.make1D(
423     "semiLeptonic_hasTwoJ_jetE",
424     idJets[1].p4.E(
425     ), hasTwoJ, EqB(50, 0., 500.), title="Sub-leading_Jet_Energy"))
426
427 plots.append(Plot.make1D(
428     "semiLeptonic_hastwoJ_mjj", op.invariant_mass(
429     idJets[0].p4, idJets[1].p4), hasTwoJ,
430     EqB(100, 0., 1000.), title="Di-jet_invariant_mass"))
431 plots.append(Plot.make1D(
432     "semiLeptonic_hasthreeJ_mjj", op.invariant_mass(
433     idJets[1].p4, idJets[2].p4), hasthreeJ,
434     EqB(100, 0., 1000.), title="Tri-jet_invariant_mass"))
435
436 c3 = hasInvM.refine("hasOneTauNoLept", cut=op.AND(
437     nTau == 1, op.rng_len(idElectrons) == 0,
438     op.rng_len(isoMuons) == 0)
439 )
440 c3_OneJ = c3.refine("c3hasOneJ", cut=op.rng_len(idJets) == 1)
441 c3_TwoJ = c3.refine("c3hasTwoJ", cut=op.rng_len(idJets) == 2)
442
443 plots.append(Plot.make1D("c3_nJet", nJet, c3, EqB(
444     10, 0., 10.), title="Number_of_Jets"))
445 plots.append(Plot.make1D("c3_nbJet", op.rng_len(bJets), c3,
446     EqB(10, 0., 10.), title="Number_of_b-tagged_Jets"))
447 plots.append(Plot.make1D(
448     "c3_met", met[0].pt, c3, EqB(
449     80, 0., 800.), title="Missing_Transverse_Energy"))
450 plots.append(Plot.make1D(
451     "c3_pt_mgg", pT_mGGL, c3, EqB(
452     100, 0., 5.),
453     title="Leading_Photon_p_{T}/m_{\gamma\gamma}"))
454 plots.append(Plot.make1D(
455     "c3_SLpt_mgg", pT_mGGSL, c3, EqB(

```



```

456         100, 0., 5.),
457         title="Sub-leading_Photon_p_{T}/m_{\gamma\gamma}")
458 plots.append(Plot.make1D(
459     "c3_leadingphoton_eta", idPhotons[0].eta, c3, EqB(
460     80, -4., 4.), title="Leading_Photon_\eta"))
461 plots.append(Plot.make1D(
462     "c3_subleadingphoton_eta", idPhotons[1].eta, c3, EqB(
463     80, -4., 4.), title="Sub-leading_Photon_\eta"))
464 plots.append(Plot.make1D(
465     "c3_leadingphoton_phi", idPhotons[0].phi, c3, EqB(
466     100, -3.5, 3.5), title="Leading_Photon_\phi"))
467 plots.append(Plot.make1D(
468     "c3_subleadingphoton_phi", idPhotons[1].phi, c3, EqB(
469     100, -3.5, 3.5), title="Sub-leading_Photon_\phi"))
470 plots.append(Plot.make1D(
471     "c3_LE_mgg", E_mGGL, c3, EqB(
472     100, 0., 5.),
473     title="Leading_Photon_E/m_{\gamma\gamma}"))
474 plots.append(Plot.make1D(
475     "c3_SLE_mgg", E_mGGSL, c3, EqB(
476     100, 0., 5.),
477     title="Sub-leading_Photon_E/m_{\gamma\gamma}"))
478 plots.append(Plot.make1D(
479     "c3_leadingtau_E", isoTaus[0].p4.E(), c3, EqB(
480     50, 0., 500.), title="Leading_Tau_Energy"))
481 plots.append(Plot.make1D(
482     "c3_leadingtau_eta", isoTaus[0].eta, c3, EqB(
483     80, -4., 4.), title="Leading_Tau_\eta"))
484 plots.append(Plot.make1D(
485     "c3_leadingtau_phi", isoTaus[0].phi, c3, EqB(
486     100, -3.5, 3.5), title="Leading_Tau_\phi"))
487 plots.append(Plot.make1D(
488     "c3_leadingtau_pt", isoTaus[0].pt, c3, EqB(
489     50, 0., 500.), title="Leading_Tau_p_T"))
490 plots.append(Plot.make1D(
491     "c3_oneJet_Ljetpt", idJets[0].pt, c3_OneJ, EqB(
492     50, 0., 500.), title="Leading_Jet_p_T"))
493 plots.append(Plot.make1D(
494     "c3_oneJet_Ljeteta", idJets[0].eta, c3_OneJ, EqB(
495     80, -5.5, 5.5), title="Leading_Jet_\eta"))
496 plots.append(Plot.make1D(
497     "c3_twoJet_SLjetpt", idJets[1].pt, c3_TwoJ, EqB(
498     50, 0., 500.), title="Sub-leading_Jet_p_T"))
499 plots.append(Plot.make1D(
500     "c3_twoJet_SLjeteta", idJets[1].eta, c3_TwoJ, EqB(
501     80, -5.5, 5.5), title="Sub-leading_Jet_\eta"))
502
503 plots.append(Plot.make1D(
504     "fullyLeptonic_mgg", mGG, fullyLeptonic, EqB(
505     80, 100., 180.),
506     title="Di-photon_invariant_mass_[GeV]"))
507 plots.append(Plot.make1D(
508     "Inv_mass_ggfullyLeptonic", mGG, fullyLeptonic,
509     EqB(80, 100., 180.),
510     title="m_{\gamma\gamma}"))
511 plots.append(Plot.make1D(
512     "mGG_c3", mGG, c3, EqB(80, 100, 180),
513     title="M_{\gamma\gamma}",

```

```

514         plotopts={"log-y": True}))
515 plots.append(Plot.make1D(
516     "mGG_c4", mGG, c4, EqB(80, 100, 180),
517     title="M_{\gamma\gamma}",
518     plotopts={"log-y": True}))
519
520 mvaVars_semiLeptonic = {
521     "weight": noSel.weight,
522     "Eta_ph1": idPhotons[0].eta,
523     "Phi_ph1": idPhotons[0].phi,
524     "E_mGG_ph1": E_mGGL,
525     "pT_mGG_ph1": pT_mGGL,
526     "Eta_ph2": idPhotons[1].eta,
527     "Phi_ph2": idPhotons[1].phi,
528     "E_mGG_ph2": E_mGGSL,
529     "pT_mGG_ph2": pT_mGGSL,
530     "Electron_E": op.switch(
531         op.rng_len(idElectrons) == 0,
532         op.c_float(0.), idElectrons[0].p4.E()),
533     "Electron_pT": op.switch(
534         op.rng_len(idElectrons) == 0,
535         op.c_float(0.), idElectrons[0].pt),
536     "Electron_Eta": op.switch(
537         op.rng_len(idElectrons) == 0,
538         op.c_float(0.), idElectrons[0].eta),
539     "Electron_Phi": op.switch(
540         op.rng_len(idElectrons) == 0,
541         op.c_float(0.), idElectrons[0].phi),
542     "Muon_E": op.switch(
543         op.rng_len(isoMuons) == 0,
544         op.c_float(0.), isoMuons[0].p4.E()),
545     "Muon_pT": op.switch(
546         op.rng_len(isoMuons) == 0,
547         op.c_float(0.), isoMuons[0].pt),
548     "Muon_Eta": op.switch(
549         op.rng_len(isoMuons) == 0,
550         op.c_float(0.), isoMuons[0].eta),
551     "Muon_Phi": op.switch(
552         op.rng_len(isoMuons) == 0,
553         op.c_float(0.), isoMuons[0].phi),
554     "nJets": nJet,
555     "E_jet1": op.switch(
556         op.rng_len(idJets) == 0,
557         op.c_float(0.), idJets[0].p4.E()),
558     "pT_jet1": op.switch(
559         op.rng_len(idJets) == 0,
560         op.c_float(0.), idJets[0].pt),
561     "Eta_jet1": op.switch(
562         op.rng_len(idJets) == 0,
563         op.c_float(0.), idJets[0].eta),
564     "Phi_jet1": op.switch(
565         op.rng_len(idJets) == 0,
566         op.c_float(0.), idJets[0].phi),
567     "E_jet2": op.switch(
568         op.rng_len(idJets) < 2,
569         op.c_float(0.), idJets[1].p4.E()),
570     "pT_jet2": op.switch(
571         op.rng_len(idJets) < 2,

```

```

572         op.c_float(0.), idJets[1].pt),
573     "Eta_jet2": op.switch(
574         op.rng_len(idJets) < 2,
575         op.c_float(0.), idJets[1].eta),
576     "Phi_jet2": op.switch(
577         op.rng_len(idJets) < 2,
578         op.c_float(0.), idJets[1].phi),
579     "InvM_jet": op.switch(
580         op.rng_len(idJets) < 2,
581         op.c_float(0.), mJets),
582     "InvM_jet2": op.switch(
583         op.rng_len(idJets) < 3,
584         op.c_float(0.), mJets_SL),
585     "met": metPt
586 }
587
588 mvaVars_singleTau = {
589     "weight": noSel.weight,
590     # Event level variables
591     "nJets": nJet,
592     "nBJets": op.rng_len(bJets),
593     "metPt": metPt,
594     # Photon and di-Photon variables
595     "L_pt_mGG": pT_mGGL,
596     "L_photon_eta": idPhotons[0].eta,
597     "L_photon_phi": idPhotons[0].phi,
598     "E_mGG_ph1": E_mGGL,
599     "E_mGG_ph2": E_mGGSL,
600     "SL_pt_mGG": pT_mGGSL,
601     "SL_photon_eta": idPhotons[1].eta,
602     "SL_photon_phi": idPhotons[1].phi,
603     "LTauE": isoTaus[0].p4.E(),
604     "LtauPt": isoTaus[0].pt,
605     "LtauEta": isoTaus[0].eta,
606     "LtauPhi": isoTaus[0].phi,
607     "Ljet_Pt": op.switch(
608         nJet == 0, op.c_float(0.), idJets[0].pt),
609     "Ljet_Eta": op.switch(
610         nJet == 0, op.c_float(0.), idJets[0].eta),
611     "SLjet_Pt": op.switch(
612         nJet < 2, op.c_float(0.), idJets[1].pt),
613     "SLjet_Eta": op.switch(
614         nJet < 2, op.c_float(0.), idJets[1].eta),
615 }
616
617 # save mvaVars to be retrieved later in the
618 # postprocessor and save in a parquet file
619 if self.args.mvaSkim:
620     from bamboo.plots import Skim
621     plots.append(Skim("Skim", mvaVars_semiLeptonic,
622         semiLeptonic))
623     plots.append(Skim("c3", mvaVars_singleTau, c3))
624
625 # evaluate dnn model on data
626 if self.args.mvaEval:
627     #from IPython import embed
628     WW_DNNmodel_path_even =
629         "~/DNN_WW/even_model_test3.onnx"

```

```

630 WW_DNNmodel_path_odd =
631     "~/DNN_WW/odd_model_test3.onnx"
632 tt_DNNmodel_path_even =
633     "~/DNN_Tau/even_model_test3.onnx"
634 tt_DNNmodel_path_odd =
635     "~/DNN_Tau/odd_model_test3.onnx"
636
637 mvaVars_semiLeptonic.pop("weight", None)
638 mvaVars_singleTau.pop("weight", None)
639 mvaVars_semiLeptonic_FH.pop("weight", None)
640 from bamboo.root import loadHeader
641 loadHeader(
642     "~/WWGG/header_split.h")
643
644 split_evaluator = op.extMethod('split::Ph1_phi')
645 split = split_evaluator(idPhotons[0].phi)
646
647 if split == 0:
648     tt_model = tt_DNNmodel_path_even
649     WW_model = WW_DNNmodel_path_even
650 else:
651     tt_model = tt_DNNmodel_path_odd
652     WW_model = WW_DNNmodel_path_odd
653
654 dnn_ww = op.mvaEvaluator(
655     WW_model, mvaType="ONNXRuntime",
656     otherArgs="predictions")
657 inputs_ww = op.array(
658     'float',
659     *[op.static_cast('float', val)
660       for val in mvaVars_semiLeptonic.values()])
661 output_ww = dnn_ww(inputs_ww)
662
663 dnn_tt = op.mvaEvaluator(
664     tt_model, mvaType="ONNXRuntime", otherArgs="predictions")
665 inputs_tt = op.array(
666     'float',
667     *[op.static_cast('float', val)
668       for val in mvaVars_singleTau.values()])
669 output_tt = dnn_tt(inputs_tt)
670
671 ##### semiLeptonic_DNN #####
672
673 hasDNNscore1 = semiLeptonic.refine(
674     "hasDNNscore1", cut=op.in_range(0.1, output_ww[0], 0.6))
675 yields_OneL.add(hasDNNscore1, title='hasDNNscore1')
676
677 hasDNNscore2 = semiLeptonic.refine(
678     "hasDNNscore2", cut=op.in_range(0.6, output_ww[0], 0.8))
679 yields_OneL.add(hasDNNscore2, title='hasDNNscore2')
680
681 hasDNNscore3 = semiLeptonic.refine(
682     "hasDNNscore3", cut=op.in_range(0.8, output_ww[0], 0.92))
683 yields_OneL.add(hasDNNscore3, title='hasDNNscore3')
684
685 hasDNNscore4 = semiLeptonic.refine(
686     "hasDNNscore4", cut=output_ww[0] > 0.92)
687 yields_OneL.add(hasDNNscore4, title='hasDNNscore4')

```

```

688
689 ##### OneTau_DNN #####
690
691 hasDNNscore1_tt = c3.refine(
692     "hasDNNscore_tt", cut=op.in_range(0.1, output_tt[0],
693     0.65))
694 yields_OneTau.add(hasDNNscore1_tt, title='hasDNNscore_tt')
695
696 hasDNNscore2_tt = c3.refine(
697     "hasDNNscore2_tt", cut=output_tt[0] > 0.65)
698 yields_OneTau.add(hasDNNscore2_tt, title='hasDNNscore2_tt')
699
700 plots.append(Plot.make1D("dnn_score_ww",
701     output_ww[0], semiLeptonic, EqB(40, 0, 1.)))
702 plots.append(Plot.make1D("dnn_score_tt",
703     output_tt[0], c3, EqB(40, 0, 1.)))
704
705 plots.append(Plot.make1D("Inv_mass_semiLeptonic_DNN_1",
706     mGG, hasDNNscore1, EqB(
707     80, 100., 180.)), title="m_{\gamma\gamma}")
708 plots.append(Plot.make1D("Inv_mass_semiLeptonic_DNN_2",
709     mGG, hasDNNscore2, EqB(
710     80, 100., 180.)), title="m_{\gamma\gamma}")
711 plots.append(Plot.make1D("Inv_mass_semiLeptonic_DNN_3",
712     mGG, hasDNNscore3, EqB(
713     80, 100., 180.)), title="m_{\gamma\gamma}")
714 plots.append(Plot.make1D("Inv_mass_semiLeptonic_DNN_4",
715     mGG, hasDNNscore4, EqB(
716     80, 100., 180.)), title="m_{\gamma\gamma}")
717
718 plots.append(Plot.make1D("Inv_mass_singleTau_c3_DNN_1",
719     mGG, hasDNNscore1_tt, EqB(
720     80, 100., 180.)), title="m_{\gamma\gamma}")
721 plots.append(Plot.make1D("Inv_mass_singleTau_c3_DNN_2",
722     mGG, hasDNNscore2_tt, EqB(
723     80, 100., 180.)), title="m_{\gamma\gamma}")
724
725 from bamboo.plots import Skim
726
727 return plots

```

Listing A.2 : DNN setup for the $WW\gamma\gamma$ semi-leptonic final state

```

1 import os
2 import yaml
3 import importlib
4 import matplotlib.pyplot as plt
5 from matplotlib.backends.backend_pdf import PdfPages
6 import numpy as np
7 import pyarrow.parquet as pq
8 import pandas as pd
9 from sklearn.model_selection import train_test_split
10 from sklearn.preprocessing import LabelEncoder, OneHotEncoder
11 import tensorflow as tf
12 from tensorflow import keras
13 from tensorflow.keras import layers
14 from tensorflow.keras.callbacks import EarlyStopping,
15     ReduceLROnPlateau
16

```

```

17 ### Parameters of the training ###
18 #split = "even"
19 split = "odd"
20
21 suffix = 'test3'
22
23 quantile = 1.0
24
25 tags = ['HH', 'background', 'single_H']
26 importance = [1., 1., 1.] # Importance of each category
27
28 # DNN hyperparameters #
29 parameters = {
30     'epochs'           : 200,
31     'lr'               : 0.001,
32     'batch_size'       : 256,
33     'n_layers'         : 3,
34     'n_neurons'        : 64,
35     'hidden_activation': 'relu',
36     'output_activation': 'softmax',
37     'l2'               : 1e-6,
38     'dropout'          : 0.,
39     'batch_norm'       : True,
40 }
41
42 # Input variables
43 input_vars=["Eta_ph1",
44            "Phi_ph1",
45            "E_mGG_ph1",
46            "pT_mGG_ph1",
47            "Eta_ph2",
48            "Phi_ph2",
49            "E_mGG_ph2",
50            "pT_mGG_ph2",
51            "Electron_E",
52            "Electron_pT",
53            "Electron_Eta",
54            "Electron_Phi",
55            "Muon_E",
56            "Muon_pT",
57            "Muon_Eta",
58            "Muon_Phi",
59            "nJets",
60            "E_jet1",
61            "pT_jet1",
62            "Eta_jet1",
63            "Phi_jet1",
64            "E_jet2",
65            "pT_jet2",
66            "Eta_jet2",
67            "Phi_jet2",
68            "InvM_jet",
69            "InvM_jet2",
70            "met"]
71 # Load the required data #
72 outputPath = '~/bamboodev/WWGG/Clean_Skim'
73 skimFile = os.path.join(outputPath, 'results', 'Skim.parquet')
74 yamlFile = os.path.join(outputPath, 'plots.yml')

```

```

75
76 # Load dataframe from parquet #
77 df = pd.read_parquet(skimFile)
78
79 # Load samples + plots data from yaml file #
80 with open(yamlFile, 'r') as handle:
81     config = yaml.full_load(handle)
82
83 # Cut negative event weights #
84 df = df[df['weight']>0]
85
86 # Add tag column #
87 if 'tag' in df.columns:
88     del df['tag']
89
90 df['tag'] = 'background'
91 for singleH in ['VBFH', 'VH', 'THQ', 'GluGluHTo', 'ttHJet']:
92     df.loc[df.process.str.contains(singleH), ['tag']] = 'single_H'
93 df.loc[df.process.str.contains('HH'), ['tag']] = 'HH'
94
95 for tag in tags:
96     if tag in df.columns:
97         del df[tag]
98
99 assert len(set(tags).intersection(set(pd.unique(df['tag'])))) == \
100     len(tags)
101 label_encoder = LabelEncoder()
102 onehot_encoder = OneHotEncoder(sparse=False)
103 label_encoder.fit(tags)
104 integers = label_encoder.transform(df['tag']).reshape(-1, 1)
105 onehot_encoder.fit(np.arange(len(tags)).reshape(-1, 1))
106 onehot = onehot_encoder.transform(integers)
107 onehot_df = pd.DataFrame(onehot, columns=tags, index=df.index)
108
109 df = pd.concat((df, onehot_df), axis=1)
110
111 df['event_weight'] = df['weight'].copy()
112
113 if (df['event_weight'] < 0).sum() > 0:
114     raise RuntimeError(f"There_are_{(df['event_weight']<0).sum()} \
115     _____events_with_negative_event_weight,_this_should_not_happen")
116
117 # Compute training weight #
118 if 'training_weight' in df.columns:
119     del df['training_weight']
120 df['training_weight'] = df['event_weight'].copy()
121
122 for itag, tag in enumerate(tags):
123     df.loc[df['tag']==tag, 'training_weight'] *= importance[itag] \
124         * df.shape[0] / len(tags) / df[df['tag']==tag] /
125         ['event_weight'].sum()
126
127 df = df[(df['weight']>0) & (df['training_weight']<200)]
128
129 import utils
130 importlib.reload(utils)
131 print ('Using_event_weight')
132 utils.checkBatches(df, label_column='tag', weight_column=

```

```

133 'event_weight', batch_size=parameters['batch_size'])
134 print ('Using_training_weight')
135 utils.checkBatches(df, label_column='tag', /
136 weight_column='training_weight', \
137     batch_size=parameters['batch_size'])
138
139 # Plot the background and signal weights #
140 fig, axes = plt.subplots(figsize=(16, 8), nrows=len(tags), ncols=3)
141 fig.subplots_adjust(left=0.1, right=0.9, top=0.98, bottom=0.1, /
142 wspace=0.2, hspace=0.4)
143 for irow, tag in enumerate(tags):
144     for icol, column in enumerate(['weight', 'event_weight', /
145     'training_weight']):
146         axes[irow, icol].hist(df[df['tag']==tag][column], /
147         bins=100, color='b')
148         axes[irow, icol].set_title(f"Category_{tag}")
149         axes[irow, icol].set_xlabel(column)
150         axes[irow, icol].set_yscale('log')
151 fig.savefig("event_weights_A.pdf", dpi = 300)
152
153 # Determine splitting variable #
154 split_var = df['Phi_ph1'].copy()
155 split_var = np.abs(split_var)
156 split_var *= 1e4
157 split_var -= np.floor(split_var)
158 split_var = (split_var*1e1).astype(int)
159 split_var = split_var % 2 == 0
160 print (f'Even_set_has_{df[split_var].shape[0]:10d}_events_\
161 ____[{df[split_var].shape[0]/df.shape[0]*100:5.2f}%]')
162 print (f'Odd_set_has_{df[~split_var].shape[0]:10d}_events_\
163 ____[{df[~split_var].shape[0]/df.shape[0]*100:5.2f}%]')
164
165 # Sets splitting #
166 print (f'Using_split_type_{split}')
167 if split == 'even':
168     train_df = df[~split_var] # Trained on odd
169     test_df = df[split_var] # Evaluated on even
170 elif split == 'odd':
171     train_df = df[split_var] # Trained on even
172     test_df = df[~split_var] # Evaluated on odd
173 else:
174     raise RuntimeError(f'Split_needs_to_be_either_odd_or_even')
175
176 # Randomise for training
177 train_df = train_df.sample(frac=1)
178
179 # Quantile corrections
180 quantile_lim = train_df['training_weight'].quantile(quantile)
181 print (f'{(1-quantile)*100:5.2f}%_right_quantile_is_when/
182 weight_is_at_{quantile_lim}')
183 print ('->_These_events_will_be_repeated_and_their/
184 learning_weights_reduced_accordingly_to_avoid_unstability')
185
186 # Select the events #
187 idx_to_repeat = train_df['training_weight'] >= quantile_lim
188 events_excess = train_df[idx_to_repeat].copy()
189
190 saved_columns = train_df[['training_weight', 'process']].copy()

```



```

191
192 # Compute multiplicative factor #
193 factor = (events_excess['training_weight']/quantile_lim)/
194 .values.astype(np.int32)
195
196 # Correct the weights of events already in df #
197 train_df.loc[idx_to_repeat,'training_weight'] /= factor
198
199 # Add N-1 copies #
200 arr_to_repeat = train_df[idx_to_repeat].values
201 repetition = np.repeat(np.arange(arr_to_repeat.shape[0]), factor-1)
202 df_repeated = pd.DataFrame(np.take(arr_to_repeat,repetition,axis=0),
203 columns=train_df.columns)
204 df_repeated = df_repeated.astype(train_df.dtypes.to_dict())
205 train_df = pd.concat((train_df,df_repeated),axis=0,/
206 ignore_index=True).sample(frac=1).reset_index()
207
208 # Printout #
209 print ('Changes_per_process_in_training_set')
210 for process in pd.unique(train_df['process']):
211     N_before = saved_columns[saved_columns['process']==process]/
212     .shape[0]
213     N_after = train_df[train_df['process']==process].shape[0]
214     if N_before != N_after:
215         print (f"{process:20s}")
216         print (f"..._{N_before:6d}_events_{sum_weight=_\
217 _____{saved_columns[saved_columns['process']==process]\
218 _____['training_weight'].sum():14.6f}]",end='_->_')
219         print (f"{N_after:6d}_events_{sum_weight=_\
220 _____{train_df[train_df['process']==process]\
221 _____['training_weight'].sum():14.6f}]" )
222
223 print ()
224 print (f"Total_entries_{_}{saved_columns.shape[0]:14d}_->_\
225 ____{train_df.shape[0]:14d}")
226 print (f"Total_event_sum_{_}{saved_columns['training_weight']\
227 _____.sum():14.6f}_->_{train_df['training_weight'].sum():14.6f}")
228
229 # Validation split #
230 train_df,val_df = train_test_split(train_df,test_size=0.3)
231
232 # Printout #
233 print ('\nFinal_sets')
234 print (f'Training_set_____={train_df.shape[0]}')
235 print (f'Validation_set_____={val_df.shape[0]}')
236 print (f'Testing_set_____={test_df.shape[0]}')
237 print (f'Total_set_____={df.shape[0]}')
238
239 # Plot the background and signal weights #
240 fig,axs = plt.subplots(figsize=(16,8),nrows=1,ncols=2)
241 fig.subplots_adjust(left=0.1, right=0.9, top=0.96, bottom=0.1,
242                     wspace=0.2,hspace=0.3)
243
244 if split == 'even':
245     axs[0].hist(df[~split_var]['training_weight'],bins=100,color='b')
246 elif split == 'odd':
247     axs[0].hist(df[split_var]['training_weight'],bins=100,color='b')
248 axs[0].set_title("Before_correction")

```

```

249 axs[0].set_xlabel("Training_weight")
250 axs[0].set_yscale('log')
251 axs[1].hist(train_df['training_weight'],bins=100,color='b')
252 axs[1].set_title("After_correction")
253 axs[1].set_xlabel("Training_weight")
254 axs[1].set_yscale('log')
255 fig.savefig("event_weights_C.pdf", dpi = 300)
256
257 # Input layer #
258 inputs = keras.Input(shape=(len(input_vars),), name="particles")
259
260 # Preprocessing layer
261 from tensorflow.keras.layers.experimental import preprocessing
262 normalizer = preprocessing.Normalization\
263     (mean      = train_df[input_vars].mean(axis=0),
264      variance  = train_df[input_vars].var(axis=0),
265      name      = 'Normalization')(inputs)
266     # this layer does the preprocessing (x-mu)/std for each input
267 # Dense (hidden) layers #
268 x = normalizer
269 for i in range(parameters['n_layers']):
270     x = layers.Dense(
271         units          = parameters['n_neurons'],
272         activation      = parameters['hidden_activation'],
273         activity_regularizer = tf.keras.regularizers/
274             .l2(parameters['l2']),
275         name           = f"dense_{i}") (x)
276     if parameters['batch_norm']:
277         x = layers.BatchNormalization()(x)
278     if parameters['dropout'] > 0.:
279         x = layers.Dropout(parameters['dropout'])(x)
280 # Output layer #
281 outputs = layers.Dense(
282     units          = 3,
283     activation      = parameters['output_activation'],
284     activity_regularizer = tf.keras.regularizers/
285         .l2(parameters['l2']),
286     name           = "predictions") (x)
287
288 # Registering the model #
289 model = keras.Model(inputs=inputs, outputs=outputs)
290
291 model_preprocess = keras.Model(inputs=inputs, outputs=normalizer)
292 out_test = model_preprocess.predict(train_df[input_vars]/
293     ,batch_size=5000)
294 print ('Input_(after_normalization)_mean_(should/
295 be_close_to_0)')
296 print(out_test.mean(axis=0))
297 print ('Input_(after_normalization)_variance_(should/
298 be_close_to_1)')
299 print(out_test.var(axis=0))
300
301 model.compile(
302     #optimizer=keras.optimizers.RMSprop(),
303     optimizer=keras.optimizers.Adam(lr=parameters['lr']),
304     # Loss function to minimize
305     loss=keras.losses.CategoricalCrossentropy(),
306     # List of metrics to monitor

```

```

307     metrics=[keras.metrics.BinaryAccuracy(),
308               tf.keras.metrics.AUC(),
309               tf.keras.metrics.Precision(),
310               tf.keras.metrics.Recall()],
311 )
312
313 model.summary()
314
315 # Callbacks #
316 early_stopping = EarlyStopping(monitor = 'val_loss',
317                                min_delta = 0.001,
318                                patience = 20,
319                                verbose=1,
320                                mode='min',
321                                restore_best_weights=True)
322 # Stop the learning when val_loss stops increasing
323 # https://keras.io/api/callbacks/early_stopping/
324
325 reduce_plateau = ReduceLROnPlateau(monitor = 'val_loss',
326                                     factor = 0.1,
327                                     min_delta = 0.001,
328                                     patience = 8,
329                                     min_lr = 1e-8,
330                                     verbose=2,
331                                     mode='min')
332 # reduce LR if not improvement for some time
333 # https://keras.io/api/callbacks/reduce_lr_on_plateau/
334 import History
335 importlib.reload(History)
336 loss_history = History.LossHistory()
337
338 history = model.fit(
339     train_df[input_vars],
340     train_df[tags],
341     verbose=2,
342     batch_size=parameters['batch_size'],
343     epochs=parameters['epochs'],
344     sample_weight=train_df['training_weight'],
345     # Pass some validation for
346     # monitoring validation loss and metrics
347     # at the end of each epoch
348     validation_data=(val_df[input_vars],val_df[tags],
349                     val_df['training_weight']),
350     callbacks = [early_stopping, reduce_plateau, loss_history],
351 )
352
353 History.PlotHistory(loss_history,params=parameters,
354                     outputName=f'loss_{suffix}_{split}.png')
355 # Params is a dict of parameters with name and values
356 # used for plotting
357
358 # Produce output on the test set as new column #
359 output = model.predict(test_df[input_vars],batch_size=5000)
360 output_tags = [f'output_{tag}' for tag in tags]
361 for output_tag in output_tags:
362     if output_tag in test_df.columns:
363         del test_df[output_tag]
364

```

```

365 test_df = pd.concat((test_df, pd.DataFrame(
366     output, columns=output_tags, index=test_df.index)), axis=1)
367
368 # Make the discriminator #
369 if 'd_HH' in test_df.columns:
370     del test_df['d_HH']
371
372 signal_idx = [i for i, tag in enumerate(tags) if 'HH' in tag]
373
374 # d_HH = ln (P(HH) / (P(single H) + P(background)))
375
376 #test_df['d_HH'] = pd.Series(np.ones(test_df.shape[0]))
377
378 # Numerator #
379 num = pd.DataFrame((test_df[[output_tags[i]\
380     for i in range(len(tags)) if i in signal_idx]]).sum(axis=1))
381 # Denominator #
382 den = pd.DataFrame(test_df[[output_tags[i]\
383     for i in range(len(tags)) if i not in signal_idx]].sum(axis=1))
384 # Ln #
385 d_HH = np.log(num / den)
386 test_df['d_HH'] = d_HH
387
388 if df[df.isin([np.nan, np.inf, -np.inf]).any(1)].shape[0] > 0:
389     raise RuntimeError('Some_nan_of_inf_values_in_d_HH')
390
391 import roc
392 importlib.reload(roc) # Reload in case file has changed
393 for tag in tags:
394     print (f'ROC_curve_of_binary_classification_of_{tag}\
395     _____node_versus_all_the_others')
396     roc.rocAndSig(y_true = test_df[tag],
397                  y_pred = test_df[f'output_{tag}'],
398                  w_roc = test_df['training_weight'],
399                  w_sig = test_df['event_weight'],
400                  show_significance = 'HH' in tag,
401                  outputName = f'roc_{suffix}_{split}_{tag}.pdf')
402
403 print (f'ROC_curve_of_binary_classification_of_d_HH')
404 roc.rocAndSig(y_true = test_df['HH'],
405              y_pred = test_df['d_HH'],
406              w_roc = test_df['training_weight'],
407              w_sig = test_df['event_weight'],
408              show_significance = True,
409              outputName = f'roc_{suffix}_{split}_d_HH.pdf')
410
411 for tag in tags:
412     print (f'Multi_roc_curve_for_{tag}')
413     tags_order = [tag] + [t for t in tags if t != tag]
414     roc.multiRoc(
415         outputs = [test_df[test_df['tag']==tag][f'output_{tag}'] for tag in tags_order],
416         tags = tags_order,
417         weights = [test_df[test_df['tag']==tag][f'training_weight'] for tag in tags_order],
418         title = f'Using_node_{tag}',
419         outputName = f'multi_roc_{suffix}_{split}_output_{tag}.pdf')
420
421
422

```

```

423 fig, axes = plt.subplots(figsize=(12,25),nrows=len(tags)+1,ncols=2)
424 fig.subplots_adjust(left=0.1, right=0.9, top=0.98,
425                     bottom=0.1, wspace=0.3,hspace=0.5)
426
427 tag_df = {tag:test_df[test_df['tag']==tag] for tag in tags}
428 colors = ['g','r','b']
429
430 # Manual binning to help compute significance #
431 n_bins = 50
432
433 def get_bin_content(bins,y,w):
434     digitized = np.digitize(y,bins)
435     return np.array([w[digitized==i].sum() for i in range(1,/
436                     len(bins))])
437
438 for irow,output_tag in enumerate(output_tags+['d_HH']):
439     for icol,weight in enumerate(['event_weight','training_weight']):
440         # Fill the bins myself #
441         bins = np.linspace(test_df[output_tag].min(),
442                             test_df[output_tag].max(),n_bins+1)
443         centers = (bins[1:]+bins[:-1])/2
444         widths = np.diff(bins)
445
446         tag_content = {tag:get_bin_content(bins,tag_df[tag]/
447                                         [output_tag],
448                                         tag_df[tag][weight])for tag in tags}
449         tag_cumsum_left = {tag:np.cumsum(tag_content[tag])/\
450                             tag_content[tag].sum() for tag in tags}
451         tag_cumsum_right = {tag:np.cumsum(tag_content[tag]\
452                                         [::-1])[::-1]/tag_content[tag].sum() for tag in tags}
453         for i,(tag,content) in enumerate(tag_content.items()):
454             axes[irow,icol].bar(x=centers,height=content,
455                                width=widths,fill=False,edgecolor=colors[i],label=tag)
456         axes[irow,icol].set_title(f"Using_{weight}")
457         axes[irow,icol].set_xlabel(output_tag)
458         axes[irow,icol].set_ylabel('Yield')
459         axes[irow,icol].set_ylim(1e-5,max([content.max() for\
460                                         content in tag_content.values()])*100)
461         axes[irow,icol].set_yscale('log')
462         axes[irow,icol].legend(loc='upper_left')
463 fig.savefig(f"prediction_{suffix}_{split}.pdf", dpi = 300)
464
465 # evaluate the model
466 scores = model.evaluate(test_df[input_vars],
467                         test_df[tags],
468                         sample_weight = test_df['training_weight'],
469                         batch_size = 5000,
470                         verbose=2)
471 print("%s: %.2f%%" % (model.metrics_names[1], scores[1]*100))
472
473 # save model and architecture to single file
474 modelName = f"model_{suffix}_{split}"
475 model.save(modelName)
476 print(f"Saved_model_to_disk_as_{modelName}")

```

APPENDIX A.2

Table A.1 : MC samples used in the analysis.

Dataset	Nevents	X-section \times BR (fb)
Signal		
<i>GluGluToHHHTo2G2Tau_node_cHHH1_TuneCP5_14TeV-powheg-pythia8_200PU</i>	1999866	0.00104441
<i>GluGluToHHHTo2G2Qnu_node_cHHH1_TuneCP5_14TeV-powheg-pythia8_200PU</i>	1898894	0.0156981
<i>GluGluToHHHTo2G2l2nu_node_cHHH1_TuneCP5_14TeV-powheg-pythia8_200PU</i>	1885835	0.0037234
Resonant Background		
<i>VHToGG_M125_TuneCP5_14TeV-amcatnloFFX-madspin-pythia8_200PU</i>	1830426	5.44326
<i>ttHJetToGG_M125_TuneCP5_14TeV-amcatnloFFX-madspin-pythia8_200PU</i>	5971707	1.393764
<i>GluGluHToGG_M125_TuneCP5_14TeV-amcatnloFFX-pythia8_200PU</i>	444658	114.798
<i>VBFHToGG_M125_TuneCP5_14TeV-amcatnlo-pythia8_200PU</i>	1712900	9.51216
<i>THQ_ctcvcp_HToGG_M125_TuneCUETP8M1_14TeV-madgraph-pythia8_200PU</i>	789129.0	0.205428
Continuum Background		
<i>DiPhotonJetsBox_MGG-40to80_14TeV-Sherpa_200PU</i>	20677034.0	332804
<i>DiPhotonJetsBox_MGG-80toInf_14TeV-Sherpa_200PU</i>	19933297	98670
<i>GJet_Pt-20toInf_DoubleEMEnriched_MGG-40to80_TuneCUETP8M2T4_14TeV-pythia8_200PU</i>	19985496	3901000
<i>GJet_Pt-40toInf_DoubleEMEnriched_MGG-80toInf_TuneCUETP8M2T4_14TeV_Pythia8_200PU</i>	20033932	998100
<i>GJet_Pt-20to40_DoubleEMEnriched_MGG-80toInf_TuneCP5_14TeV-pythia8_200PU</i>	14313734	260850
<i>QCD_Pt-30toInf_DoubleEMEnriched_MGG-40to80_TuneCUETP8M2T4_14TeV_Pythia8_200PU</i>	19581853	295700000
<i>QCD_Pt-40toInf_DoubleEMEnriched_MGG-80toInf_TuneCUETP8M2T4_14TeV_Pythia8_200PU</i>	7141716	141200000
<i>QCD_Pt-30to40_DoubleEMEnriched_MGG-80toInf_TuneCUETP8M2T4_14TeV_Pythia8_200PU</i>	30793791	16510000
<i>TT_TuneCUETP8M2T4_14TeV-powheg-pythia8_200PU</i>	49398942	864400
<i>TTWJetsToLNu_TuneCUETP8M1_14TeV-amcatnloFFX-madspin-pythia8_200PU</i>	5040836.0	225.3
<i>TTGamma_Dilept_TuneCUETP8M2T4_14TeV-madgraph-pythia8_200PU</i>	2999843	623.1
<i>TTGamma_Hadronic_TuneCUETP8M2T4_14TeV-madgraph-pythia8_200PU</i>	2999836	799

Table A.2 : MC samples used in the analysis (cont'd).

Continuum Background			
<i>TTGamma_SingleLeptFromT_TuneCUETP8M2T4_14TeV-madgraph-pythia8_200PU</i>	2939839	770.9	
<i>TTGamma_SingleLeptFromTbar_TuneCUETP8M2T4_14TeV-madgraph-pythia8_200PU</i>	2939844	769	
<i>TTGG_0Jets_TuneCUETP8M1_14TeV_amcatnlo_madspin_pythia8_200PU</i>	1101895	18.64	
<i>DYJets_incl_MLL-50_TuneCUETP8M1_14TeV-madgraphMLM-pythia8_200PU</i>	76952612.0	5711000	
<i>W1JetsToLNu_TuneCUETP8M1_14TeV-madgraphMLM-pythia8_200PU</i>	77486992.0	10370000	
<i>W2JetsToLNu_TuneCUETP8M1_14TeV-madgraphMLM-pythia8_200PU</i>	43222285.0	2965000	
<i>W3JetsToLNu_TuneCUETP8M1_14TeV-madgraphMLM-pythia8_200PU</i>	5674591.0	1268000	
<i>WGToLNuG_PtG-40_TuneCUETP8M1_14TeV-madgraphMLM-pythia8_200PU</i>	11776400	18790	
<i>ZGTo2LG_TuneCUETP8M1_14TeV-amcatnloFFX-pythia8_200PU</i>	30301987	145200	
<i>WGGJets_TuneCP5_14TeV_madgraphMLM_pythia8_200PU</i>	1981569.0	1884	
<i>WGJJToLNu_FWK_QCD_TuneCP5_14TeV-madgraph-pythia8_200PU</i>	1801596.0	6032	
<i>WW_TuneCUETP8M1_14TeV-pythia8_200PU</i>	99484471.0	70440	

Table A.3 : Cut-flow report showing number of events, before selections, in the semi-leptonic channel and in its categories. Percentages in brackets show the total selection efficiency that is number of events of the final state divided by the total number of events in the samples $\times 100$.

Samples	No sel.	Semi-leptonic F.S.	DNN Cat. 1	DNN Cat. 2	DNN Cat. 3	DNN Cat. 4
$HH \rightarrow WW\gamma\gamma$	$4.69e+01$	$6.40e+00$ (13.6%)	$1.98e+00$ (4.2%)	$1.35e+00$ (2.8%)	$1.62e+00$ (3.4%)	$1.16e+00$ (2.4%)
$HH \rightarrow WW\gamma\gamma(FL)$	$1.12e+01$	$2.35e+00$ (21.0%)	$6.52e-01$ (5.8%)	$4.94e-01$ (4.4%)	$6.40e-01$ (5.7%)	$4.41e-01$ (3.9%)
$HH \rightarrow \tau\tau\gamma\gamma$	$3.13e+00$	$3.44e-01$ (10.9%)	$9.46e-02$ (3.0%)	$7.05e-02$ (2.2%)	$9.09e-02$ (2.9%)	$6.98e-02$ (2.2%)
Signal	$1.10e+02$	$9.39e+00$ (8.5%)	$2.88e+00$ (2.6%)	$1.98e+00$ (1.8%)	$2.39e+00$ (2.1%)	$1.69e+00$ (1.5%)
$GGH \rightarrow \gamma\gamma$	$3.44e+05$	$8.94e+02$ (0.2%)	$2.64e+02$ (0.0%)	$2.22e+01$ (0.0%)	$9.73e-01$ (0.0%)	$2.78e-01$ (0.0%)
$VBFH \rightarrow \gamma\gamma$	$2.85e+04$	$1.11e+02$ (0.3%)	$5.86e+01$ (0.2%)	$7.96e+00$ (0.0%)	$4.57e+00$ (0.0%)	$5.92e-01$ (0.0%)
$ttH \rightarrow \gamma\gamma$	$4.18e+03$	$3.97e+02$ (9.4%)	$2.19e+02$ (5.2%)	$3.45e+01$ (0.8%)	$1.81e+01$ (0.4%)	$5.99e+00$ (0.1%)
$VH \rightarrow \gamma\gamma$	$1.63e+04$	$5.91e+02$ (3.6%)	$2.77e+02$ (1.6%)	$7.33e+01$ (0.4%)	$4.61e+01$ (0.2%)	$1.58e+01$ (0.0%)
THQ	$6.16e+02$	$3.68e+01$ (5.9%)	$1.86e+01$ (3.0%)	$6.52e+00$ (1.0%)	$5.09e+00$ (0.8%)	$2.08e+00$ (0.3%)
$\gamma\gamma + jets_{80} - Inf$	$2.96e+08$	$1.72e+05$ (0.0%)	$5.57e+04$ (0.0%)	$4.42e+03$ (0.0%)	$1.38e+03$ (0.0%)	$2.80e+02$ (0.0%)
$\gamma\gamma + jets_{40} - 80$	$9.98e+08$	$4.96e+03$ (0.0%)	$1.02e+03$ (0.0%)	$6.36e+01$ (0.0%)	$2.82e+01$ (0.0%)	$7.06e+00$ (0.0%)
$G + jets$	$2.99e+09$	$2.82e+05$ (0.0%)	$6.74e+04$ (0.0%)	$3.89e+03$ (0.0%)	$1.49e+03$ (0.0%)	$4.48e+02$ (0.0%)
$G + jets_{20} - 40GeV$	$7.83e+08$	$1.91e+04$ (0.0%)	$1.20e+03$ (0.0%)	$0.00e+00$ (0.0%)	$5.47e+01$ (0.0%)	$0.00e+00$ (0.0%)
$G + jets_{20} - Inf$	$1.17e+10$	$5.91e+04$ (0.0%)	$1.05e+04$ (0.0%)	$0.00e+00$ (0.0%)	$1.17e+03$ (0.0%)	$0.00e+00$ (0.0%)
$W1Jets \rightarrow Lv$	$3.11e+10$	$3.25e+04$ (0.0%)	$2.81e+03$ (0.0%)	$4.02e+02$ (0.0%)	$0.00e+00$ (0.0%)	$0.00e+00$ (0.0%)
$W2Jets \rightarrow Lv$	$8.90e+09$	$2.70e+04$ (0.0%)	$7.42e+03$ (0.0%)	$0.00e+00$ (0.0%)	$2.06e+02$ (0.0%)	$0.00e+00$ (0.0%)
$W3Jets \rightarrow Lv$	$3.80e+09$	$3.36e+04$ (0.0%)	$1.34e+04$ (0.0%)	$2.01e+03$ (0.0%)	$6.72e+02$ (0.0%)	$0.00e+00$ (0.0%)
$WGJJ$	$1.81e+07$	$7.09e+03$ (0.0%)	$3.09e+03$ (0.0%)	$2.71e+02$ (0.0%)	$1.10e+02$ (0.0%)	$3.01e+01$ (0.0%)
$WGGJets$	$5.65e+06$	$7.65e+03$ (0.1%)	$3.30e+03$ (0.0%)	$3.77e+02$ (0.0%)	$1.68e+02$ (0.0%)	$5.13e+01$ (0.0%)
$DYJets$	$1.71e+10$	$1.87e+05$ (0.0%)	$3.74e+04$ (0.0%)	$1.78e+03$ (0.0%)	$2.23e+02$ (0.0%)	$0.00e+00$ (0.0%)

Table A.4 : Cut-flow report showing number of events, before selections, in the semi-leptonic channel and in its categories. Percentages in brackets show the total selection efficiency that is number of events of the final state divided by the total number of events in the samples $\times 100$. (cont'd).

Samples	No sel.	Semi-leptonic F.S.	DNN Cat. 1	DNN Cat. 2	DNN Cat. 3	DNN Cat. 4
<i>ZG</i>	$4.36e+08$	$9.73e+04$ (0.0%)	$3.27e+04$ (0.0%)	$2.10e+03$ (0.0%)	$9.39e+02$ (0.0%)	$1.13e+02$ (0.0%)
<i>WW</i> (inclusive)	$2.11e+08$	$2.04e+03$ (0.0%)	$7.97e+02$ (0.0%)	$9.35e+01$ (0.0%)	$5.74e+01$ (0.0%)	$8.50e+00$ (0.0%)
<i>t\bar{t}</i> (inclusive)	$2.59e+09$	$5.43e+04$ (0.0%)	$2.37e+04$ (0.0%)	$1.52e+03$ (0.0%)	$6.83e+02$ (0.0%)	$1.58e+02$ (0.0%)
<i>ttGJets</i>	$1.37e+07$	$1.36e+04$ (0.0%)	$6.51e+03$ (0.0%)	$8.85e+02$ (0.0%)	$4.03e+02$ (0.0%)	$7.83e+01$ (0.0%)
<i>ttGG</i>	$5.59e+04$	$6.60e+02$ (1.1%)	$3.18e+02$ (0.5%)	$3.35e+01$ (0.0%)	$1.85e+01$ (0.0%)	$7.43e+00$ (0.0%)
<i>ttW</i>	$6.76e+05$	$9.66e+01$ (0.0%)	$3.99e+01$ (0.0%)	$4.04e+00$ (0.0%)	$1.26e+00$ (0.0%)	$1.01e+00$ (0.0%)
Background	$8.10e+10$	$1.00e+06$ (0.0%)	$2.68e+05$ (0.0%)	$1.80e+04$ (0.0%)	$7.67e+03$ (0.0%)	$1.20e+03$ (0.0%)

APPENDIX A.3

Table A.5 : Input variables used to train semi-leptonic final state DNN.

Feature	Description
Leading Photon $p_T / m_{\gamma\gamma}$	p_T of the leading good photon scaled to diphoton mass.
Leading Photon Energy / $m_{\gamma\gamma}$	Energy of the leading good photon scaled to diphoton mass.
Leading Photon η	Pseudorapidity of the leading good photon
Leading Photon ϕ	Direction in the transverse plane of the leading good photon
Sub-leading Photon $p_T / m_{\gamma\gamma}$	p_T of the sub-leading good photon scaled to diphoton mass.
Sub-leading Photon Energy / $m_{\gamma\gamma}$	Energy of the sub-leading good photon scaled to diphoton mass.
Sub-leading Photon η	Pseudorapidity of the sub-leading good photon
Sub-leading Photon ϕ	Direction in the transverse plane of the sub-leading good photon
Electron p_T	p_T of the selected electron
Electron η	Pseudorapidity of the selected electron
Electron ϕ	Direction in the transverse plane of the selected electron
Electron Energy	Energy of the selected electron
Muon p_T	p_T of the selected muon
Muon η	Pseudorapidity of the selected muon
Muon ϕ	Direction in the transverse plane of the selected muon
Muon Energy	Energy of the selected muon
Jet Multiplicity	Number of jets in the event (flavour inclusive)
MET	Missing transverse energy in the event
Leading Jet p_T	p_T of the leading good jet
Leading Jet η	Pseudorapidity of the leading good jet
Leading Jet ϕ	Direction in the transverse plane of the leading good jet
Leading Jet Energy	Energy of the leading good jet
Sub-leading Jet p_T	p_T of the sub-leading good jet
Sub-leading Jet η	Pseudorapidity of the sub-leading good jet
Sub-leading Jet ϕ	Direction in the transverse plane of the sub-leading good jet
Sub-leading Jet Energy	Energy of the sub-leading good jet
m_{j_0,j_1}	The invariant mass of the leading and sub-leading jets
m_{j_1,j_2}	The invariant mass of the leading and sub-leading jets

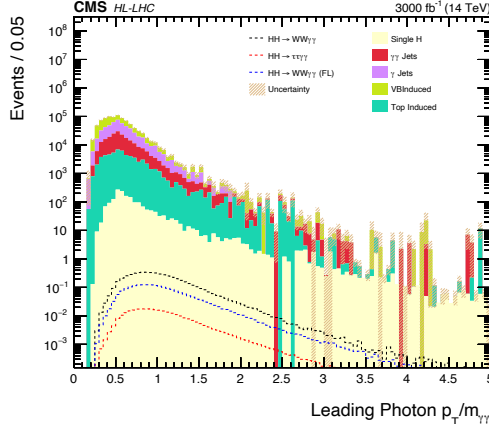


Figure A.1a : Leading Photon $p_T/m_{\gamma\gamma}$

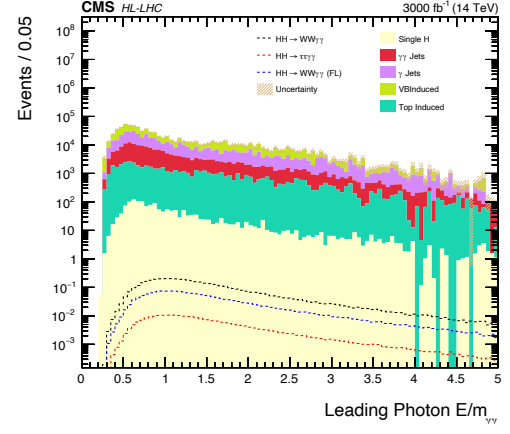


Figure A.1b : Leading Photon $E/m_{\gamma\gamma}$

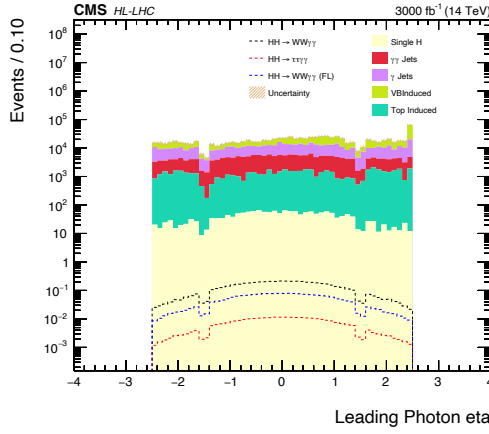


Figure A.1c : Leading Photon η

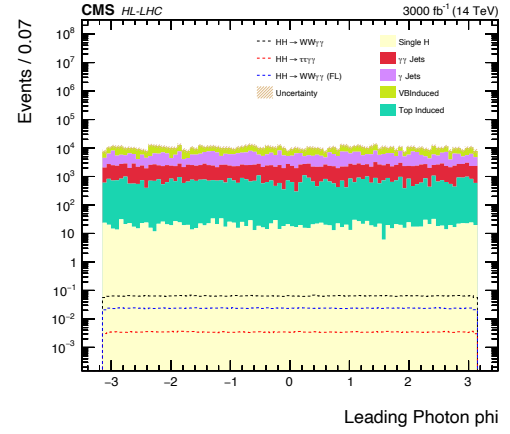


Figure A.1d : Leading Photon ϕ

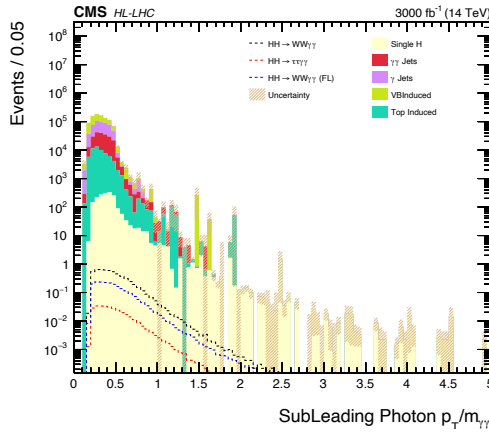


Figure A.1e : Sub-leading Photon
 $p_T/m_{\gamma\gamma}$

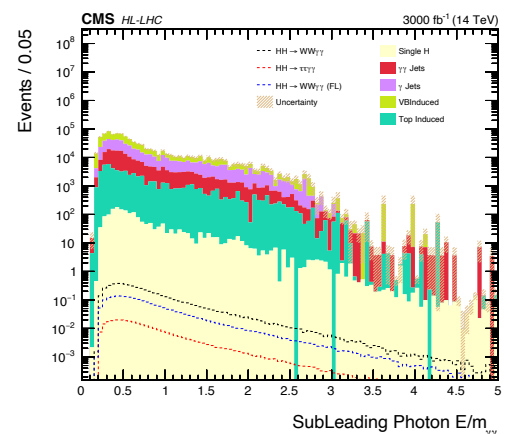


Figure A.1f : Sub-leading Photon $E/m_{\gamma\gamma}$

Figure A.1 : DNN input distributions for the semi-leptonic final state.

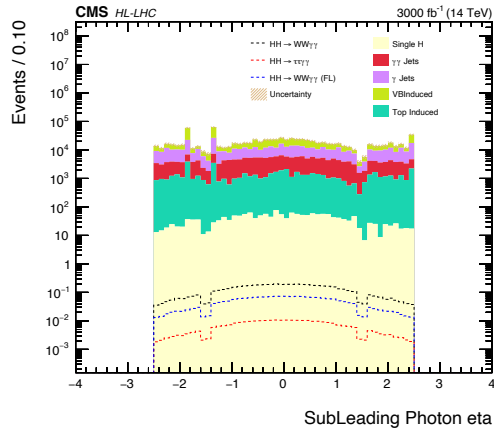


Figure A.2a : Sub-leading Photon η

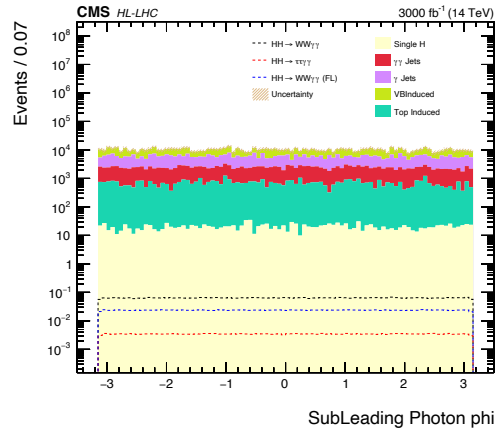


Figure A.2b : Sub-leading Photon ϕ

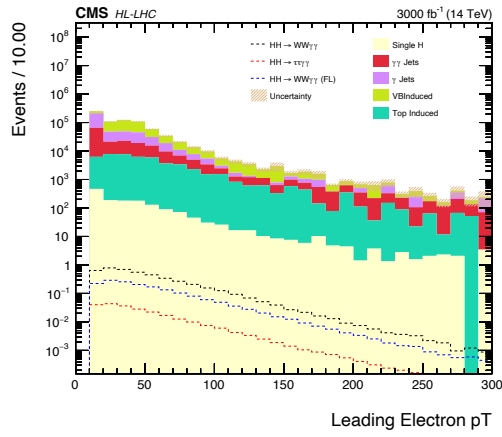


Figure A.2c : Leading Electron p_T

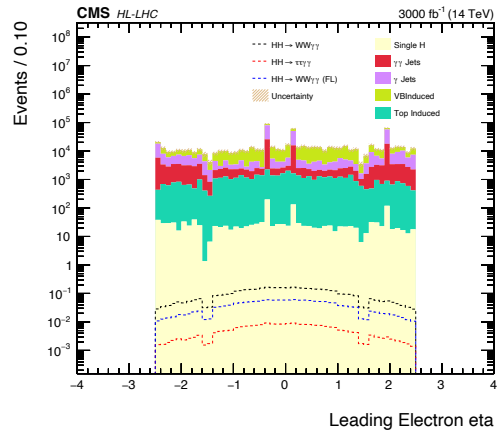


Figure A.2d : Leading Electron η

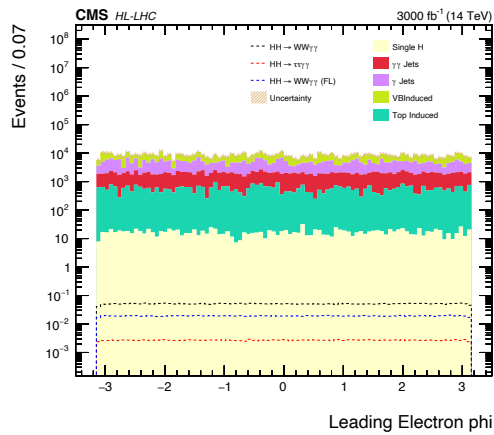


Figure A.2e : Leading Electron ϕ

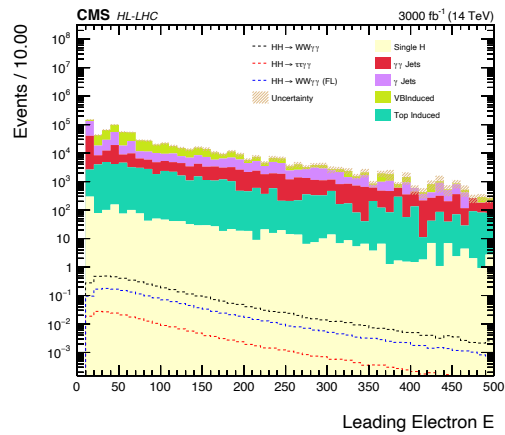


Figure A.2f : Leading Electron energy

Figure A.2 : DNN input distributions for the semi-leptonic final state (continued).

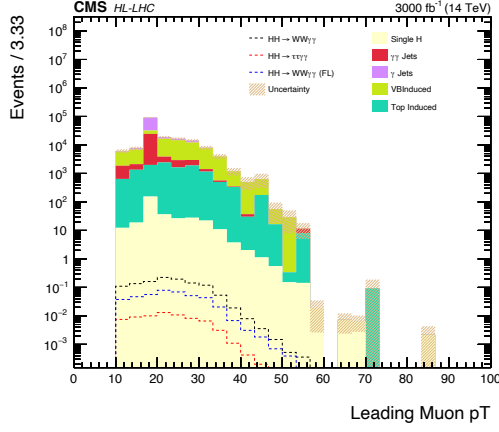


Figure A.3a : Leading Muon p_T

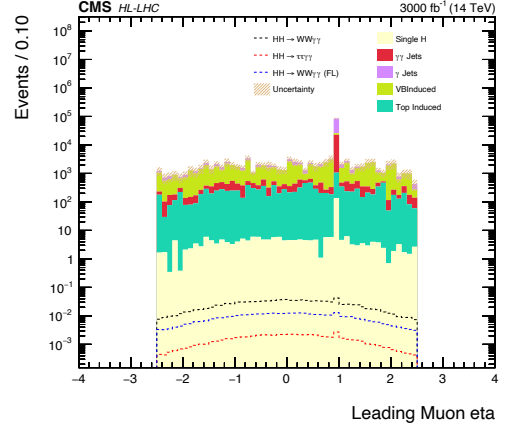


Figure A.3b : Leading Muon η

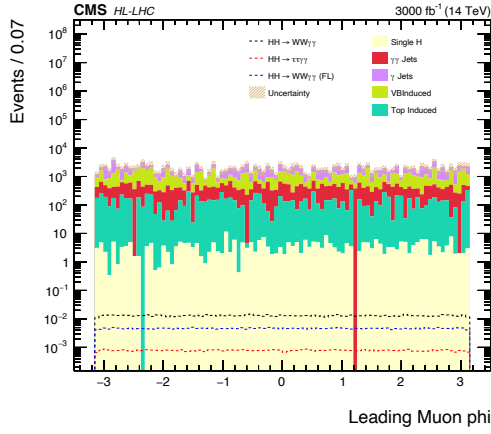


Figure A.3c : Leading Muon ϕ

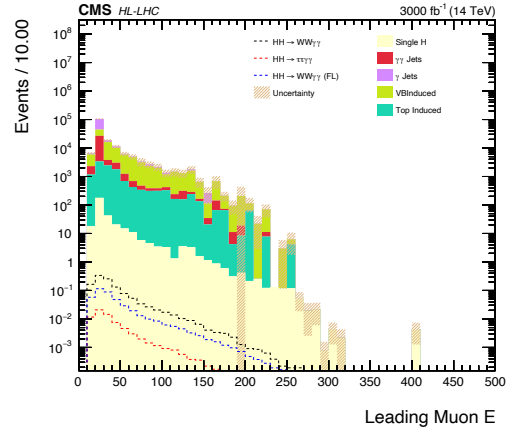


Figure A.3d : Leading Muon energy

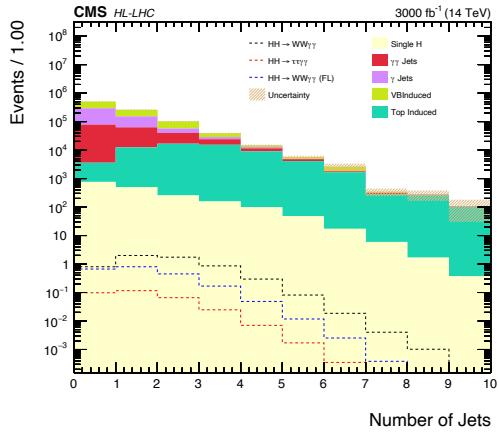


Figure A.3e : Jet multiplicity

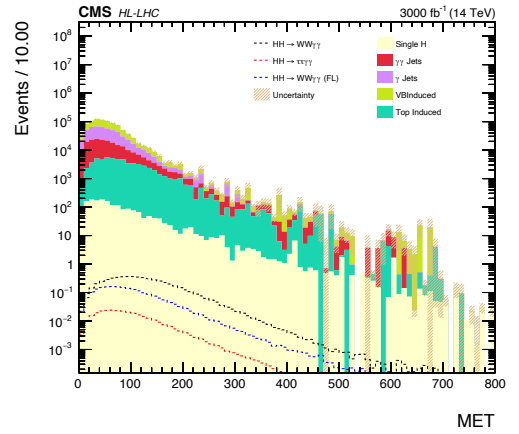


Figure A.3f : E_T^{miss}

Figure A.3 : DNN input distributions for the semi-leptonic final state (continued).

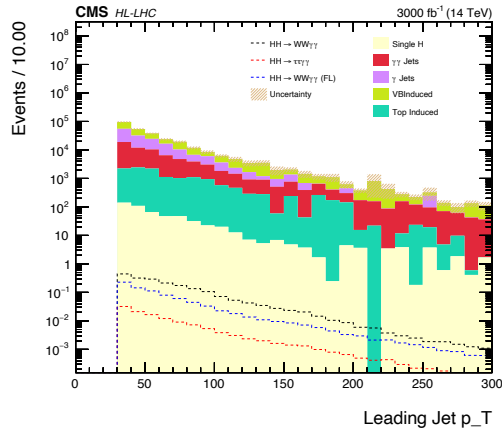


Figure A.4a : Leading jet p_T

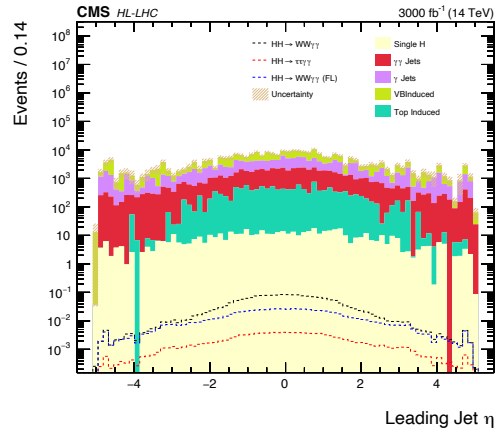


Figure A.4b : Leading jet η

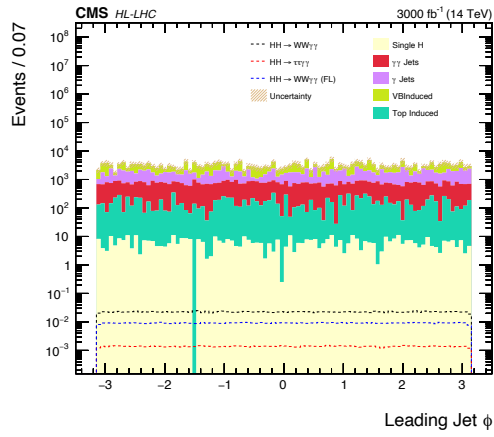


Figure A.4c : Leading jet ϕ

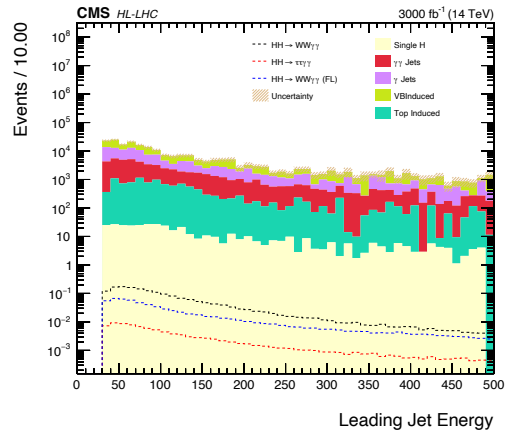


Figure A.4d : Leading jet energy

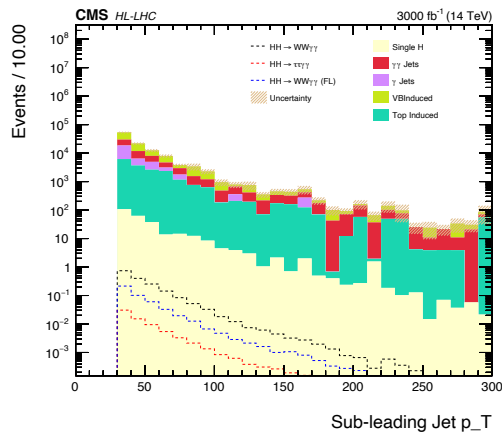


Figure A.4e : Sub-leading jet p_T

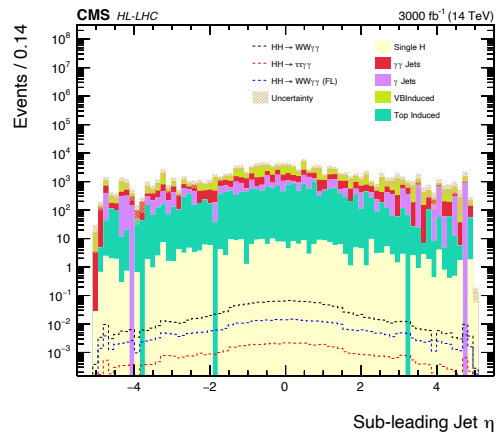


Figure A.4f : Sub-leading jet η

Figure A.4 : DNN input distributions for the semi-leptonic final state (continued).

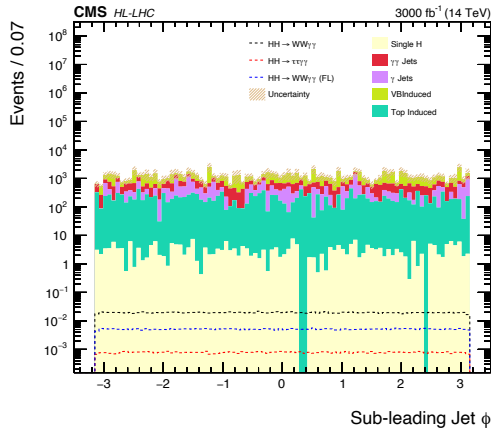


Figure A.5a : Sub-leading jet ϕ

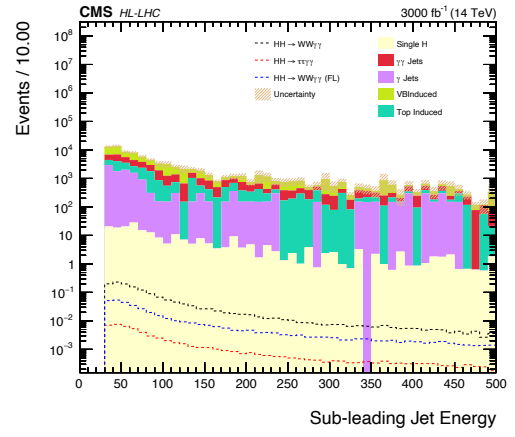


Figure A.5b : Sub-leading jet energy

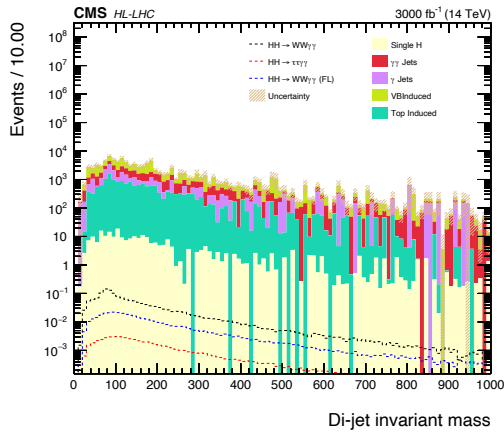


Figure A.5c : m_{j_0, j_1}

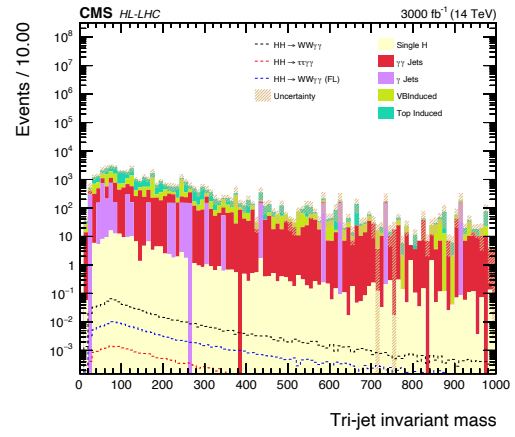


Figure A.5d : m_{j_1, j_2}

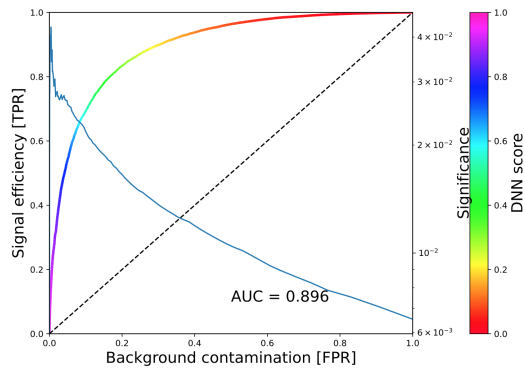


Figure A.5e : Odd training

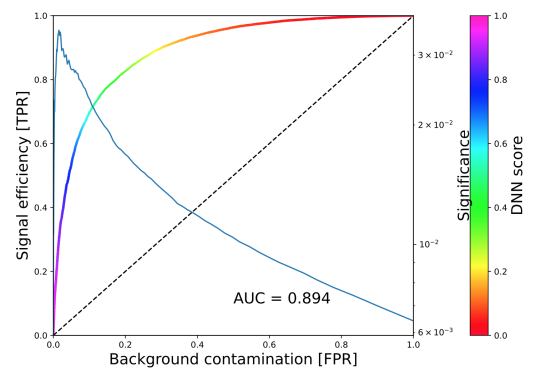


Figure A.5f : Even training

Figure A.5 : DNN input distributions (a,b,c,d) and the ROC curves (e,f) for the semi-leptonic final state (continued).

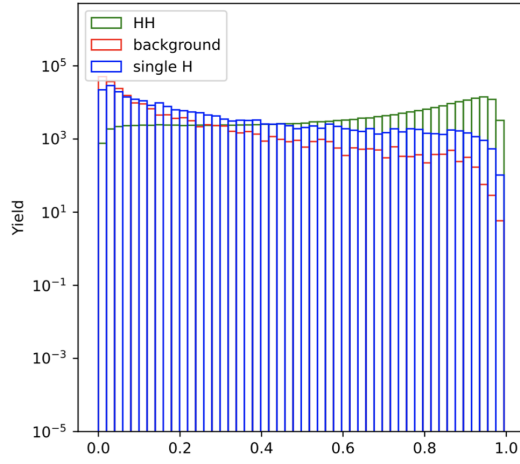


Figure A.6a : Odd training

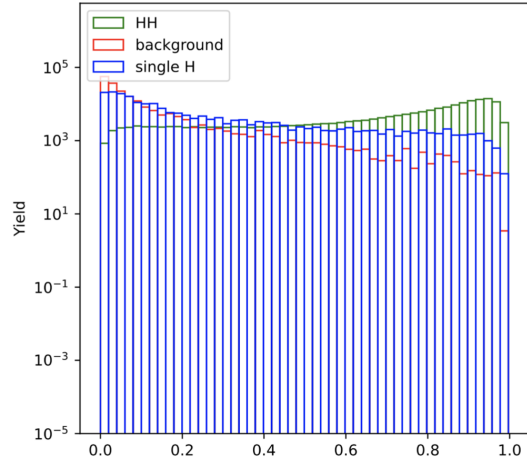


Figure A.6b : Even training

Figure A.6 : DNN evaluations for the semi-leptonic channel final state.

Table A.6 : Hyper-parameter settings for the DNN performed for semi-leptonic channel.

Hyper-parameter	Setting
Epochs	200
Batch size	256
Learning rate	0.001
Optimiser	Adam
Loss function	Categorical Cross Entropy
Hidden layer activation functions	ReLU
Output layer activation function	sigmoid

APPENDIX A.4

Table A.7 : Cut-flow report showing number of events, before selections, in the semi-leptonic channel and in its categories. Percentages in brackets show the total selection efficiency.

Samples	No selection	Fully-leptonic final state
$HH \rightarrow WW\gamma\gamma$	$4.69e+01$	$2.14e-02$ (0.046%)
$HH \rightarrow WW\gamma\gamma(FL)$	$1.12e+01$	$3.25e-01$ (2.911%)
$HH \rightarrow \tau\tau\gamma\gamma$	$3.13e+00$	$1.64e-02$ (0.525%)
Signal	$1.10e+02$	$3.63e-01$ (0.331%)
$GGH \rightarrow \gamma\gamma$	$3.44e+05$	$0.00e+00$ (0.000%)
$VBFH \rightarrow \gamma\gamma$	$2.85e+04$	$5.38e-02$ (0.000%)
$ttH \rightarrow \gamma\gamma$	$4.18e+03$	$2.39e+00$ (0.057%)
$VH \rightarrow \gamma\gamma$	$1.63e+04$	$3.03e+00$ (0.019%)
THQ	$6.16e+02$	$7.81e-02$ (0.013%)
$\gamma\gamma + jets_{80-Inf}$	$2.96e+08$	$2.87e+02$ (0.000%)
$\gamma\gamma + jets_{40-80}$	$9.98e+08$	$0.00e+00$ (0.000%)
$G + jets$	$2.99e+09$	$0.00e+00$ (0.000%)
$G + jets_{20-40GeV}$	$7.83e+08$	$0.00e+00$ (0.000%)
$G + jets_{20-Inf}$	$1.17e+10$	$0.00e+00$ (0.000%)
$W1Jets \rightarrow Lv$	$3.11e+10$	$0.00e+00$ (0.000%)
$W2Jets \rightarrow Lv$	$8.90e+09$	$0.00e+00$ (0.000%)
$W3Jets \rightarrow Lv$	$3.80e+09$	$0.00e+00$ (0.000%)
$WGJJ$	$1.81e+07$	$1.00e+01$ (0.000%)
$WGGJets$	$5.65e+06$	$5.70e+00$ (0.000%)
$DYJets$	$1.71e+10$	$0.00e+00$ (0.000%)
ZG	$4.36e+08$	$6.20e+02$ (0.000%)
$WW(inclusive)$	$2.11e+08$	$1.91e+01$ (0.000%)
$t\bar{t}(inclusive)$	$2.59e+09$	$5.25e+01$ (0.000%)
$ttGJets$	$1.37e+07$	$7.83e+01$ (0.001%)
$ttGG$	$5.59e+04$	$4.89e+00$ (0.009%)
ttW	$6.76e+05$	$2.52e-01$ (0.000%)
Background	$8.10e+10$	$1.08e+03$ (0.000%)

APPENDIX A.5

Table A.8 : Input variables used to train single τ final state DNN.

Feature	Description
Leading Photon $p_T / m_{\gamma\gamma}$	p_T of the leading good photon scaled to diphoton mass.
Leading Photon Energy / $m_{\gamma\gamma}$	Energy of the leading good photon scaled to diphoton mass.
Leading Photon η	Pseudorapidity of the leading good photon
Leading Photon ϕ	Direction in the transverse plane of the leading good photon
Sub-leading Photon $p_T / m_{\gamma\gamma}$	p_T of the sub-leading good photon scaled to diphoton mass.
Sub-leading Photon Energy / $m_{\gamma\gamma}$	Energy of the sub-leading good photon scaled to diphoton mass.
Sub-leading Photon η	Pseudorapidity of the sub-leading good photon
Sub-leading Photon ϕ	Direction in the transverse plane of the sub-leading good photon
Leading Tau p_T	p_T of the leading good Tau
Leading Tau η	Pseudorapidity of the leading good Tau
Leading Tau ϕ	Direction in the transverse plane of the leading good Tau
Leading Tau Energy	Energy of the leading good Tau
Jet Multiplicity	Number of jets in the event (flavour inclusive)
B-tagged jet Multiplicity	Number of b-tagged jets in the event (flavour inclusive)
Leading Jet p_T	p_T of the leading good jet
Leading Jet η	Pseudorapidity of the leading good jet
Sub-leading Jet p_T	p_T of the sub-leading good jet
Sub-leading Jet η	Pseudorapidity of the sub-leading good jet
MET	Missing transverse energy in the event

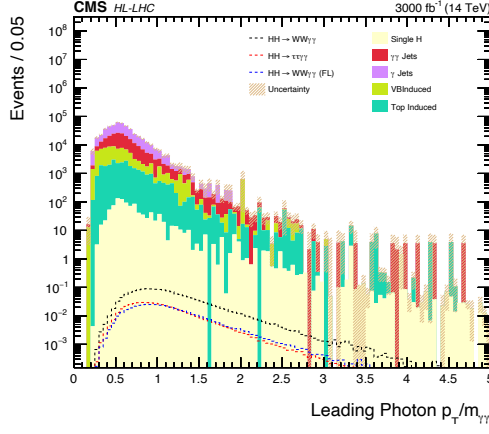


Figure A.7a : Leading Photon $p_T/m_{\gamma\gamma}$

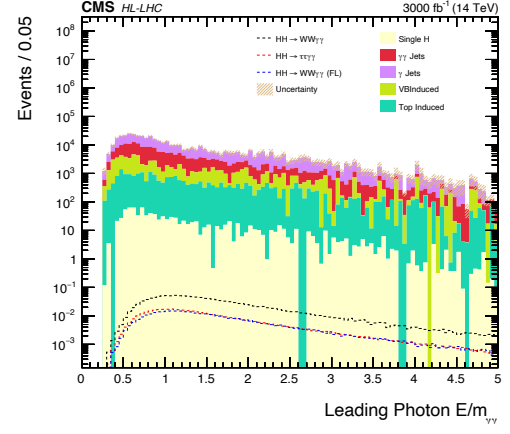


Figure A.7b : Leading Photon $E/m_{\gamma\gamma}$

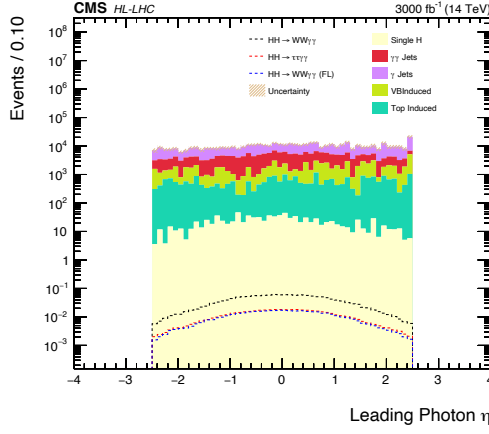


Figure A.7c : Leading Photon η

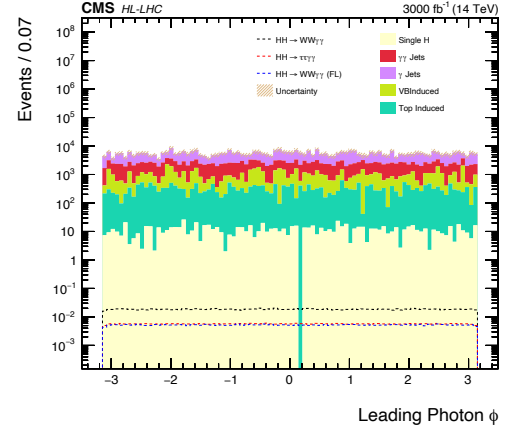


Figure A.7d : Leading Photon ϕ

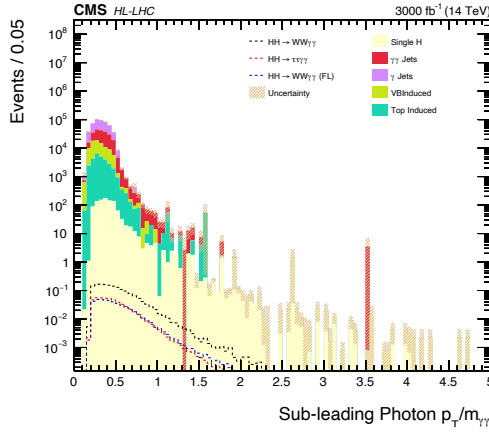


Figure A.7e : Sub-leading Photon $p_T/m_{\gamma\gamma}$

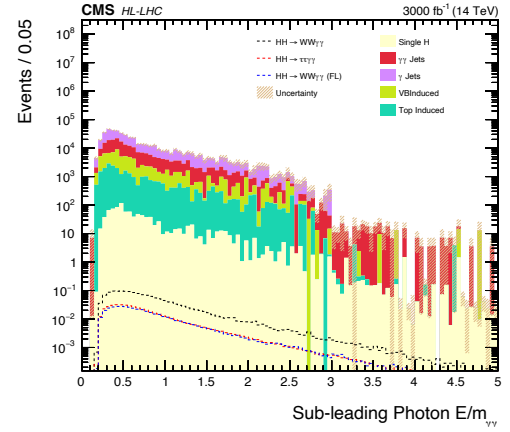


Figure A.7f : Sub-leading Photon $E/m_{\gamma\gamma}$

Figure A.7 : DNN input distributions for the single τ final state.

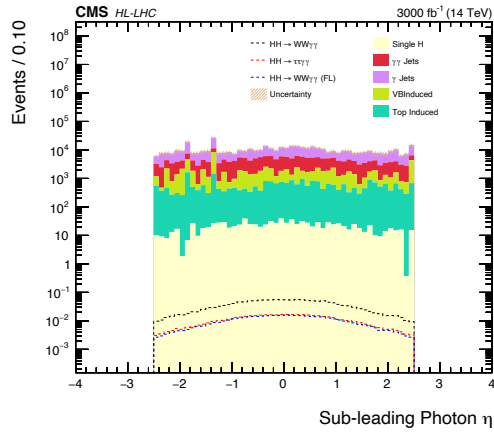


Figure A.8a : Sub-leading Photon η

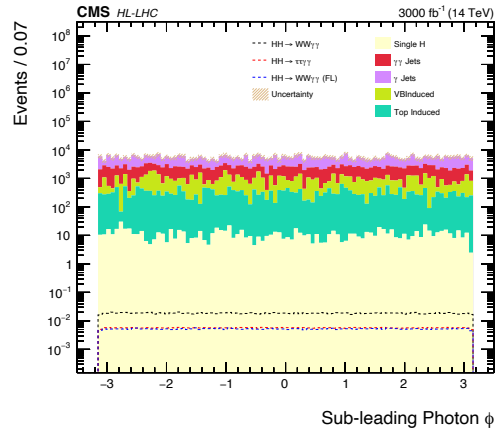


Figure A.8b : Sub-leading Photon ϕ

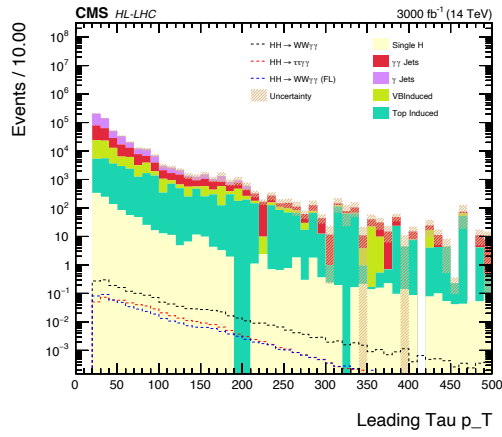


Figure A.8c : Leading Tau p_T

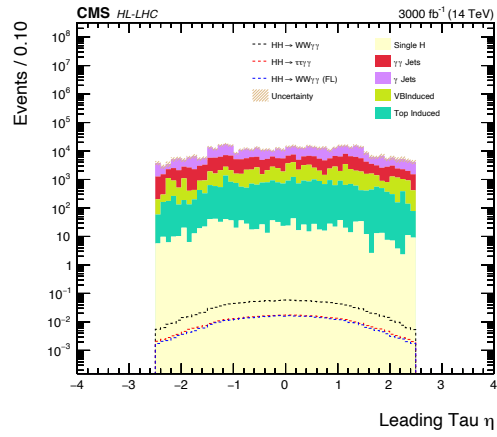


Figure A.8d : Leading Tau η

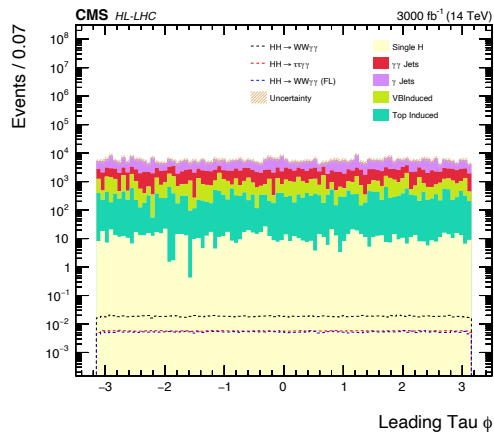


Figure A.8e : Leading Tau ϕ

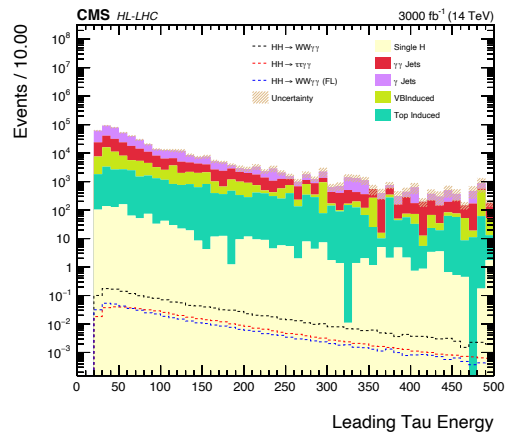


Figure A.8f : Leading Tau energy

Figure A.8 : DNN input distributions for the single τ final state (continued).

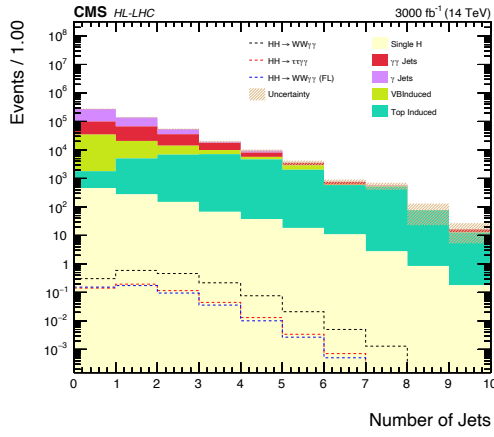


Figure A.9a : Jet multiplicity

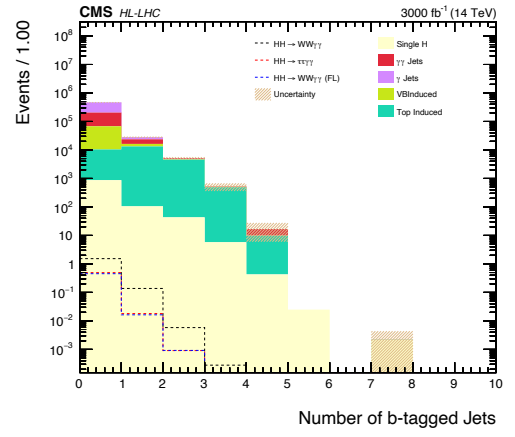


Figure A.9b : B-tagged jet multiplicity

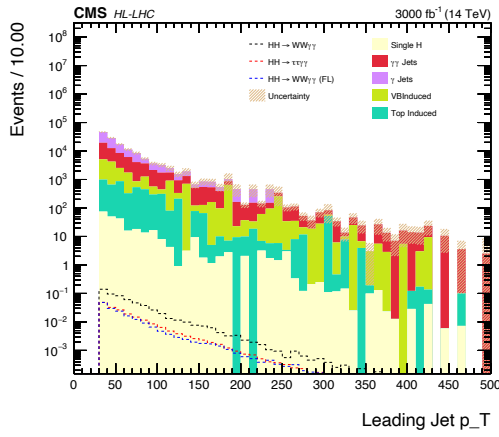


Figure A.9c : Leading jet p_T

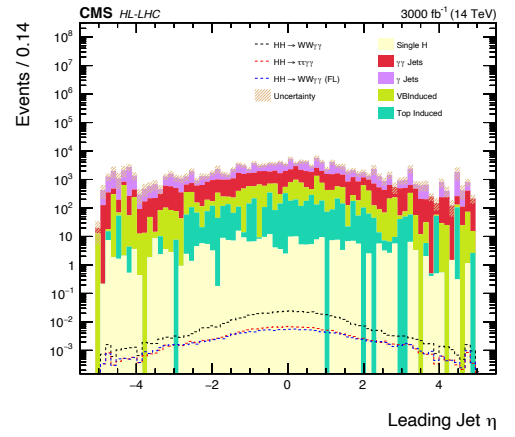


Figure A.9d : Leading jet η

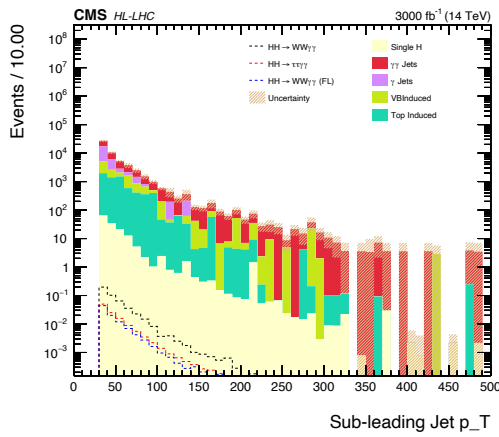


Figure A.9e : Sub-leading jet p_T

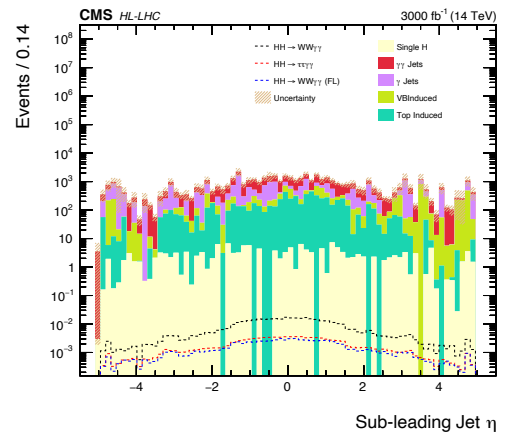


Figure A.9f : Sub-leading jet η

Figure A.9 : DNN input distributions for the single τ final state (continued).

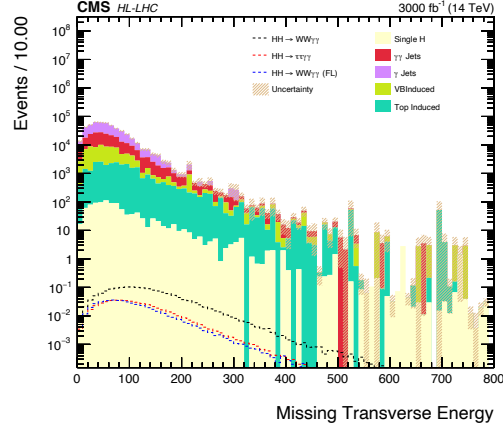


Figure A.10a : E_T^{miss}

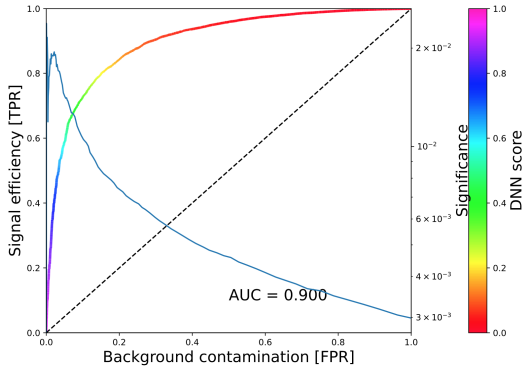


Figure A.10b : Odd ROC curve

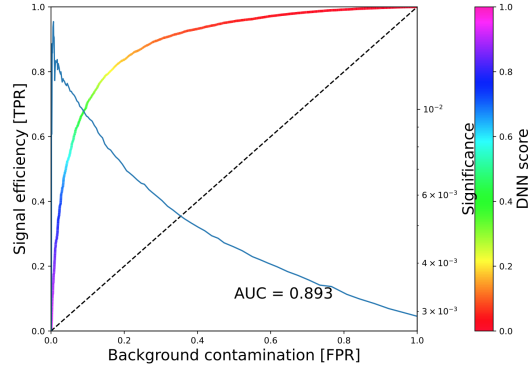


Figure A.10c : Even ROC curve

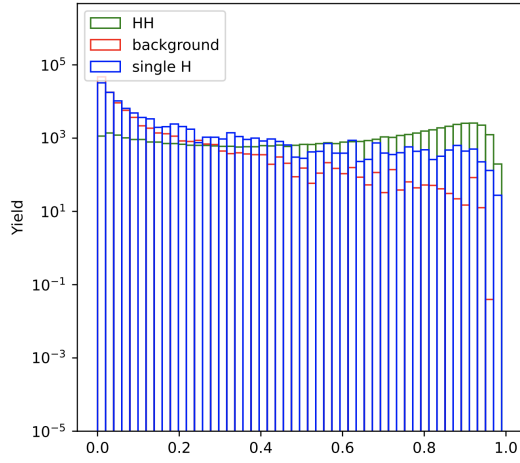


Figure A.10d : Odd training weight

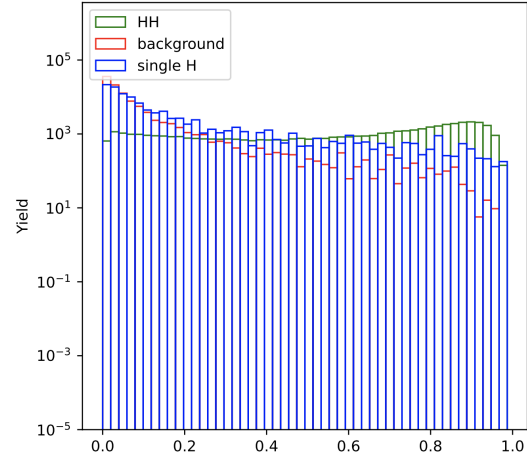


Figure A.10e : Even training weight

Figure A.10 : DNN input distributions for the single τ final state (continued) (a), corresponding ROC curves (b,c) and DNN evaluation on odd (d) and even (e) data.

Table A.9 : Cut-flow report showing number of events, before selections, in the single τ final state and in its categories. Percentages in brackets show the total selection efficiency.

Samples	No Selection	Single τ	DNN Cat. 1	DNN Cat. 2
$HH \rightarrow WW\gamma\gamma$	4.69e+01	1.82e+00 (3.879%)	7.97e-01 (1.699%)	8.60e-01 (1.833%)
$HH \rightarrow WW\gamma\gamma(FL)$	1.12e+01	4.98e-01 (4.459%)	1.91e-01 (1.710%)	2.57e-01 (2.304%)
$HH \rightarrow \tau\tau\gamma\gamma$	3.13e+00	5.24e-01 (16.734%)	2.32e-01 (7.410%)	2.34e-01 (7.457%)
Signal	1.10e+02	3.19e+00 (2.908%)	1.44e+00 (1.310%)	1.43e+00 (1.303%)
$GGH \rightarrow \gamma\gamma$	3.44e+05	6.34e+02 (0.184%)	1.53e+02 (0.044%)	8.34e+00 (0.002%)
$VBFH \rightarrow \gamma\gamma$	2.85e+04	7.20e+01 (0.252%)	3.47e+01 (0.122%)	3.28e+00 (0.012%)
$ttH \rightarrow \gamma\gamma$	4.18e+03	1.36e+02 (3.250%)	4.68e+01 (1.120%)	6.89e+00 (0.165%)
$VH \rightarrow \gamma\gamma$	1.63e+04	1.72e+02 (1.050%)	6.92e+01 (0.424%)	2.81e+01 (0.172%)
THQ	6.16e+02	1.35e+01 (2.198%)	6.20e+00 (1.007%)	1.97e+00 (0.319%)
$\gamma\gamma + jets_{80-Inf}$	2.96e+08	1.42e+05 (0.048%)	3.58e+04 (0.012%)	1.82e+03 (0.001%)
$\gamma\gamma + jets_{40-80}$	9.98e+08	1.99e+03 (0.000%)	3.46e+02 (0.000%)	2.82e+01 (0.000%)
$G + jets$	2.99e+09	2.10e+05 (0.007%)	3.33e+04 (0.001%)	1.49e+03 (0.000%)
$G + jets_{20-40GeV}$	7.83e+08	1.12e+04 (0.001%)	1.64e+02 (0.000%)	0.00e+00 (0.000%)
$G + jets_{20-Inf}$	1.17e+10	4.16e+04 (0.000%)	1.17e+03 (0.000%)	0.00e+00 (0.000%)
$W1Jets \rightarrow Lv$	3.11e+10	8.04e+03 (0.000%)	8.04e+02 (0.000%)	0.00e+00 (0.000%)
$W2Jets \rightarrow Lv$	8.90e+09	5.56e+03 (0.000%)	4.12e+02 (0.000%)	2.06e+02 (0.000%)
$W3Jets \rightarrow Lv$	3.80e+09	6.04e+03 (0.000%)	1.34e+03 (0.000%)	6.72e+02 (0.000%)
$WGJJ$	1.81e+07	2.00e+03 (0.011%)	7.63e+02 (0.004%)	1.71e+02 (0.001%)
$WGGJets$	5.65e+06	1.81e+03 (0.032%)	5.76e+02 (0.010%)	7.42e+01 (0.001%)
$DYJets$	1.71e+10	2.14e+04 (0.000%)	2.67e+03 (0.000%)	4.46e+02 (0.000%)
ZG	4.36e+08	1.52e+04 (0.003%)	3.76e+03 (0.001%)	2.63e+02 (0.000%)

Table A.10 : Cut-flow report showing number of events, before selections, in the single τ final state and in its categories (cont'd).

Samples	No Selection	Single τ	DNN Cat. 1	DNN Cat. 2
$WW(inclusive)$	$2.11e+08$	$5.65e+02$ (0.000%)	$1.66e+02$ (0.000%)	$2.34e+01$ (0.000%)
$t\bar{t}(inclusive)$	$2.59e+09$	$2.29e+04$ (0.001%)	$5.51e+03$ (0.000%)	$4.20e+02$ (0.000%)
$t\bar{t}GJets$	$1.37e+07$	$4.50e+03$ (0.033%)	$1.30e+03$ (0.009%)	$8.62e+01$ (0.001%)
$t\bar{t}GG$	$5.59e+04$	$2.47e+02$ (0.441%)	$6.72e+01$ (0.120%)	$7.81e+00$ (0.014%)
$t\bar{t}W$	$6.76e+05$	$2.85e+01$ (0.004%)	$6.31e+00$ (0.001%)	$2.52e-01$ (0.000%)
Background	$8.10e+10$	$4.96e+05$ (0.001%)	$8.85e+04$ (0.000%)	$5.76e+03$ (0.000%)

APPENDIX A.6

Table A.11 : Cut-flow report showing number of events, before selections and in the double τ final state of $HH \rightarrow \tau\tau\gamma\gamma$.

Samples	noSel	Two Taus
$HH \rightarrow WW\gamma\gamma$	$4.69e+01$	$8.29e-02$ (0.177%)
$HH \rightarrow WW\gamma\gamma(FL)$	$1.12e+01$	$2.48e-02$ (0.222%)
$HH \rightarrow \tau\tau\gamma\gamma$	$3.13e+00$	$1.10e-01$ (3.495%)
Signal	$1.10e+02$	$2.22e-01$ (0.202%)
$GGH \rightarrow \gamma\gamma$	$3.44e+05$	$1.39e+00$ (0.000%)
$VBFH \rightarrow \gamma\gamma$	$2.85e+04$	$1.08e-01$ (0.000%)
$ttH \rightarrow \gamma\gamma$	$4.18e+03$	$7.17e+00$ (0.171%)
$VH \rightarrow \gamma\gamma$	$1.63e+04$	$4.29e+00$ (0.026%)
THQ	$6.16e+02$	$3.41e-01$ (0.055%)
$\gamma\gamma + jets_{80-Inf}$	$2.96e+08$	$5.14e+02$ (0.000%)
$\gamma\gamma + jets_{40-80}$	$9.98e+08$	$7.06e+00$ (0.000%)
$G + jets$	$2.99e+09$	$0.00e+00$ (0.000%)
$G + jets_{20-40GeV}$	$7.83e+08$	$0.00e+00$ (0.000%)
$G + jets_{20-Inf}$	$1.17e+10$	$0.00e+00$ (0.000%)
$W1Jets \rightarrow Lv$	$3.11e+10$	$0.00e+00$ (0.000%)
$W2Jets \rightarrow Lv$	$8.90e+09$	$0.00e+00$ (0.000%)
$W3Jets \rightarrow Lv$	$3.80e+09$	$0.00e+00$ (0.000%)
$WGJJ$	$1.81e+07$	$1.00e+01$ (0.000%)
$WGGJets$	$5.65e+06$	$8.56e+00$ (0.000%)
$DYJets$	$1.71e+10$	$2.23e+02$ (0.000%)
ZG	$4.36e+08$	$7.33e+02$ (0.000%)
$WW(inclusive)$	$2.11e+08$	$1.49e+01$ (0.000%)
$t\bar{t}(inclusive)$	$2.59e+09$	$1.05e+03$ (0.000%)
$ttGJets$	$1.37e+07$	$1.45e+02$ (0.001%)
$ttGG$	$5.59e+04$	$9.98e+00$ (0.018%)
ttW	$6.76e+05$	$2.52e+00$ (0.000%)
Background	$8.10e+10$	$2.73e+03$ (0.000%)

APPENDIX A.7

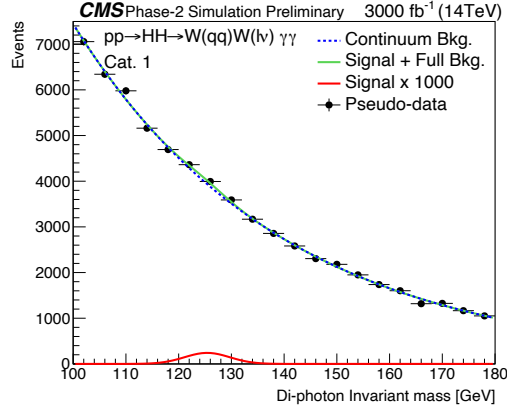


Figure A.11a : Cat. 1

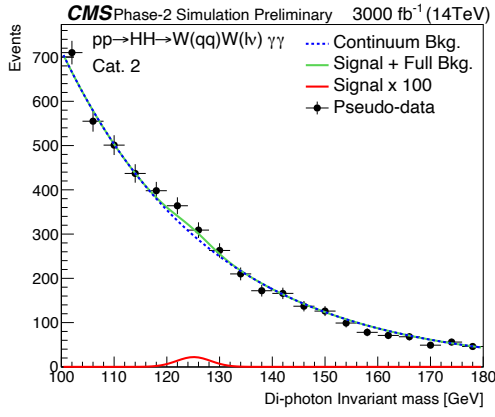


Figure A.11b : Cat. 2

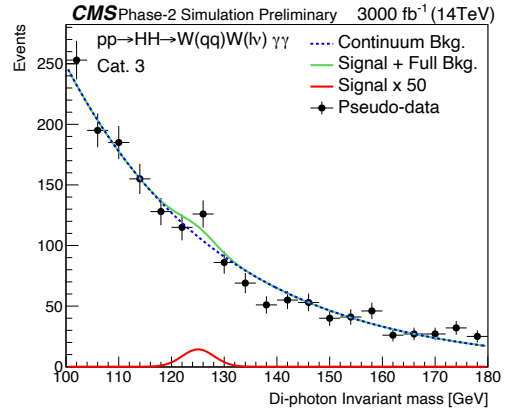


Figure A.11c : Cat. 3

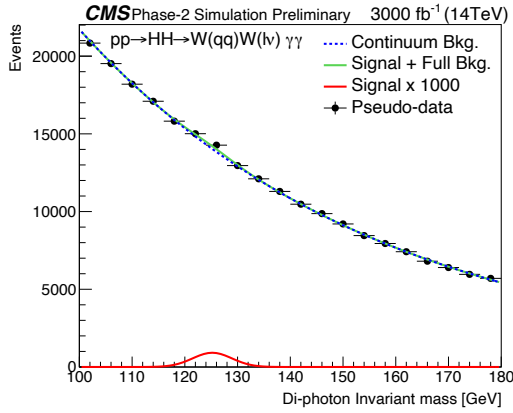


Figure A.11d : Semi-leptonic

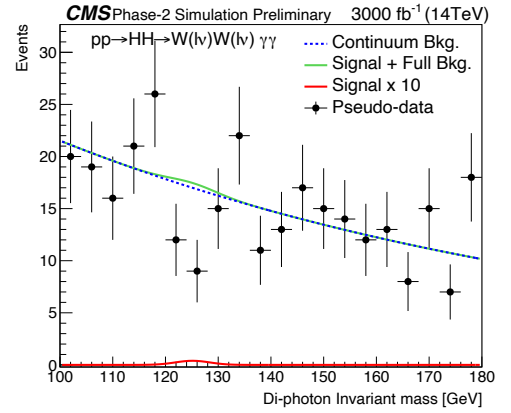


Figure A.11e : Fully-leptonic

Figure A.11 : Fitted distributions of semi-leptonic channel (a,b,c) and fully-leptonic (d,e).

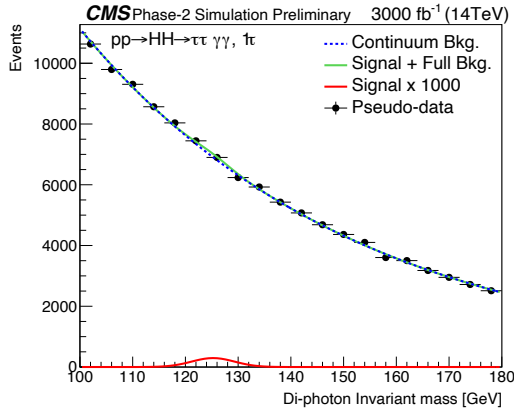


Figure A.12a : Single τ

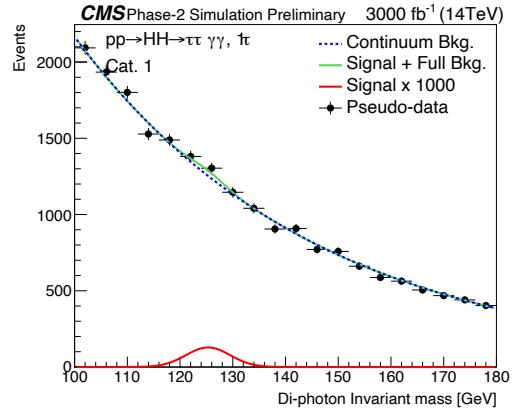


Figure A.12b : Single τ Cat. 1

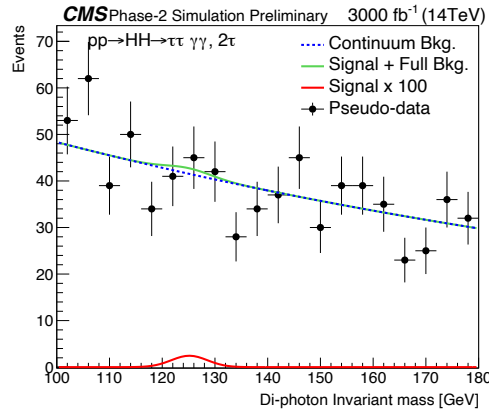


Figure A.12c : Double τ

Figure A.12 : Fitted distributions of single τ (a,b) and double τ final states.

CURRICULUM VITAE



Name Surname : Ahmet Oğuz Güzel
Place and Date of Birth : Turkey, 1995
E-Mail : guzelah@itu.edu.tr

EDUCATION

- **M.Sc.** : 2022, Istanbul Technical University, Graduate School, Department of Physics Engineering, Physics Engineering Programme
- **B.Sc.** : 2019, Istanbul Technical University, Faculty of Aeronautics and Astronautics, Department of Astronautical Engineering

PROFESSIONAL EXPERIENCE

- CMS User, CERN
- Associate researcher at the Turkish Atomic Energy Authority, TAEK

PUBLICATIONS

- **The ATLAS and CMS Collaborations** 2022. Snowmass White Paper Contribution: Physics with the Phase-2 ATLAS and CMS Detectors, *ATL-PHYS-PUB-2022-018*, *CMS-PAS-FTR-22-001*, <https://cds.cern.ch/record/2806962?ln=en>.
- **The CMS Collaboration** 2022. Prospects for HH measurements in the $WW\gamma\gamma$ and $\tau\tau\gamma\gamma$ final states in proton-proton collisions at $\sqrt{s} = 14$ TeV at the HL-LHC, *CMS-PAS-FTR-21-003*, <https://cds.cern.ch/record/2804003?ln=en>.
- **Altan Cakir and Oguz Guzel** 2019. A Brief Review of Plasma Wakefield Acceleration, *arXiv*, <https://arxiv.org/abs/1908.07207>.

A. O. GÜZEL

PROSPECTS OF NONRESONANT HIGGS BOSON PAIR PRODUCTION MEASUREMENT
IN THE $W W \gamma \gamma$ CHANNEL AT THE HL-LHC WITH THE PHASE-II CMS DETECTOR

2022