

University of Southampton Research Repository

Copyright © and Moral Rights for this thesis and, where applicable, any accompanying data are retained by the author and/or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This thesis and the accompanying data cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder/s. The content of the thesis and accompanying research data (where applicable) must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holder/s.

When referring to this thesis and any accompanying data, full bibliographic details must be given, e.g.

Thesis: Author (Year of Submission) "Full thesis title", University of Southampton, name of the University Faculty or School or Department, PhD Thesis, pagination.

Data: Author (Year) Title. URI [dataset]

UNIVERSITY OF SOUTHAMPTON

Faculty of Engineering and Physical Sciences

School of Physics & Astronomy

**A new L1 trigger algorithm for the HL-LHC
upgrade
&
Search for light Higgs bosons via NMSSM
cascade decays at the CMS experiment.**

by

Giovanna Salvi

ORCID: [0000-0002-2787-1063](https://orcid.org/0000-0002-2787-1063)

A thesis for the degree of

Doctor of Philosophy

30 July 2024

University of Southampton

Abstract

Faculty of Engineering and Physical Sciences
School of Physics & Astronomy

Doctor of Philosophy

A new L1 trigger algorithm for the HL-LHC upgrade

&

Search for light Higgs bosons via NMSSM cascade decays at the CMS experiment.

by Giovanna Salvi

This thesis describes three different studies conducted within the framework of the Compact Muon Solenoid (CMS) experiment at CERN.

The performance of the CMS experiment at the Large Hadron Collider relies heavily on the accurate calibration of its subdetectors. Among these, the Electromagnetic Calorimeter, made of scintillating crystals, plays the crucial role of measuring the energy of photons and electrons. The transparency of these crystals degrades due to radiation exposure, necessitating frequent corrections to maintain optimal performances. The correction process was improved from Run 2 to Run 3 and is presented here.

Furthermore, this thesis discusses a search for light Higgs bosons in the framework of the Next-to-Minimal Supersymmetric Standard Model. In particular, it focuses on identifying the events characterized by minimal missing transverse momentum, where the light Higgs bosons decay into boosted pairs of b quarks. This work describes the reduction and estimation of the main background sources for this analysis, with particular attention to the $t\bar{t}$ background.

Lastly, a new Level 1 trigger algorithm, developed for potential implementation in the CMS experiment for the High-Luminosity Large Hadron Collider era (HL-LHC), is introduced. HL-LHC will see a massive upgrade to the detector with a completely new L1 trigger system. For the first time, tracks will be reconstructed at the L1 trigger. This new algorithm will exploit reconstructed track information enabling the identification of hard interaction vertices.

Acknowledgements

I would like to thank my supervisors, Stefano Moretti and Claire Shepherd-Themistocleous for their guidance and support throughout the course of my doctoral research.

A special thanks goes to Jacob Linacre. Thank you for all the knowledge shared, for the guidance in the NMSSM analysis, for your patience in answering all my questions and for being a brilliant scientist and an example for me. Your support has been crucial in this journey.

Special thanks to Emmanuel Olaiya for the support and guidance with the trigger project. Thanks to all the RAL CMS group people! Thanks for the advice and help along the way.

Thanks to all my Oxford and my CERN and Geneva friends!

Grazie Balc, per il tuo amore, la tua pazienza e il tuo supporto. Grazie per essere stato così importante in questo ultimo anno di vita.

Grazie a Zio Peppe, che mi ha insegnato che per vincere bisogna saper perdere. Che la vera vittoria arriva quando siamo noi a fare un passo indietro di fronte ai nostri muri e fare muro contro muro non porta a niente.

Grazie a Luciana, Concezio e Alberto e a tutta la mia famiglia allargata! Grazie ai miei amici e le mie amiche di sempre e a whatsapp che anche se lontani ci ha permesso di aggiornarci sulle nostre vite. Grazie per il vostro supporto, il vostro amore e il vostro incoraggiamento.

Declaration of Authorship

I declare that this thesis and the work presented in it is my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. None of this work has been published before submission

Signed:.....

Date:.....

Author Contributions

Part I provides an introduction to the theoretical and experimental framework within which this thesis work is developed and does not contain original work.

Part II discusses the upgrade of the transparency correction strategy applied to the Electromagnetic Calorimeter of CMS. This is an original work performed by me supervised by Uttiya Sarkar, David Petyt and Amina Zghiche. This work was documented in an internal CMS note and the results have been implemented in the Run 3 data-taking procedure.

In Part III I performed the full Monte Carlo production and the related studies described in Chapter 5. The comparison between the previous and new analysis tools as well as the test of the previous QCD background estimation strategy using the new tools were performed by Mei-Li Holmberg exploiting the Monte Carlo samples produced by me. I conducted the full $t\bar{t}$ background estimation study described in Section 7.2.

Part IV discusses a new L1 trigger algorithm, thought to be implemented during HL-LHC in the CMS detector. The design, C++ implementation, test and FPGA emulation of the algorithm is my original work. The VHDL implementation is done by Thomas Schuh.

Contents

Acknowledgements	v
Declaration of Authorship	vii
Author Contributions	ix
List of Figures	xv
List of Tables	xxvii
I Theory and Experimental Layout	1
1 Theoretical framework	3
1.1 The Standard Model	3
1.1.1 The SM Lagrangian	5
1.1.2 Open questions	9
1.2 Supersymmetry	13
1.2.1 The Minimal Supersymmetric Standard Model	15
1.2.2 The Next-to-Minimal Supersymmetric Standard Model	18
2 The CMS Experiment at the LHC	21
2.1 The Large Hadron Collider	21
2.1.1 The accelerating system	22
2.1.2 Operations	23
2.2 The CMS Experiment	25
2.2.1 Coordinate system	26
2.2.2 Structure of the detector	27
2.2.3 Solenoid Magnet	28
2.2.4 Tracker	29
2.2.5 Electromagnetic calorimeter	31
2.2.6 Hadronic calorimeter	32
2.2.7 Muon system	33
2.2.8 Trigger system	33
2.3 Physics Object Identification and Reconstruction	34
2.3.1 The Particle Flow algorithm	34
2.3.2 Jets	35

II	The CMS Electromagnetic Calorimeter: transparency corrections updates	39
3	Transparency corrections updates in the Electromagnetic Calorimeter	41
3.1	Crystal transparency loss mechanisms	42
3.1.1	Laser monitoring system	42
3.1.2	Response losses measured in LHC Run 1 and Run 2	43
3.2	Transparency corrections	44
3.2.1	L1/HLT corrections history in Run 1 and Run 2	45
3.2.2	Planned L1/HLT update frequency in Run 3	46
3.3	Impact on ECAL energy scale	47
3.4	Impact on L1 and HLT rates	47
3.4.1	L1 rates	47
3.4.2	HLT rates	50
3.5	Impact on L1/HLT resolutions	51
3.5.1	L1 resolution	51
3.5.2	HLT resolution	52
3.6	Workflow of the transparency updates	53
3.7	Procedure implementation in Run 3	57
3.8	Conclusion of the study	58
III	Search for Light Higgs Bosons from SUSY Decay Cascades	59
4	Theoretical and experimental motivation	61
4.1	Introduction	61
4.2	The Signal Model	63
5	Monte Carlo production	67
5.1	Signal MC	67
5.2	MadGraph and Pythia	69
5.3	The merging problem	69
5.4	Comparison between the two MC scenarios	72
5.5	MC production chain	74
6	Event selection	77
6.1	The analysis framework	77
6.2	Double-b-Tagging	77
6.2.1	The btagHbb double-b-tagger	78
6.2.2	ParticleNet-tagger	79
6.3	Jet Mass Reconstruction	79
6.3.1	The Softdrop algorithm	80
6.3.2	The ParticleNet mass	80
6.4	Kinematic Event Selection	81
6.5	Performance Comparison and Cut Criteria	82
6.5.1	Double-b-tag Selection	83
6.5.2	Mass Event Selection	85
6.5.3	Yields in the Signal Mass Region	87

7	Background estimations	93
7.1	QCD Background Estimation	93
7.1.1	Data-driven QCD Estimation	95
7.2	The $t\bar{t}$ Background Estimation	96
7.2.1	The Electron study	99
7.2.2	The Muon study	102
7.2.3	The Tau study	106
7.2.4	The Hadronic study	108
7.2.5	Final choice of $t\bar{t}$ veto cuts	110
7.2.6	Uncertainties associated with the $t\bar{t}$ veto cuts	112
7.2.7	Data-driven $t\bar{t}$ Estimation	115
7.2.8	Discussion of Expected Limits	116
IV	A new algorithm for the L1 trigger at CMS at HL-LHC	119
8	The CMS Phase-2 upgrade	121
8.1	The High-Luminosity Large Hadron Collider	121
8.1.1	Overview of the CMS detector upgrade	121
8.1.2	Physics motivations and objectives	122
8.1.3	Upgrade of the Outer Tracker	123
8.2	The upgrade of the CMS Trigger System	124
8.2.1	The Global Track Trigger	128
9	The Peak Finding Algorithm	129
9.1	Designing the algorithm	129
9.1.1	Distribution of the residuals and algorithm's parameters	131
9.1.2	Adding the η information	138
9.1.3	Exploring supplementary discrimination factors	146
9.2	Comparison between the PFA and FastHisto	153
9.3	The implementation of the algorithm on an FPGA	157
9.3.1	Field-Programmable Gate Array	157
9.3.2	FPGA at LHC	159
9.3.3	Implementation of Peak Finding Algorithm on an FPGA	160
9.4	Conclusion of the study	166
V	Conclusions	169
10	Conclusions	171
Appendix A	DJR plots	173
Appendix A.1	First scenario	173
Appendix A.2	Second scenario	174
Appendix B	Supplementary discrimination factor for PFA	177
Appendix B.0.1	QCD 30-50 GeV	178
Appendix B.0.2	QCD 80-120 GeV	183

Bibliography**189**

List of Figures

1.1	Scheme of the Standard Model constituents. Figure from Ref. [1].	4
1.2	Shape of the Higgs potential with spontaneous symmetry breaking. Figure from Ref. [2].	8
1.3	The Feynman diagrams of the one-loop mass corrections to the photon (a), electron (b) and the scalar Higgs boson (c).	11
1.4	Mass ranges of the lightest CP-even Higgs in various supersymmetric extensions of the SM, as resulted from a random scan of the parameters space, accounting experimental and theoretical constraints. Together with the MSSM, there is the Next-to-Minimal Supersymmetric Standard Model (NMSSM), the new Minimally-extended Supersymmetric SM (nMSSM) and UMSSM which is a $U(1)'$ extension of the MSSM. The motivations of the bounds are also provided. Th. = theoretical bound; Scan = limit value is sensitive to the chosen parameter space; LEP = experimental constraints from LEP. Figure from Ref. [3].	19
2.1	The accelerator complex and the various experiments operating at CERN. Figure from Ref. [4].	22
2.2	The LHC plan spanning from 2011 to 2029, with luminosity and energy goals, reported in red at the bottom and at the top, respectively. At the end of Run 3 in 2025, CMS will undergo the Phase-2 upgrade in preparation for the High-Luminosity LHC (HL-LHC). Figure from Ref. [5]. . .	24
2.3	On the left, the CMS integrated luminosity data per year from 2010 to July 2023. On the right, focus on Run 3 integrated luminosity until July 2023. The CMS experiment recorded a total of about 67 fb^{-1} with a peak instantaneous luminosity of $2.22 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$, more than doubling the LHC design luminosity. Figure from Ref. [6].	25
2.4	The CMS coordinate system. Figure from Ref. [1].	26
2.5	The scheme of the CMS detector. Key subsystems are highlighted. Figure from Ref. [7].	27
2.6	The Phase-1 tracker configuration is shown, illustrating one quadrant of the tracker in the r-z plane, centred around the interaction point. The pixel detector is represented in green, while the single-sided and double-sided strip detectors are displayed in red and blue, respectively. Figure from Ref. [8].	30
2.7	Layout of the CMS electromagnetic calorimeter showing the arrangement of crystal modules (in green), supermodules and endcaps. Figure from Ref. [9].	31

2.8	The results of the three main jet reconstruction algorithms, performed on the same data and with the same radius R , is shown. Different colours are used to represent the different jets. The k_T algorithm produces highly irregular clusters since it prioritizes combining soft and collinear particles. The Cambridge/Aachen algorithm neglects momentum information and looks only at angular distances. The anti- k_T algorithm produces regular clusters, conically shaped, centred around the hardest particle. Figure from Ref. [10].	37
2.9	Sketch of a heavy-flavour jet and its main features. The jet results in a cluster of charged particles (often including a soft lepton) coming from a secondary vertex. Figure from Ref. [10].	38
3.1	PbWO ₄ crystal with APD photodetectors attached. The inset shows a capsule with two APDs. Figure from Ref. [11].	41
3.2	Schematic of the ECAL laser monitoring system. Figure from Ref. [11].	44
3.3	Top: Relative ECAL response to laser light over the period 2011-2018, averaged in bins of pseudorapidity η . Bottom: LHC instantaneous luminosity which has increased during the years. Figure from Ref. [12].	45
3.4	Laser response over two weeks from October 14th 2017 to October 26th 2017, for a single crystal in the endcap (EE) at $\eta=+2.5$	46
3.5	Impact of more frequent ECAL transparency updates on the time stability of the di-electron invariant mass m_{ee} distribution for CMS data taken during a period of one week in July 2018 using $Z \rightarrow ee$ events. Both electrons are required to be in the ECAL barrel. The top plot displays the ratio of the median m_{ee} calculated with the twice-per-week (blue) transparency corrections and once-per-LHC-fill (red) corrections, to the median m_{ee} calculated with the full refined calibration (corrections applied every 40 minutes). The two blue vertical dashed lines indicate when the twice-per-week update was applied. The panel at the bottom shows the corresponding history of the LHC instantaneous luminosity during this period, with a clear LHC fill structure. The bottom plot displays the distribution of the ratio of the median m_{ee} calculated with the twice-per-week (blue) corrections and the once-per-LHC-fill (red) to the median m_{ee} calculated with the full refined calibration for this data-taking period.	48
3.6	Variation of the L1_SingleIsoEG28er2p5 trigger rate for a one-week period in July 2018. The rates are emulated for two transparency update frequencies (once-per-LHC-fill shown in red and twice-per-week shown in blue). The L1 electron/photon candidate are chosen to have a transverse energy greater than 28 GeV, pass an isolation requirement, and be within $ \eta < 2.5$	49
3.7	Rate vs PU plot for fill numbers 6956-6961. A fit for all the fills for seed L1_SingleIsoEG28er2p5 is presented for the once-per-LHC-fill (top) and the twice-per-week (bottom) updates.	50

- 3.8 The plot displays the average L1 electron/photon rate calculated with the twice-per-week (blue) corrections and the once-per-LHC-fill (red) corrections for CMS data taken during a period of one week in July 2018. The electron/photon candidates are required to have a transverse energy greater than 28 GeV, pass an isolation requirement, and be within $|\eta| < 2.5$. The same dataset was used to produce the history of the L1 electron/photon rate using two different sets of ECAL transparency corrections to demonstrate the impact of more frequent transparency updates on the stability of the L1 rate. The ECAL transparency corrections are updated either twice-per-week, emulating the update frequency that was applied at L1/HLT in 2018, or once-per-LHC-fill, emulating the L1/HLT update frequency that is applied from Run 3 onwards. 51
- 3.9 The plot displays the average total L1 rate calculated with the Twice-per-week (blue) corrections and the Once-per-LHC-fill (red) corrections for CMS data taken during a period of one week in July 2018. The ECAL transparency corrections are updated either twice-per-week, emulating the update frequency that was applied at L1/HLT in 2018, or once-per-LHC-fill, emulating the L1/HLT update frequency that is applied from Run 3 onwards. 52
- 3.10 The same dataset as used in Fig. 3.9 was also used to produce the profile distribution of the L1 rate using two different sets of ECAL transparency corrections to demonstrate the impact of more frequent transparency updates on the stability of the ECAL energy scale and resolution. The plot displays the distribution of the median of average rate calculated with the twice-per-week (blue) conditions and the once-per-LHC-fill (red). The left plot shows the profile distribution of the total rate (L1APhysics) and the right plot shows the profile distribution of electron/photon rate (L1.SingleIsoEG28er2p5). 53
- 3.11 Time stability of the rate of electrons reconstructed at HLT with $p_T > 30$ GeV per LHC fill, for CMS data taken during a period of one week in July 2018. The first set of corrections corresponds to the refined calibration performed in 2019, where the ECAL transparency corrections are updated every 40 minutes. For the other two sets, the ECAL transparency corrections are updated either twice-per-week, emulating the update frequency that was applied at L1/HLT in 2018, or once-per-LHC-fill, emulating the L1/HLT update frequency that is applied in Run 3 (2022 onwards). The plot displays the ratio of the HLT rates calculated with the twice-per-week (blue) conditions and the once-per-LHC-fill (red) to the rate calculated with the full refined calibration for this data-taking period. 54
- 3.12 The plots display the ratio between the L1 electron/photon E_T and the offline electron/photon E_T for the once-per-LHC-fill conditions (left) or the twice-per-week conditions (right) for CMS data taken during the 21st and the 22nd of July 2018 corresponding to four consecutive LHC fills 6956-6961. The same dataset was used to reconstruct L1 electron/photon candidates using two different sets of ECAL transparency corrections, to demonstrate the impact of more frequent transparency updates on the stability of the ECAL energy scale and resolution. The electrons are required to be within pseudo rapidity $|\eta| \leq 1.5$ 55

3.13	The p_T resolution obtained for electrons reconstructed at HLT with $p_T > 30$ GeV, for CMS data taken during a period of one week in July 2018. The plot displays the fractional difference between the HLT electron p_T calculated with the twice-per-week (blue) conditions or the once-per-LHC-fill (red) conditions, compared to the p_T calculated with the full refined calibration for this data-taking period.	56
3.14	The plot shows the latency of the insertion time of last tag (the processing of the corrections starts right after the completion of each laser sequence) and the prompt tag (the processing waits until one more laser sequence has been completed) in the DB for a one week period in September 2022.	56
3.15	The left plot shows the time taken for each step of the validation process to run relative to the start of each LHC fill for the Run 3 commissioning and data taking period between fill numbers 7650 to 8020. The right plot shows the total number of validation process during this period and the average time of each step. Before fill 7960, the laser data was collected after 2 hours from the start of a fill, later it was modified to be the start of stable beams. Thus an additional latency is observed after fill 7960. The longer latency peaks are usually caused by either ECAL not participating in the run or due to a delay in the Jenkins job.	57
4.1	Feynman diagram of the signal model following squark pair production. The decay arm can arise either from $\tilde{q}$ or from $\tilde{g} \rightarrow \tilde{q} + q$ states. $\tilde{\chi}_2^0$ is the Next-to Lightest Supersymmetric Particle which decays to a singlino-like Lightest Supersymmetric Particle, $\tilde{\chi}_1^0$, and the light CP-even neutral Higgs boson H_1	63
4.2	SM-like Higgs boson branching ratios for various SM particle final states as a function of the Higgs mass. Figure from Ref. [13]	66
5.1	An example of a DJR plot, generated by adjusting XQCUT and QCUT values in MC simulations. The plot illustrates the logarithm of the transverse momentum of the first additional jet in our dataset. Since this is a physical quantity, the distribution on the plot should be continuous.	71
5.2	These three plots in a row represent from left to right: the H_T distribution, the mass distribution of the two fat jets and the double-b-tag distribution of the two fat jets for m_{Squark} or m_{SUSY} equal 1200 GeV, $m_{H_1} = 60$ GeV.	72
5.3	These three plots in a row represent from left to right: the H_T distribution, the mass distribution of the two fat jets and the double-b-tag distribution of the two fat jets for m_{Squark} or m_{SUSY} equal 2400 GeV, $m_{H_1} = 60$ GeV.	73
5.4	These three plots in a row represent from left to right: the H_T distribution, the mass distribution of the two fat jets and the double-b-tag distribution of the two fat jets for m_{Squark} or m_{SUSY} equal 3200 GeV, $m_{H_1} = 60$ GeV.	73
5.5	These three plots in a row represent from left to right: the H_T distribution, the mass distribution of the two fat jets and the double-b-tag distribution of the two fat jets for m_{Squark} or m_{SUSY} equal 3200 GeV after the reweight of m_{SUSY} to m_{Squark} following the H_T distribution ratio.	74

6.1	The H_T distribution analyzed across diverse background and signal events. Every event fulfils the H_T trigger selection criteria.	81
6.2	The two-dimensional space of btagHbb (left) and ParticleNet tagger (right) is divided into the signal region, also called Tag Region (TR) in green, the Control Region (CR) in yellow, and the Validation Regions (VR) in red. These latter are designated to estimate background events.	83
6.3	The double-b-tag discriminator values distribution for the two chosen AK8 jets for the signal sample with $m_{H_1} = 80$ GeV and $m_{SUSY} = 2800$ GeV, comparing btagHbb algorithm (A) and ParticleNet (B).	84
6.4	The double-b-tag discriminator values distribution for the two chosen AK8 jets for QCD background events, comparing btagHbb algorithm (A) and ParticleNet (B). The empty bin in the ParticleNet distribution is due to the limited number of events in the Monte Carlo sample.	84
6.5	The double-b-tag discriminator values distribution for the two chosen AK8 jets for $t\bar{t}$ background events, comparing btagHbb algorithm (A) and ParticleNet (B).	85
6.6	Comparison of the mass reconstructed by the Softdrop and ParticleNet algorithm in a signal sample with $m_{H_1} = 80$ GeV, $m_{SUSY} = 2800$ GeV. . . .	86
6.7	Scheme of the mass regions in the 2D jet mass plane. Signal mass regions are labelled as S_i , while sideband mass regions are identified as U_i	86
6.8	Distribution of the reconstructed jet mass for the two selected AK8 jets across various m_{H_1} masses and before and after the double-b-tagger selection, for the Softdrop (left) and the ParticleNet algorithm (right). . . .	88
6.9	The distribution of the reconstructed jet mass for the two selected AK8 jets in the QCD background for the combination of btagHbb and Softdrop (left) and ParticleNet algorithms (right).	89
6.10	The distribution of the reconstructed jet mass for the two selected AK8 jets in the $t\bar{t}$ background for the combination of btagHbb and Softdrop (left) and ParticleNet algorithms (right).	89
6.11	The distribution of the reconstructed jet mass for the two selected AK8 jets in the W+jets (A and B) and Z+jets (C and D) for the combination of btagHbb and Softdrop (left) and ParticleNet algorithms (right).	90
6.12	Yields in the signal mass regions after applying event selections with btagHbb and Softdrop algorithms (left) and the ParticleNet algorithms (right). The plotted signal and background yields are based on estimates derived from simulated events. The hatched bands in the figures denote the statistical uncertainty associated with the total background, while systematic uncertainties are not shown.	91
7.1	Predicted and Observed yields in the TR (left) and VR (right) regions for the QCD MC background. The shaded bands and error bars show the statistical uncertainties of predicted and observed values, respectively. . . .	94
7.2	Predicted and observed yields in the VR region using the actual yields in Run 2 data from 2018. $H_T \geq 2000$ GeV is applied. The shaded bands and error bars show the statistical uncertainties of predicted and observed values, respectively.	95
7.3	Scheme of the new mass regions in the 2D jet mass plane (left). ParticleNet, signal sample with $m_{H_1} = 80$ GeV, $m_{SUSY} = 2800$ GeV, after double-b-tag selection with new grid (right).	97

7.4	$t\bar{t}$ contributions to the fat jet mass distributions. The plot is normalized so that the area under each histogram integrates to 1. The $t\bar{t}$ dileptonic decay (A) and $t\bar{t}$ hadronic decay (B) are divided into four different contributions which are the combination of two different conditions: with or without B hadrons, with $(300 < p_T \leq 430)$ GeV, or $p_T > 430$ GeV. The latter is the approximate threshold for a top quark decay to be fully merged inside an AK8 jet.	98
7.5	The fraction of signal vs the fraction of $t\bar{t}$ with at least one electron at the generator level removed by each selection. In this particular instance, the NMSSM sample being plotted corresponds to $m_{\text{SUSY}} = 2800$ GeV and $m_{H_1} = 80$ GeV.	100
7.6	The distribution of $\min_e(\min_{A,B} \Delta R_{j,e})$ for NMSSM and $t\bar{t}$ events after applying the different selections.	101
7.7	Significance values vs $\min_e(\min_{A,B} \Delta R_{j,e})$ for different selections studied for the electron case.	102
7.8	The fraction of signal vs the fraction of $t\bar{t}$ with at least one electron at the generator level removed by each selection after applying the $\min_e(\min_{A,B} \Delta R_{j,e})$ cuts.	103
7.9	The fraction of signal vs the fraction of $t\bar{t}$ with muons at the generator level removed by each selection. In plot A, both the ID cuts and the MVA selections are shown, whereas plot B exclusively shows the MVA selections.	104
7.10	Plot A is obtained after applying the MVA.L selection. There is no obvious tail in this plot. There is no clear difference in $\min_\mu(\min_{A,B} \Delta R_{j,\mu})$ between signal and background. Plot B shows the significance curve vs $\min_\mu(\min_{A,B} \Delta R_{j,\mu})$, illustrating that there is no obvious peak in this case. . .	104
7.11	Bar graph showing the significance values for all the studied selections for the muon case.	105
7.12	The fraction of signal vs the fraction of $t\bar{t}$ with at least one tau removed by each selection. In this particular instance, the NMSSM sample being plotted corresponds to $m_{\text{SUSY}} = 2800$ GeV and $m_{H_1} = 80$ GeV.	106
7.13	Significance values vs $\min_\tau(\min_{A,B} \Delta R_{j,\tau})$ for different selections studied for the τ case.	107
7.14	The fraction of signal vs the fraction of $t\bar{t}$ with at least one tau removed by each selection, after applying the $\min_\tau(\min_{A,B} \Delta R_{j,\tau})$ cuts.	108
7.15	The 1D discriminant spans a range from 0 to 1, while the 2D discriminant, derived from the sum of two fat jets, ranges from 0 to 2. Figure (A) shows the distribution of $t\bar{t}$ background for both variables, while Fig. (B) illustrates the distribution of a signal sample with a Higgs boson mass of $m_{H_1} = 125$ GeV and $m_{\text{SUSY}} = 2800$ GeV for both variables.	109
7.16	Significance vs the 2D discriminant scan for a signal mass of 30 GeV for various 1D-cut discriminant values (coloured curved).	110
7.17	Significance vs the 2D discriminant scan for a signal mass of 80 GeV for various 1D-cut discriminant values (coloured curved).	110
7.18	Significance vs the 2D discriminant scan for a signal mass of 125 GeV for various 1D-cut discriminant values (coloured curved).	111

7.19	H_T distribution of 2018 data, QCD and $t\bar{t}$ MC background events in the VR.	115
7.20	Observed and predicted yields in the TR for each of the 30 search regions, aggregated across the three years of data collection (2016-2018). The multijet background (QCD) is estimated from data while other backgrounds are simulated. Example signal distributions are provided for $m_{H_1} = 70$ GeV and $m_{SUSY} = 1200, 2000, \text{ and } 2800$ GeV. Error bars indicate statistical uncertainties, while the hatched bands represent systematic uncertainties. Figure from Ref. [14]	117
8.1	One-quarter of the Phase-2 CMS tracker in the $r - z$ plane. The inner tracker is represented in green and yellow, with varying pixel sizes across different modules, with smaller pixels positioned closer to the interaction point. The outer tracker is depicted in blue and red. The blue sections correspond to pixel strip modules, utilizing both pixel and strip detectors. The red sections represent two strip module detectors that solely utilize strip detectors [8].	123
8.2	An example of the p_T module concept. Low- p_T particles can be rejected by correlating signals from closely spaced sensors; the green-coloured channels indicate the selection window that defines an acceptable stub. [8]124	
8.3	Functional diagram of the CMS L1 Phase-2 upgraded trigger design. The Phase-2 L1 trigger integrates inputs from the calorimeters, muon spectrometers, and the track finder. Calorimeter trigger inputs encompass data from the barrel calorimeter (BC), high-granularity calorimeter (HGCAL), and hadron forward calorimeter (HF). This comprises a barrel calorimeter trigger (BCT) and a global calorimeter trigger (GCT). The muon trigger receives inputs from detectors including drift tubes (DT), resistive plate chambers (RPC), cathode strip chambers (CSC), and gas electron multipliers (GEM). It consists of a barrel layer-1 processor and muon track finders, processing data from three distinct pseudorapidity regions-BMTF, OMTF, and EMTF for barrel, overlap, and endcap, respectively. Muon track finders transmit muon candidates to the global muon trigger (GMT), allowing combination with tracking information. The track finder (TF) provides tracks to various components, including the global track trigger (GTT). The correlator trigger (CT) at the centre (yellow area) has two layers dedicated to particle-flow reconstruction. All objects feed into the global trigger (GT) for the final L1 trigger decision. External triggers, such as potential inputs from the MTD, are also depicted. Dashed lines represent links with potential exploitation. Components under development within the Phase-2 L1 trigger project are grouped in the light blue area. Various processing levels are indicated on the right: trigger primitives (TP), local and global trigger reconstruction, particle-flow trigger reconstruction (PF), and global decision. . . .	126
9.1	Sketch of the transverse impact parameters (d_0) and longitudinal impact parameters (z_0) of a track inside the detector. Figure from Ref. [15]. . . .	130
9.2	Visual description of the simplest implementation of the Peak Finding Algorithm.	132
9.3	Image representing two steps of the PFA algorithm in a realistic event. .	133

- 9.4 The light blue distribution with $\phi > 0.1$ represents the events where at least 10% of tracks originate from the true MC vertex. Conversely, the surrounding data depicted in red denotes noise, often stemming from events with significant pileup occurrences ($\phi < 0.1$), indicating that less than 10% of the tracks used for reconstruction are sourced from the true MC vertex. 134
- 9.5 Log-scale plot illustrating the distribution of $Z_{\text{reco}} - Z_{\text{true}}$ for events with fixed $\delta = 0.05$ cm and $\sigma = 0.2$ cm. The distribution is characterized by two distinct Gaussian fits. A narrower distribution corresponds to events with a successful reconstruction, juxtaposed with a broader distribution encompassing events falling below, reflecting varying degrees of reconstruction quality. 135
- 9.6 The impact of varying δ is explored while maintaining σ at 0.15 cm. The algorithm's performance is inherently tied to the value of δ , serving as a limitation on achievable resolution. Decreasing δ initially improves resolution until reaching the experimental resolution limit. This plot shows that beyond $\delta = 0.05$, the gains in resolution become insignificant. . . . 136
- 9.7 Exploring the impact of varying σ while setting the step size to 0.05 cm. The plot depicts the behaviour of the algorithm with σ assuming three different values. Interestingly, the analysis suggests an optimal range for σ between 0.05 cm and 0.30 cm, with 0.15 cm the best performance among the three values. 136
- 9.8 Bi-dimensional scan of the algorithm across various σ values and $Z_{\text{reco}} - Z_{\text{true}}$ ranges at fixed $\delta = 0.05$ cm. The colour scale illustrates the fraction of reconstructed events within different ranges. The plot on the right (B) is a zoom of the one on the left (A), to show more granularity between $Z_{\text{reco}} - Z_{\text{true}} \leq 0.3$ and $Z_{\text{reco}} - Z_{\text{true}} \leq 1$ 137
- 9.9 In Figure (A), the algorithm's efficiency varies with σ , peaking at 0.22 cm of 0.96. The plot is generated with $Z_{\text{reco}} - Z_{\text{true}}$ fixed at 0.5 cm. The plot on the right side (B) fixes σ at 0.22 cm, displaying varying efficiency for different $Z_{\text{reco}} - Z_{\text{true}}$ ranges. Beyond 0.5 cm, the curve nearly plateaus, indicating marginal efficiency improvements attributed to the addition of low-quality reconstructed vertices with numerous tracks from the pileup. 137
- 9.10 The Z_0 track parameter resolutions for muons in $t\bar{t}$ + pileup events as a function of the muon p_T [8]. 138
- 9.11 The z_0 resolution for tracks in $t\bar{t}$ events, with an average pileup of 200, is depicted as a function of $|\eta|$. The z_0 resolutions are represented by intervals encompassing 68% (indicated by filled black markers) or 90% (represented by open red markers) of all tracks with $p_T > 2$ GeV [16]. . 139
- 9.12 The figures illustrate a simplified parabolic fit representing the track Z_0 resolution of the 68% resolution curve on the left and of the 90% resolution curve on the right. 140
- 9.13 Algorithm's efficiency in three different cases for $t\bar{t}$ sample. The old algorithm in red is compared with the one with the η variable implementation following the 68% resolution in blue and the 90% resolution in green. 141

9.14	The plot on the left (A) and the plot on the right-hand side (B) represent $t\bar{t}$ residual in the three different cases: PFA without η contribution (in red), η contribution following the curve at 68% resolution (blue) and 90% resolution (green). The left and the right plot are based on the same data. The left plot uses a log scale for the y-axis with an x-axis range of -20 to 20 cm. The right plot uses a linear y-axis scale with an x-axis range of -0.5 to 0.5 cm.	141
9.15	Residual plot (A) and efficiency plot (B) of the algorithm, using the QCD 30-50 GeV sample. A comparison between the original PFA and the one using the η distribution at 68% resolution is shown.	142
9.16	Residual plot (A) and efficiency plot (B) of the algorithm, using the QCD 50-80 GeV sample. A comparison between the original PFA and the one using the η distribution at 68% resolution is shown.	142
9.17	Residual plot (A) and efficiency plot (B) of the algorithm, using the QCD 80-120 GeV sample. A comparison between the original PFA and the one using the η distribution at 68% resolution is shown.	143
9.18	Residual plot (A) and efficiency plot (B) of the algorithm, using the QCD 120-170 GeV sample. A comparison between the original PFA and the one using the η distribution at 68% resolution is shown.	143
9.19	Residual plot (A) and efficiency plot (B) of the algorithm, using the QCD 170-300 GeV sample. A comparison between the original PFA and the one using the η distribution at 68% resolution is shown.	144
9.20	Normalized representation of the η of the Jets associated with the PV for three different samples: QCD 30-50 GeV, QCD 80-120 GeV and $t\bar{t}$ sample.	145
9.21	Normalized representation of the η of the Jets associated with the PV for different samples: QCD 30-50 GeV, QCD 50-80 GeV, QCD 80-120 GeV, QCD 120-170 GeV and $t\bar{t}$ sample.	145
9.22	Residuals, showing the resolution of the algorithm with η for different samples: QCD 30-50 GeV, QCD 50-80 GeV, QCD 80-120 GeV, QCD 120-170 GeV and $t\bar{t}$ sample.	146
9.23	In Fig. (A) and (B) the sum of the transverse momentum ($\sum p_T$) of tracks reconstructing the second peak vs the first peak is shown. Figure (A) represents the case in which the first peak is closer to the true MC PV than the second peak. While Fig. (B) describes the opposite scenario. The same concept is depicted in Figures (C) and (D) for the average p_T quantity.	148
9.24	In Figs. (A) and (B) the multiplicity of tracks reconstructing the second peak vs the first peak is shown. Figure (A) represents the case in which the first peak is closer to the true MC PV than the second peak. While Fig. (B) describes the opposite scenario. The same concept is depicted in Figs. (C) and (D) for the average standard deviation quantity.	149
9.25	The sum of transverse momenta ($\sum p_T$), average transverse momentum (p_T), multiplicity, and standard deviation are depicted for the best reconstructable peak compared to the first one.	150

9.26	The four plots represent the ratio between four different quantities (sum of transverse momenta (p_T), average p_T , standard deviation, and multiplicity) calculated for the first and the second peak. The blue curve showcases situations where the first peak is closer to the MC PV than the second peak, while the red curve illustrates cases where the second peak is closer than the first to the true MC primary vertex.	151
9.27	After implementing the described cuts and swaps, the algorithm's efficiency demonstrates an enhancement of approximately 0.7%.	152
9.28	The FastHisto histogram is shown for a single $t\bar{t}$ event with 200 PU, illustrating the Z distribution in 256 bins, weighted by p_T for each track. Different track types are distinguished by colours: blue represents evenly spread PU tracks, red denotes randomly distributed fake tracks, and green signifies primary vertex tracks clustered around $z_0 = 2$ cm, corresponding to the true vertex location (indicated by the purple dashed line). The reconstructed vertex by FastHisto is depicted in black, closely aligned with the true vertex, signifying a well-reconstructed vertex. The success of this reconstruction can be attributed to the high p_T peak at the vertex, with clearly visible high p_T PV tracks, simplifying the vertex reconstruction process for FastHisto [17].	153
9.29	Residual plot on the left and efficiency plot on the right, comparing the PFA and the FastHisto algorithm using the QCD 30-50 GeV sample. . . .	154
9.30	Residual plot on the left and efficiency plot on the right, comparing the PFA and the FastHisto algorithm using the QCD 50-80 GeV sample. . . .	155
9.31	Residual plot on the left and efficiency plot on the right, comparing the PFA and the FastHisto algorithm using the QCD 80-120 GeV sample. . . .	155
9.32	Residual plot on the left and efficiency plot on the right, comparing the PFA and the FastHisto algorithm using the QCD 120-170 GeV sample. . .	156
9.33	Residual plot on the left and efficiency plot on the right, comparing the PFA and the FastHisto algorithm using the $t\bar{t}$ sample.	156
9.34	Illustration of the L1 Track Trigger system. Track Finder system, GTT, Correlator and GT are presented.	160
9.35	Chain scheme used to implement the PFA algorithm on FPGA.	161
9.36	Scheme of a possible Router, several workers and a final Combiner. . . .	162
9.37	Residual plot on the left and efficiency plot on the right, comparing the PFA algorithm, its version implemented on FPGA (with a different bin size of 0.94 mm) and the FastHisto algorithm using the QCD 30-50 GeV sample.	164
9.38	Residual plot on the left and efficiency plot on the right, comparing the PFA algorithm, its version implemented on FPGA (with a different bin size of 0.94 mm) and the FastHisto algorithm using the QCD 50-80 GeV sample.	164
9.39	Residual plot on the left and efficiency plot on the right, comparing the PFA algorithm, its version implemented on FPGA (with a different bin size of 0.94 mm) and the FastHisto algorithm using the QCD 80-120 GeV sample.	165

9.40	Residual plot on the left and efficiency plot on the right, comparing the PFA algorithm, its version implemented on FPGA (with a different bin size of 0.94 mm) and the FastHisto algorithm using the QCD 120-170 GeV sample.	165
9.41	Residual plot on the left and efficiency plot on the right, comparing the PFA algorithm, its version implemented on FPGA (with a different bin size of 0.94 mm) and the FastHisto algorithm using the $t\bar{t}$ sample.	166
Appendix A.1	DJR plots for $m_{\tilde{g}} = (1 + 1\%)m_{SUSY}$ where $m_{SUSY} = 800$ GeV, XQCUT=60, QCUT varying	173
Appendix A.2	DJR plots for $m_{\tilde{g}} = (1 + 1\%)m_{SUSY}$ where $m_{SUSY} = 2000$ GeV, XQCUT=60, QCUT varying	174
Appendix A.3	DJR plots for $m_{\tilde{g}} = (1 + 1\%)m_{SUSY}$ where $m_{SUSY} = 2800$ GeV, XQCUT=60, QCUT varying	174
Appendix A.4	DJR plots for gluino decoupled, $m_{Squark} = 800$ GeV, XQCUT=60, QCUT varying	175
Appendix A.5	DJR plots for gluino decoupled $m_{Squark} = 2400$ GeV, XQCUT=60, QCUT varying	175
Appendix A.6	DJR plots for gluino decoupled, $m_{Squark} = 3200$ GeV, XQCUT=60, QCUT varying	175
Appendix B.1	In Fig. (A) and (B) the sum of the transverse momentum ($\sum p_T$) of tracks reconstructing the second peak vs the first peak is shown. Figure (A) represents the case in which the first peak is closer to the true MC PV than the second peak. While Fig. (B) describes the opposite scenario. The same concept is depicted in Figures (C) and (D) for the average p_T quantity.	178
Appendix B.2	In Fig. (A) and (B) the multiplicity of tracks reconstructing the second peak vs the first peak is shown. Figure (A) represents the case in which the first peak is closer to the true MC PV than the second peak. While Fig. (B) describes the opposite scenario. The same concept is depicted in Figures (C) and (D) for the average standard deviation quantity.	179
Appendix B.3	The sum of transverse momenta ($\sum p_T$), average transverse momentum (p_T), multiplicity, and standard deviation are depicted for the best reconstructable peak compared to the first one.	180
Appendix B.4	The four plots represent the ratio between four different quantities (sum of transverse momenta (p_T), average p_T , standard deviation, and multiplicity) calculated for the first and the second peak. The blue curve showcases situations where the first peak is closer to the MC PV than the second peak, while the red curve illustrates cases where the second peak is closer than the first to the true MC primary vertex.	181
Appendix B.5	Efficiency curve before and after implementing the described cuts and swaps described in Section 9.1.3.	182

Appendix B.6	In Fig. (A) and (B) the sum of the transverse momentum ($\sum p_T$) of tracks reconstructing the second peak vs the first peak is shown. Figure (A) represents the case in which the first peak is closer to the true MC PV than the second peak. While Fig. (B) describes the opposite scenario. The same concept is depicted in Figures (C) and (D) for the average p_T quantity.	183
Appendix B.7	In Fig. (A) and (B) the multiplicity of tracks reconstructing the second peak vs the first peak is shown. Figure (A) represents the case in which the first peak is closer to the true MC PV than the second peak. While Fig. (B) describes the opposite scenario. The same concept is depicted in Figures (C) and (D) for the average standard deviation quantity.	184
Appendix B.8	The sum of transverse momenta ($\sum p_T$), average transverse momentum (p_T), multiplicity, and standard deviation are depicted for the best reconstructable peak compared to the first one.	185
Appendix B.9	The four plots represent the ratio between four different quantities (sum of transverse momenta (p_T), average p_T , standard deviation, and multiplicity) calculated for the first and the second peak. The blue curve showcases situations where the first peak is closer to the MC PV than the second peak, while the red curve illustrates cases where the second peak is closer than the first to the true MC primary vertex.	186
Appendix B.10	Efficiency curve before and after implementing the described cuts and swaps described in Section 9.1.3.	187

List of Tables

1.1	The Standard Model particles representations and their relative dimensions and charge.	6
1.2	Chiral superfields within the MSSM consist of various multiplets. The initial five rows represent the standard multiplets of SM, while the last two rows depict the two Higgs doublets along with its supersymmetric partner. The final column provides information on the representation of the gauge group to which each multiplet belongs.	16
1.3	Vector superfields of the MSSM. Within these superfields, the third column designates the Standard Model gauge bosons, and the second column represents their respective supersymmetric partners. The last column provides information about the representation of the multiplet within the gauge group.	16
1.4	Mass eigenstates of the MSSM compared with the NMSSM [18][19]. . . .	20
3.1	The mean and RMS of each of the gaussian distributions from Fig. 3.12. .	52
4.1	Comparison between published analysis and ongoing analysis. Secondary backgrounds are estimated from MC samples. *Plan to consider also rare backgrounds such as ttH, ttZ, ZZ.	63
4.2	Inclusive cross sections for $pp \rightarrow \tilde{q}\tilde{q}, \tilde{q}\tilde{q}, \tilde{q}\tilde{g}, \tilde{g}\tilde{g}$ as a function of m_{SUSY} with uncertainties. The cross sections are calculated using NNLL-fast 1.1 [20] for squark masses and gluino masses set to 1% larger than m_{SUSY}	64
4.3	Branching fraction of $H_1 \rightarrow b\bar{b}$ for different values of m_{H_1} . The branching fractions were computed utilizing the HDECAY software [21] [22]. . . .	65
5.1	Cross sections for $pp \rightarrow \tilde{q}\tilde{q}$ as a function of m_{Squark} with uncertainties. The cross sections are calculated using NNLL-fast 1.1 [20].	68
5.2	Comparison of QCUT values resulting in smooth DJR curves for two distinct scenarios. The first scenario involves $\tilde{q}\tilde{q}, \tilde{q}\tilde{q}, \tilde{q}\tilde{g}, \tilde{g}\tilde{g}$ processes, while the second involves the $\tilde{q}\tilde{q}$ production.	71
7.1	Decay modes of W^+ with relative branching fractions and uncertainties. W^- has the charge conjugate of these modes [23].	97
7.2	This table showcases the top ten values derived from the average, arranged in descending order of percentage of maximum significance. The 1D and 2D cuts are presented along with corresponding fractions for signal, QCD, and $t\bar{t}$	111
9.1	Resource utilization of the PFA algorithm.	163
9.2	Resource utilization of the FastHisto algorithm [17].	163

To the 15-year-old me who first fell in love with CERN. To all the past versions of myself who fueled this journey with passion and determination and to the present me. Thank you for embracing the good moments and finding forgiveness during the challenging ones.
To the future me: keep going, girl.

Part I

Theory and Experimental Layout

Chapter 1

Theoretical framework

1.1 The Standard Model

The Standard Model (SM) is a gauge field theory which represents the most advanced effort to classify and describe the elementary particles known so far and the interactions between them. The Standard Model was built during the 60's and 70's from a theoretical point of view and has been experimentally tested since then, showing its extremely powerful predicting nature. All phenomena from the scale of $\mathcal{O}(1eV)$ to $\mathcal{O}(100GeV)$ are well described by the theory, which takes into account three of the four fundamental forces of nature: the strong, electromagnetic, and weak forces, respectively. Gravitational effects appear to be on an energy scale significantly different compared with the other forces and they are currently not described by the SM.

The particles included in the SM are grouped into two different classes:

- Fermions, which are half-integer spin particles and matter components;
- Bosons, which are integer spin particles and mediate interactions between the fermions.

All the particles of the SM are listed in Fig. 1.1. The three first columns represent the families of fermions. The particles of the second and third columns can decay to the stable particles of the first column. In the next three columns, the anti-particles are shown, those have opposite internal (independent from space-time coordinates) quantum numbers compared with particles. In the last two columns, the bosons are presented. As shown in Fig. 1.1, the quark can exist in three different configurations of colours, while the anti-quarks exist in three different configurations of anti-colours. Gluons have a combination of colour charges in a superposition of states.

Quarks are experimentally detected as colour-neutral combinations, forming either mesons (comprising two quarks) or baryons (comprising three quarks). This phenomenon occurs because Quantum Chromodynamics (QCD), the theory describing the

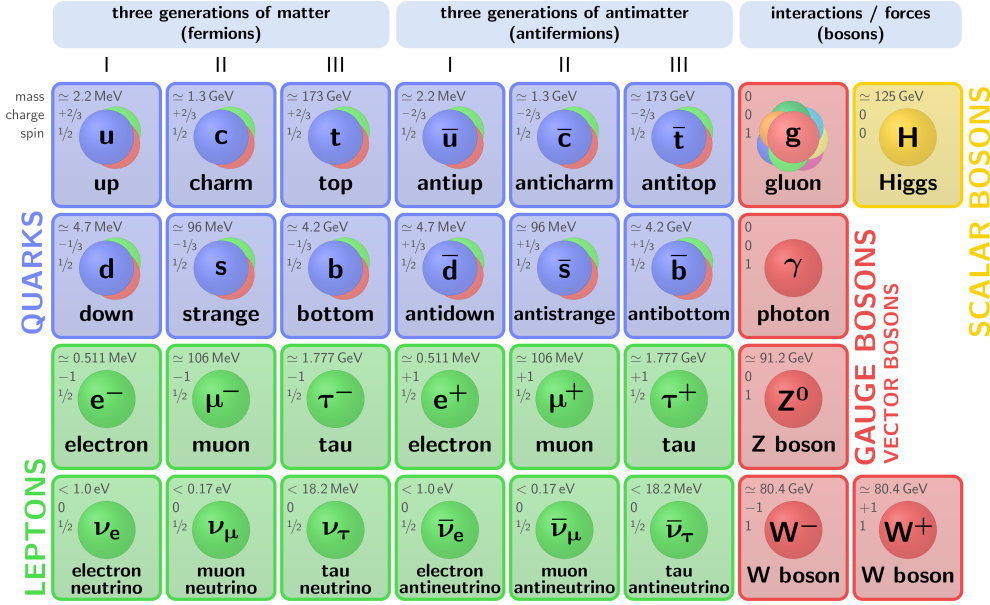


FIGURE 1.1: Scheme of the Standard Model constituents. Figure from Ref. [1].

strong interaction, features a coupling constant α_s that becomes small at high energies. This leads to asymptotic freedom, where quarks and gluons interact weakly and move almost freely. Conversely, at low energies, α_s increases, causing quarks and gluons to be strongly bound together, a phenomenon known as confinement. This results in hadronization, where quarks and gluons form hadrons such as baryons and mesons. In high-energy collisions, the strong force causes the formation of jets-collimated streams of particles that emerge from the fragmentation of high-energy quarks and gluons (jets will be better described in Section 2.3.2).

The SM Lagrangian is invariant under local gauge transformations of the fields with respect to three symmetry groups. The $SU(3)_C$ group, for the strong interactions, and the $SU(2)_W \otimes U(1)_Y$ groups, for the electroweak interactions. Here the subscripts C, W, Y stand for colour charge, weak isospin, and weak hypercharge, respectively.

The final symmetry group is given by $G_{SM} = SU(3)_C \otimes SU(2)_W \otimes U(1)_Y$.

1.1.1 The SM Lagrangian

The dynamics of the SM can be captured by the following Lagrangian density:

$$\mathcal{L}_{\text{SM}} = \mathcal{L}_{\text{gauge}} + \mathcal{L}_{\text{fermion}} + \mathcal{L}_{\text{Higgs}} + \mathcal{L}_{\text{Yukawa}}$$

where: $\mathcal{L}_{\text{gauge}}$ is the gauge boson term, $\mathcal{L}_{\text{fermion}}$ is the fermion term, $\mathcal{L}_{\text{Higgs}}$ is the Higgs term and $\mathcal{L}_{\text{Yukawa}}$ is the Yukawa term.

Gauge Boson Term

The term

$$\mathcal{L}_{\text{gauge}} = -\frac{1}{4}F_a^{\mu\nu}F_{\mu\nu}^a \quad (1.1)$$

describes the kinetic energy and interaction of gauge bosons. $F_a^{\mu\nu}$ represents the field strength tensor defined as:

$$F_{\mu\nu} = \partial_\mu F_\nu - \partial_\nu F_\mu \pm ig[F_\mu, F_\nu]. \quad (1.2)$$

The index a in Eq. 1.1 runs over the three different types of vector gauge bosons: G , W and B .

$$G_{\mu\nu}^a = \partial_\mu G_\nu^a - \partial_\nu G_\mu^a - g_s f_{abc} G_\mu^b G_\nu^c, \quad (1.3)$$

$$W_{\mu\nu}^i = \partial_\mu W_\nu^i - \partial_\nu W_\mu^i - g_W \epsilon^{ijk} W_\mu^j W_\nu^k, \quad (1.4)$$

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu. \quad (1.5)$$

The eight gluon boson fields, G^a , are linked to the $SU(3)_C$ generators and mediate the strong force. On the other hand, the three weak interaction boson fields W_i , related to the $SU(2)_W$ generators, and the hypercharge boson field B , tied to the $U(1)_Y$ generator, are not observable physical particles.

D_μ is the covariant derivative that includes the gauge boson interactions: $D_\mu = \partial_\mu + ig_s G_\mu^a L^a + ig_W W_\mu^b T^b + ig_Y B_\mu Y$ where L^a represents the generators of the $SU(3)_C$ group for the respective representations, being $\frac{1}{2}\lambda^a$ for triplets (with λ^a denoting the 3×3 Gell-Mann matrices) and 0 for singlets; T^b signifies the generators of the $SU(2)_W$ group, where $\frac{1}{2}\sigma^b$ is for doublets (with σ^b being the 2×2 Pauli matrices) and again 0 for singlets; Y is the hypercharge. This term ensures the gauge invariance of the Lagrangian under local gauge transformations.

Notably, this term not only contains kinetic terms for gauge fields, but also cubic and quartic interaction terms, coming from the $[F_\mu, F_\nu]$ commutator and corresponding to interaction vertices with three and four gauge bosons. Since $U(1)_Y$ is an abelian group, B bosons do not self-interact.

Fermion term

The fermion fields come from five distinct representations of the G_{SM} group:

$$Q_{Li}(3,2)_{+1/6}, \quad u_{Ri}(3,1)_{+2/3}, \quad d_{Ri}(3,1)_{-1/3}, \quad L_{Li}(1,2)_{-1/2}, \quad \ell_{Ri}(1,1)_{-1}. \quad (1.6)$$

Here, the index i denotes one of the three generations, and L or R refer to left-handed and right-handed chiralities, respectively. Following each field in this notation, there are two numbers in brackets, indicating the representation dimensions of the $SU(3)_C$ and $SU(2)_W$ groups, succeeded by a subscript indicating the associated hypercharge. Decomposing the $SU(2)_W$ doublets, one finds:

$$Q_{Li} = \begin{pmatrix} u_{Li} \\ d_{Li} \end{pmatrix}, \quad L_{Li} = \begin{pmatrix} \nu_{Li} \\ e_{Li} \end{pmatrix}. \quad (1.7)$$

A summary of these representations can be found in Tab. 1.1. The electromagnetic charge, $U(1)_{em}$, is determined by the relation $Q = W + \frac{1}{2}Y$, where W symbolizes the weak isospin and Y refers to the hypercharge.

Field	$SU(3)_C$	$SU(2)_W$	$U(1)_Y$	$U(1)_{em}$
$Q_{Li}^I = \begin{bmatrix} u_{Li}^I \\ d_{Li}^I \end{bmatrix}$	3	2 ($\pm 1/2$)	1/6	$\begin{bmatrix} 2/3 \\ -1/3 \end{bmatrix}$
u_{Ri}^I	3	1	2/3	2/3
d_{Ri}^I	3	1	-1/3	-1/3
$L_{Li}^I = \begin{bmatrix} \nu_{Li}^I \\ e_{Li}^I \end{bmatrix}$	1	2 ($\pm 1/2$)	-1/2	$\begin{bmatrix} 0 \\ -1 \end{bmatrix}$
ℓ_{Ri}^I	1	1	-1	-1

TABLE 1.1: The Standard Model particles representations and their relative dimensions and charge.

Particles that are singlets under a given representation are neutral to the corresponding force and do not couple to their gauge bosons. The categorisation can be elaborated as follows:

- **quarks** (Q_{Li}, u_{Ri}, d_{Ri}) are constituents of $SU(3)_C$ triplets, thereby participating in strong interactions;
- **leptons** (L_{Li}, e_{Ri}) do not carry colour charge and are thus $SU(3)_C$ singlets, being inert to the strong force.

Furthermore, these particles can also be distinguished by their chirality:

- **left-handed** entities (Q_{Li}, L_{Li}) form $SU(2)_W$ doublets and are involved in interactions mediated by W^i bosons;
- **right-handed** counterparts $(u_{Ri}, d_{Ri}^I, e_{Ri})$ are $SU(2)_W$ singlets, and as such, they do not interact with W^i bosons.

As Tab. 1.1 suggests, these representations also possess distinct hypercharges, which determine how they transform under $U(1)_Y$. By examining the organization of these multiplets, we gain insights into how particles interact. For instance: left-handed quarks engage in both strong interactions (as $SU(3)$ triplets) and weak interactions (as $SU(2)$ doublets); right-handed quarks do not interact with W bosons (as $SU(2)$ singlets); it's noteworthy that the weak interaction violates parity, as it exclusively interacts with left-handed particles.

Finally, the fermion term in the SM Lagrangian can be written as

$$\mathcal{L}_{\text{fermion}} = \sum_{i=1}^3 Q_i^\dagger \sigma^\mu D_\mu Q_i + \bar{u}_i^\dagger \sigma^\mu D_\mu \bar{u}_i + \bar{d}_i^\dagger \sigma^\mu D_\mu \bar{d}_i + L_i^\dagger \sigma^\mu \mathcal{D}_\mu L_i + \bar{e}_i^\dagger \sigma^\mu D_\mu \bar{e}_i + \text{h.c.} \quad (1.8)$$

Higgs term

Moreover, adding to the Lagrangian a mass term for fermions or for the weak force carriers (which experimentally are massive particles) creates an explicit breaking of the gauge symmetry G_{SM} . It is therefore necessary to introduce a new mechanism which allows these particles to acquire a mass without an explicit breaking. This is done by adding a complex scalar field ϕ whose Lagrangian is given by:

$$\mathcal{L}_{\text{Higgs}} = (D_\mu \phi)^\dagger (D^\mu \phi) - V(\phi) \quad (1.9)$$

where the potential $V(\phi)$ is expressed as

$$V(\phi^\dagger \phi) = -\mu^2(\phi^\dagger \phi) + \lambda(\phi^\dagger \phi)^2. \quad (1.10)$$

The shape of the potential, shown in Fig. 1.2, is invariant under rotations, however, the minimum is not unique and, since the minimum of the potential is interpreted as the ground state of the theory, the field acquires a vacuum expectation value: $v = \frac{\mu}{\sqrt{2\lambda}}$. The weak isospin doublet can be written as

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \begin{pmatrix} \frac{\pi_1 + i\pi_2}{\sqrt{2}} \\ \frac{\sigma + i\pi_3}{\sqrt{2}} \end{pmatrix}, \quad (1.11)$$

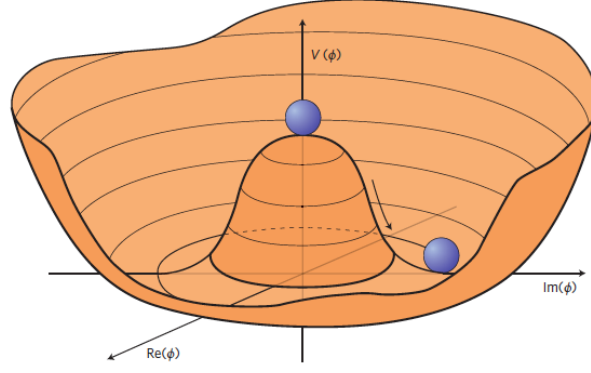


FIGURE 1.2: Shape of the Higgs potential with spontaneous symmetry breaking. Figure from Ref. [2].

where the + and 0 stand for the electric charge and σ and π_i ($i = 1, 2, 3$) are real fields. It is possible to write the σ field as

$$\sigma(x) = \tilde{\sigma}(x) + v, \quad (1.12)$$

in such a way that $\tilde{\sigma}$ represents the fluctuations of the field in the radial direction, with x indicating the space-time coordinates. These fluctuations correspond to the physical Higgs boson. The π_i fields are instead massless spin-0 states called Goldstone bosons, whose presence in the theory is related to spontaneous symmetry breaking, as stated by the Goldstone theorem [24]. The charged components of the Higgs field become the longitudinal components of the W^+ and W^- bosons, while the neutral component of the Higgs field, becomes the longitudinal component of the Z boson.

Expanding the first term of the Eq. 1.9 the following expression is derived:

$$\mathcal{L}_{mass} = g_w^2 \frac{v^2}{8} [(W_\mu^1)^2 + (W_\mu^2)^2] + \frac{v^2}{8} (g_w W_\mu^3 - g_y B_\mu)^2. \quad (1.13)$$

These are mass terms related to the four fields. Defining the physical fields in such a way that:

$$W_\mu^\pm = \frac{W_\mu^1 \pm i W_\mu^2}{\sqrt{2}}, \quad (1.14)$$

$$Z_\mu = \frac{g_w W_\mu^3 - g_y B_\mu}{\sqrt{g_w^2 + g_y^2}}, \quad (1.15)$$

$$A_\mu = \frac{g_y W_\mu^3 + g_w B_\mu}{\sqrt{g_w^2 + g_y^2}}, \quad (1.16)$$

the $W^\pm$ and the Z^0 acquire masses that are equal to:

$$M_{W^\pm}^2 = g_w^2 \frac{v^2}{4}, \quad (1.17)$$

$$M_{Z^0}^2 = g_w^2 \frac{v^2}{4} \left(1 + \frac{g_y^2}{g_w^2}\right) = M_{W^\pm}^2 / \cos^2(\theta_W), \quad (1.18)$$

where $\cos^2(\theta_W) = \frac{g_w^2}{g_w^2 + g_y^2}$, while the field A_μ associated with the photon remains massless. The Higgs boson acquires its mass through the self-interaction of the Higgs field and it depends on the parameters of the Higgs potential: $m_H = \sqrt{2\lambda}v$.

Yukawa term

The Higgs mechanism provides also a way to give mass to the fermions, both leptons and quarks, through the Yukawa term:

$$\mathcal{L}_{\text{Yukawa}} = -Y_{ij}^d \bar{Q}_{Li}^I \phi d_{Rj}^I - Y_{ij}^u \bar{Q}_{Li}^I \tilde{\phi} u_{Rj}^I - Y_{ij}^\ell \bar{L}_{Li}^I \phi \ell_{Rj}^I + \text{h.c.}$$

where $Y_{ij}^{u,d,\ell}$ are 3×3 complex matrices, with i and j indicating the generation indices, and $\tilde{\phi} = i\sigma^2 \phi^\dagger$. This term describes the Yukawa interactions between the Higgs field, lepton doublets L_i , and quark doublets Q_i . d_{Rj} , u_{Rj} and ℓ_{Rj} are the down-type quark, the up-type quark and the right-handed lepton fields, respectively. The Yukawa interactions are responsible for the masses of fermions and the mixing between different generations.

A particle consistent with the Higgs boson has been observed in 2012 by the ATLAS and the CMS collaboration [25], [26]. It has a mass of ~ 125 GeV and a spin-parity consistent with a 0^+ particle [27], [28].

1.1.2 Open questions

Despite the incredible predictive power, the Standard Model is an incomplete theory. There are fundamental physical phenomena in nature that the Standard Model does not adequately explain. Particularly noteworthy among these anomalies are neutrino masses, dark energy, dark matter, baryonic asymmetry, and notably gravity. However, the Standard Model also has certain peculiarities that, while not inherently incorrect, are rather unconventional, sometimes referred to as “aesthetic issues”.

Neutrino masses

According to the Standard Model, neutrinos exist as massless particles. However, neutrino oscillation experiments have shown that neutrinos do have mass [29] [30]. In

fact, neutrinos can fluctuate between three eigenstates of interaction, in a phenomenon known as neutrino oscillation. This inevitably leads to the modification of the SM theory. Some years ago, neutrinos were thought to be responsible for dark matter but with the present knowledge of their mass, they can only contribute an insignificant fraction of the dark matter. More importantly, neutrinos can't be dark matter because if they were, they would be too "hot", i.e., flying around with too much kinetic energy, which would spoil the formation of structure in the early universe.

Dark energy & dark matter

About 95% of the energy and matter that compose the universe is expected to be dark energy and dark matter, therefore only 5% of the universe is described by the SM [31]. In the late 1990s, observations of distant supernovae and other cosmological measurements indicated that the expansion of the universe is not slowing down, as expected due to the gravitational attraction of matter, but is instead accelerating. Dark energy was proposed as an explanation for this acceleration. The dark matter was postulated by Oort [32] and Zwicky [33] to explain the fast orbital velocities of the stars and galaxies. At that time, the measurement of the galaxy rotation curve was found to be much higher than expected in order to prevent the stars from flying away from the galactic centre as a result of the centrifugal force. The equilibrium state of the galaxies could be explained by assuming that they are formed by more matter than the visible stars. The first indirect evidence of dark matter was observed by Rubin in 1970 [34] when analyzing the spiral galaxy rotation curves. Until now, no dark matter candidates have been observed experimentally and only indirect evidence exists.

Baryonic asymmetry

In our observable universe, the prevalence of matter significantly outweighs antimatter. However, the Standard Model posits nearly perfect symmetry between the two. Does the asymmetry inherent in the Standard Model suffice to explain the existence of the observed matter, considering that most matter supposedly annihilated with antimatter in the aftermath of the Big Bang? It appears that this is not the case. Within the Standard Model, the symmetry between matter and antimatter is characterized by the invariance under simultaneous charge and parity transformations (CP symmetry). Experimental observations in the 1950s, particularly in the mixing of neutral Kaons, revealed CP asymmetry [35]. The Standard Model later attributed this asymmetry to the complex phase in the quark flavour mixing matrix, known as the Cabibbo–Kobayashi–Maskawa matrix. Despite this, the observed CP asymmetry is insufficient to account for the matter-antimatter disparity in the Universe.

Gravity

The Standard Model does not include gravity. Approaches such as string theory and loop quantum gravity are being explored, but a complete, experimentally verified quantum theory of gravity is still lacking. In analogy to the gauge bosons of the SM, gravity would be mediated by a hypothetical particle called the graviton. The graviton has not been detected, and its inclusion in the SM requires a consistent quantum field theory of gravity. Simply adding a graviton to the Standard Model does not yield a consistent or predictive theory due to fundamental issues with quantizing gravity. For example, naive quantization of gravity results in a non-renormalizable quantum field theory. The process of quantizing gravity as a perturbation theory reveals the necessity for an infinite number of independent parameters to define the theory. While choosing specific values for these parameters can provide a coherent framework for the theory, the impracticality of conducting an infinite number of experiments to precisely determine each parameter has led to arguments that, within perturbation theory, the theory lacks a fully meaningful physical interpretation.

Fine-Tuning problem

In addition to the significant limitations outlined earlier, which indicate that the Standard Model remains a highly effective theory within a specific energy range, there are certain anomalies within the model that appear “unnaturally peculiar”.

The Fine-Tuning, or Hierarchy, problem poses the question: Why is the mass of the Higgs boson at the electroweak scale, around 125 GeV, while quantum corrections suggest it should be much higher, close to the Planck scale?” If we assume the validity of the Standard Model up to the Planck scale, where gravity becomes significant, the loop corrections become orders of magnitude larger than the resulting corrected mass. This implies an extraordinary level of fine-tuning for these corrections to cancel each other precisely, yielding the observed low mass of the Higgs boson. This lack of naturalness leads to the name “the naturalness problem.”

To gain a deeper insight into the fine-tuning problem’s nature and origin, let’s examine the 1-loop self-energy corrections of a photon, electron, and scalar boson, as depicted in Fig. 1.3.

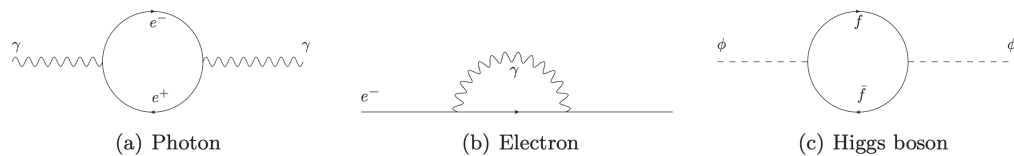


FIGURE 1.3: The Feynman diagrams of the one-loop mass corrections to the photon (a), electron (b) and the scalar Higgs boson (c).

Firstly, consider the photon self-energy (Fig. 1.3 (a)). The contribution to the 2-point function from this diagram is divergent, as is typical for all one-loop diagrams in Quantum Electrodynamics (QED). However, employing dimensional regularization results in the vanishing of this contribution. The precise $U(1)$ gauge symmetry plays a crucial role, effectively “protecting” the photon’s mass, ensuring it remains massless in all orders of perturbation theory.

The scenario is slightly different for the electron self-energy in Fig. 1.3 (b). This diagram is also quadratically divergent, and unlike the photon case, the divergence is not nullified by renormalization. Instead, it transforms into a logarithmic divergence, leading to a correction proportional to the electron mass. This phenomenon can be explained by invoking an approximate chiral symmetry precise only for massless particles. Indeed, setting the electron mass to zero entirely eliminates the contribution from diagram Fig. 1.3 (b).

On the other hand, the situation for the scalar boson (Fig. 1.3(c)) is much worse. In this case, there is no symmetry, exact or approximate, to shield it from the quadratic divergence originating from the fermion loop. As a result, the divergence persists, leading to corrections much larger than the expected mass for the Higgs boson. Achieving the correct value necessitates a fine-tuning of contributions at every order of perturbation theory, an approach that appears highly implausible. Consequently, the proposal of an additional symmetry emerged. This symmetry, known as Supersymmetry, aims to establish a symmetry between fermions and bosons, cancelling the divergence in the scalar boson mass.

1.2 Supersymmetry

Supersymmetry (SUSY) theory predicts a symmetry between fermions and bosons [36]. It was proposed as a way to solve the fine-tuning problem, however, it turns out that SUSY can address many other problems of the Standard Model. SUSY is formulated as a rather straightforward extension of the SM so it provides many testable predictions. For this reason, it has become the favourite Beyond the Standard Model (BSM) theory among many physicists.

Theoretical Framework

All space-time symmetries fall under the Poincaré group, characterized by 10 generators forming the Poincaré algebra through specific commutation relations. The 10 degrees of freedom for such transformations may be thought of as translation through time or space (four degrees); reflection through a plane (three degrees); or a “boost” in any of the three spatial directions (three degrees).

How can we incorporate a new symmetry into our theoretical framework?

The intuitive approach involves identifying generators of SUSY and establishing commutators with existing space-time symmetries.

However, the Coleman–Mandula theorem [37] asserts that is impossible to add to the Poincaré algebra a new symmetry in any way other than trivial. The SM exemplifies this limitation, featuring not only space-time symmetries but also a gauge symmetry that commutes with all space-time symmetries. Consequently, the resultant symmetry group becomes a tensor product of the Poincaré group and a gauge group.

To get around the constraints of the Coleman–Mandula theorem, an alternative path involves using anticommuting spinorial operators, Q_α , as additional generators. The Haag–Lopuszanski–Sohnius theorem [38] crucially states that Supersymmetry, a symmetry between fermions and bosons generated by this additional operator, represents the sole viable extension of the Poincaré group.

Supersymmetry, as defined by this theorem, posits the existence of a bosonic partner with the same mass for every fermion, and vice versa. These superpartners possess identical quantum numbers, such as colour charge and electric charge, as their respective Standard Model counterparts, differing only in spin. Consequently, fermions in the SM align with supersymmetric bosons as their partners, and vice versa. The quantum loop corrections of these super-partners, effectively nullify the quadratically divergent contributions from the Standard Model to the Higgs mass. This elegant resolution addresses the hierarchy problem.

Although this framework theoretically resolves the fine-tuning problems by cancelling bosonic and fermionic contributions, empirical observations challenge this notion. No superpartner, mass degenerate with SM ones, has been observed to date. In fact, the

cross-section for such particles suggests they should have been detectable by now, indicating that SUSY must be broken, so that superparticles are heavier than the known SM particles. If the scale of SUSY breaking, i.e., the mass scale of the superpartners is not excessively high (around the order of 1 TeV), it remains a plausible solution for addressing the fine-tuning problem, ensuring that corrections to the scalar mass remain within a reasonable scale.

Supersymmetry Breaking

As previously noted, SUSY seems to manifest as a broken symmetry.

The initial approach to break it, involved a mechanism similar to the Standard Model one. However, this method does not provide a phenomenologically viable model.

To overcome this challenge, many models renounce dynamic explanations for SUSY breaking and instead introduce “soft breaking terms” into the Lagrangian to account for our lack of knowledge about the breaking mechanism. A convenient way to express this in the Lagrangian is by decomposing it into two terms:

$$\mathcal{L} = \mathcal{L}_{\text{SUSY}} + \mathcal{L}_{\text{BREAK}}$$

where the first term preserves the symmetry, and the second term introduces the necessary breaking. However, to maintain the partial cancellation of the quadratic divergence of the Higgs boson mass, the term responsible for breaking the symmetry must be carefully added.

The soft SUSY breaking terms are:

- scalar mass terms: $-m_i^2|\phi_i|^2$, which provide mass to sfermions, breaking the mass symmetry with fermions;
- Gaugino mass terms, $-\frac{1}{2}m_l\bar{\lambda}_l\lambda_l$, which give mass to gauginos, differentiating them from gauge bosons;
- bilinear terms, $-B_{ij}\phi_i\phi_j$, which affect the Higgs sector, breaking the symmetry between Higgs fields and their superpartners;
- trilinear scalar interaction, $-A_{ijk}\phi_i\phi_j\phi_k$, which modify scalar interactions, breaking the symmetry in scalar sectors.

These terms explicitly break supersymmetry, resulting in higher mass for superpartners than particles, while still preserving the cancellation of quadratic divergences which is a key motivation for SUSY.

1.2.1 The Minimal Supersymmetric Standard Model

The Minimal Supersymmetric Standard Model (MSSM) is a SUSY model that introduces the fewest additional particles needed to maintain consistency with known physics.

In referring to the superpartners of the Standard Model particles, a “s” is added in front of the fermion’s name (e.g., quarks become squarks, and the top quark becomes the top squark or “stop”). For bosonic superpartners, the name is formed by appending “ino” at the end of the SM particle’s name. Collectively, all superpartners are termed “sparticles”, and a “ $\sim$ ” is placed on top of each to distinguish them from their Standard Model counterparts.

Initially, it is essential to delineate the superfields necessary for constructing the Minimal Supersymmetric Standard Model Lagrangian. As previously mentioned, Standard Model fermions are categorized into five representations of the SM gauge group. Consequently, we require five chiral superfields to represent them:

- Q containing quark and squark $SU(2)$ doublets and $SU(3)$ triplets;
- U and D containing quark and squark $SU(2)$ singlets and $SU(3)$ triplets;
- L containing lepton and slepton $SU(2)$ doublets and $SU(3)$ singlets;
- E containing lepton and slepton $SU(2)$ and $SU(3)$ singlets.

Including only the superpartners of SM particles leads to anomalies in the Higgs sector, that are resolved in the MSSM incorporating two complex Higgs doublets. One doublet provides mass to up-type quarks, and the other to down-type quarks and leptons, resulting in five physical Higgs bosons, including two SM-like Higgs bosons:

- A a neutral pseudoscalar (CP-odd);
- h “light” neutral scalars (CP-even);
- H “heavy” neutral scalars (CP-even);
- H^+ the positive charged Higgs boson;
- H^- the negative charged Higgs boson.

The chiral superfields of the MSSM are outlined in Tab.1.2.

A single vector superfield contains 8 gluons (g) and 8 gluinos ($\tilde{g}$), automatically emerging when $SU(3)$ is utilized as a gauge group during the construction of vector superfields.

Similarly, we introduce a superfield encompassing three weak vector bosons ($W^\pm$ and W^0) and three wino particles ($\tilde{W}$). We naturally require a superfield comprising the

$U(1)$ gauge boson and its corresponding supersymmetric partner: the bino $\tilde{B}$. Higgsinos ($\tilde{H}$) are spin-1/2 superpartners of Higgs bosons.

Due to Electroweak Spontaneous Symmetry Breaking (SSB), the bino, the three winos and the four higgsinos particles do not exist as mass eigenstates. Instead, they undergo mixing to give rise to observable particles: two pairs of charginos and four neutralinos, which are electrically charged and neutral, respectively. Charginos are denoted as $\tilde{\chi}_i^\pm$, and neutralinos as $\tilde{\chi}_i^0$, where i is an integer representing different mass eigenstates, with $i = 1$ referring to the lightest such particle. Superpartners for vector bosons are schematized in Tab. 1.3.

		Spin 0	Spin $\frac{1}{2}$	$(SU(3), SU(2), U(1))$
quarks and squarks	Q	$(\tilde{u}_L, \tilde{d}_R)$	(u_L, d_L)	$(3, 2, \frac{1}{6})$
	U	$\tilde{u}_R^*$	$u_R^\dagger$	$(\bar{3}, 1, -\frac{2}{3})$
	D	$\tilde{d}_R^*$	$d_R^\dagger$	$(\bar{3}, 1, \frac{1}{3})$
leptons and sleptons	L	$(\tilde{\nu}_L, \tilde{e}_L)$	(ν_L, e_L)	$(1, 2, -\frac{1}{2})$
	E	$\tilde{e}_R^*$	$e_R^\dagger$	$(1, 1, 1)$
higgs and higgsinos	H_u	$(h_u^\dagger, h_u^0)$	$(\tilde{h}_u^\dagger, \tilde{h}_u^0)$	$(1, 2, \frac{1}{2})$
	H_d	$(h_d^\dagger, h_d^0)$	$(\tilde{h}_d^\dagger, \tilde{h}_d^0)$	$(1, 2, -\frac{1}{2})$

TABLE 1.2: Chiral superfields within the MSSM consist of various multiplets. The initial five rows represent the standard multiplets of SM, while the last two rows depict the two Higgs doublets along with its supersymmetric partner. The final column provides information on the representation of the gauge group to which each multiplet belongs.

	Spin $\frac{1}{2}$	Spin 1	$(SU(3), SU(2), U(1))$
gluinos and gluons	$\tilde{g}$	g	$(8, 1, 0)$
winos and W-bosons	$\tilde{W}^\pm, \tilde{W}^0$	$W^\pm, W^0$	$(1, 3, 0)$
bino and B-boson	$\tilde{B}$	B	$(1, 1, 0)$

TABLE 1.3: Vector superfields of the MSSM. Within these superfields, the third column designates the Standard Model gauge bosons, and the second column represents their respective supersymmetric partners. The last column provides information about the representation of the multiplet within the gauge group.

R-parity

Several proposed theories suggest the existence of a stable, weakly interacting, and electrically neutral particle known as the Lightest SUSY Particle (LSP), making it a po-

tential candidate for dark matter. To prevent the decay of the LSP into any SM particle, a concept called R-parity is introduced.

R-parity is a quantum number defined as:

$$R = (-1)^{3B+L+2S} \quad (1.19)$$

where: B is the baryon number $B = \frac{1}{3}(n_q - n_{\bar{q}})$, the number of quarks minus the number of antiquarks divided by 3; L is the lepton number $L = n_\ell - n_{\bar{\ell}}$, the number of leptons minus the number of antileptons; S is the spin of the particle.

R-parity is a multiplicative quantum number that distinguishes between ordinary particles (R-parity +1) and their supersymmetric partners (R-parity -1). The conservation of R-parity has several consequences in supersymmetric models:

1. supersymmetric particles are produced or annihilated in pairs, leading to the stability of the lightest supersymmetric particle;
2. the LSP is a good dark matter candidate since, being stable and weakly interacting, it could contribute to the observed dark matter in the universe;
3. the conservation of R-parity prevents rapid proton decay, which is an experimental constraint.

However, it's important to note that the concept of R-parity is a theoretical construct, and experimental observations have not confirmed its existence. While it is a convenient tool in many SUSY models, some SUSY models allow for R-parity violation.

Limits of the MSSM

Despite its initial consideration as a natural solution, the MSSM faces some challenges. The MSSM superpotential includes a mass term, represented by the parameter μ , which is a critical component in the formulation of the model:

$$f_{MSSM} = \sum_{i,j=1}^3 [(\lambda_E)_{ij} H_d L_i E_j + (\lambda_D)_{ij} H_d Q_i D_j + (\lambda_U)_{ij} H_u Q_i U_j] + \mu H_u H_d. \quad (1.20)$$

The last term of Eq. 1.20 couples the two Higgs doublets H_u and H_d . This coupling is essential for imparting masses to Higgs bosons and higgsinos.

In the MSSM this relation between the Higgs masses m_{H_u} , m_{H_d} and the μ parameter holds [39]:

$$\frac{1}{2} M_Z^2 = \frac{(m_{H_d}^2 - \tan^2 \beta m_{H_u}^2)}{\tan^2 \beta - 1} - \mu^2, \quad (1.21)$$

where M_Z is the Z boson mass and $\tan \beta$ is the ratio of the vacuum expectation values of the two Higgs doublet fields. In the absence of fine tuning no large cancellations should occur in the right-hand side, and both terms should be of comparable magnitude with the left-hand side term. It follows that μ must be at the order of M_Z , at the electroweak scale. However, the only apparently “natural” values that μ might assume are either zero or approximately the Planck scale, both of which are disallowed and significantly distant from the electroweak scale [36].

An additional challenge comes from the so-called “little hierarchy problem” [39, 36]. This arises due to a combination of experimental constraints from collider searches, particularly those conducted at the Large Electron-Positron Collider (LEP), and theoretical considerations. Experimental searches at LEP set lower limits on Higgs scalar with a SM-like coupling to the Z boson and limits on the masses of superpartners, such as charginos, neutralinos, and sleptons. These searches did not find evidence for superpartners within the reach of the collider’s energy, imposing lower bounds on their masses. From theoretical considerations, it follows that the loop corrections to the Higgs mass need more fine-tuning the higher the masses of the superpartners in the loop. However, if the superpartner masses are pushed to relatively high scales to satisfy experimental constraints, MSSM parameters have to be tuned with a relative precision of $\sim 10^{-2}$ [39].

Another relevant issue of the MSSM comes from experimental searches for dark matter, which over the time have considerably restricted the regions within the MSSM parameter space to produce a viable dark matter candidate [40].

1.2.2 The Next-to-Minimal Supersymmetric Standard Model

The Next-to-Minimal Supersymmetric extension of the Standard Model (NMSSM) has been specifically designed to address the previously exposed issues of the MSSM. This is achieved by substituting the μ parameter with a gauge singlet scalar superfield (S), that couples the two scalar Higgs doublets: $\lambda SH_u H_d$, where λ is a coupling constant. The μ parameter is subsequently dynamically generated as this field acquires the vacuum expectation value at the electroweak scale. The NMSSM typically results in a richer Higgs sector, expanding the phenomenology of the MSSM. This additional complexity can lead to different patterns of radiative corrections to the Higgs boson mass, potentially reducing the fine-tuning required to stabilize the electroweak scale.

Moreover, the NMSSM offers mechanisms for addressing the little hierarchy problem by introducing additional degrees of freedom into the Higgs sector. Instead of pushing the Higgs mass above the LEP lower limit, in the NMSSM it is possible to find parameter space where LEP constraints do not apply [39]. Figure 1.4 shows the result of a scan of the full space of parameters that satisfy the experimental and theoretical constraints of the MSSM, NMSSM and other supersymmetric extension of the SM. Significant dif-

ferences are observed among the models for the mass range of the lightest CP-even Higgs mass. In the MSSM, the scalar Higgs mass range is very restricted, while the upper bound in the NMSSM is extended. Notably, the lower bound in the NMSSM can be much lower than the SM Higgs mass.

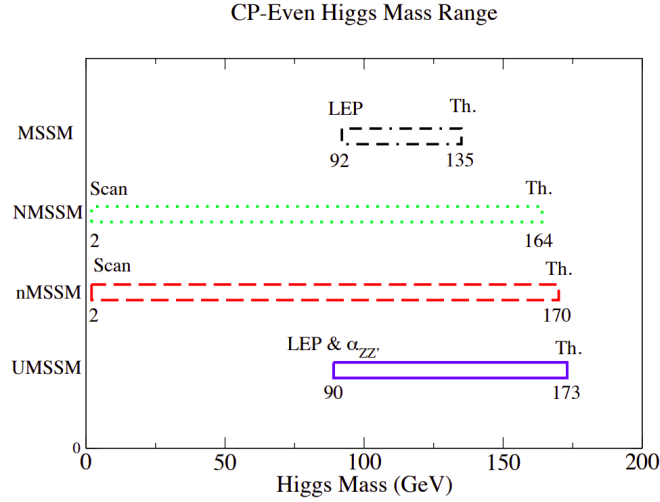


FIGURE 1.4: Mass ranges of the lightest CP-even Higgs in various supersymmetric extensions of the SM, as resulted from a random scan of the parameters space, accounting experimental and theoretical constraints. Together with the MSSM, there is the Next-to-Minimal Supersymmetric Standard Model (NMSSM), the new Minimally-extended Supersymmetric SM (nMSSM) and UMSSM which is a $U(1)'$ extension of the MSSM. The motivations of the bounds are also provided. Th. = theoretical bound; Scan = limit value is sensitive to the chosen parameter space; LEP = experimental constraints from LEP. Figure from Ref. [3].

Moreover the introduction of the additional singlet superfield, in the NMSSM, modifies the properties of the neutralino and its interactions with other particles, potentially altering its viability as a dark matter candidate. Specifically, the NMSSM can lead to scenarios where the neutralino is lighter or possesses different couplings compared to the MSSM, thereby evading some constraints imposed by experimental searches for dark matter [40].

The primary distinction between the MSSM and NMSSM lies in the superpotential, which is now formulated as [41]

$$f_{NMSSM} = \sum_{i,j=1}^3 [(\lambda_E)_{ij} H_d L_i E_j + (\lambda_D)_{ij} H_d Q_i D_j] + (\lambda_U)_{ij} H_u Q_i U_j + \lambda S H_u H_d + \frac{k}{3} S^3. \quad (1.22)$$

In contrast to Eq. 1.20, it is evident that μ has been substituted with λS , and an additional term, cubic in the singlet field S , has been introduced. Without the latter modification, the newly established global $U(1)$ symmetry of S in the term $\lambda S H_u H_d$ might have resulted in an additional massless CP-odd scalar, commonly known as the Peccei-Quinn axion [42].

The introduction of the new superfield gives rise to three additional particles: two gauge singlet Higgs bosons and one gauge singlet Higgsino. These particles undergo mixing with the original Higgs bosons and neutralinos, resulting in a distinctive particle spectrum. Unlike the MSSM, the NMSSM ultimately features:

- 3 CP-even neutral Higgs bosons: H_1 , H_2 , and H_3
- 2 CP-odd neutral Higgs bosons: A_1 and A_2
- 2 charged Higgs bosons (H^+ , H^-)
- 5 neutralinos

The spin-1/2 superpartner associated with the singlet field, known as the singlino, undergoes a mixing process with the higgsinos and neutral gauginos. This mixing results in the formation of an extra neutralino compared with the MSSM.

All other particles align with those found in the MSSM (see Tab. 1.4).

Particle	MSSM	NMSSM
Higgs Bosons	$\tilde{h}^0, \tilde{H}^0, \tilde{A}^0, \tilde{H}^\pm$	$H_1, H_2, H_3, A_1, A_2, H^\pm$
Squarks	$\tilde{u}_L, \tilde{u}_R, \tilde{d}_L, \tilde{d}_R, \tilde{c}_L, \tilde{c}_R, \tilde{s}_L, \tilde{s}_R, \tilde{t}_L, \tilde{t}_R, \tilde{b}_L, \tilde{b}_R$	same
Sleptons	$\tilde{e}_L, \tilde{e}_R, \tilde{\nu}_e, \tilde{\mu}_L, \tilde{\mu}_R, \tilde{\nu}_\mu, \tilde{\tau}_L, \tilde{\tau}_R, \tilde{\nu}_\tau$	same
Neutralinos	$\tilde{\chi}_0^1, \tilde{\chi}_0^2, \tilde{\chi}_0^3, \tilde{\chi}_0^4$	$\tilde{\chi}_0^1, \tilde{\chi}_0^2, \tilde{\chi}_0^3, \tilde{\chi}_0^4, \tilde{\chi}_0^5$
Charginos	$\tilde{\chi}_1^\pm, \tilde{\chi}_2^\pm$	same
Gluino	$\tilde{g}$	same
Gravitino	$\tilde{G}$	same

TABLE 1.4: Mass eigenstates of the MSSM compared with the NMSSM [18][19].

The experimental search reported in Chapter 4 makes use of a model in the NMSSM framework. The detailed presentation of the signal model is provided in Section 4.2. The key feature of this model is the assumption that the Lightest Supersymmetric Particle takes the form of a singlino-like neutralino. This specific configuration results in distinctive SUSY decay cascades, presenting features that might have remained unnoticed in more conventional SUSY searches.

Chapter 2

The CMS Experiment at the LHC

2.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is today the biggest and most powerful particle accelerator in the world. It is part of CERN, the European Centre for Nuclear Research, located on the French-Swiss border near the city of Geneva.

The LHC was designed to collide protons at an energy in the center of mass of 14 TeV. However, during Run 2 collisions occurred at 13 TeV, while in Run 3 the collider is running at 13.6 TeV. Additionally, the LHC can produce heavy nuclei collisions, in particular Pb-Pb and Pb- p . The collider is situated in a circular tunnel 27 km long, at a depth that varies from 45m up to 170m. At the LHC two particle beams pass in two ultra-high vacuum beam pipes and are accelerated in two opposite directions thanks to radio frequency cavities. The circular motion of the particles is allowed by the use of superconducting dipolar magnets, capable of generating a magnetic field up to 8.3 T, while the beam is collimated with multipolar magnets. In order to operate in the superconductive regime, the magnets are cooled down to 1.9 K using liquid helium. The two circulating particle beams collide in four points. In each interaction point is located a detector: ATLAS (A Toroidal Lhc ApparatuS), CMS (Compact Muon Solenoid), LHCb (LHC Beauty) and ALICE (A Large Ion Collider Experiment). ATLAS [43] and CMS [11] are multi-purpose detectors and are designed for the detection of particles produced in proton-proton collisions and both have been optimized for the search of the Higgs boson and BSM physics. ALICE is an experiment designed to study collisions between heavy nuclei and LHCb (LHC Beauty) specialises in studying the physics of the b quark and CP symmetry violation.

2.1.1 The accelerating system

CERN has a complex system of increasing energy accelerators. Each machine injects its beam into the next one which, in turn, accelerates it to a higher energy. The last element of this chain is the LHC. In addition, most accelerators in the system have an associated experimental area where beams can be extracted and used in experiments with lower energies. In Fig. 2.1, a schematic description of the CERN accelerator complex and the various experiments is presented.

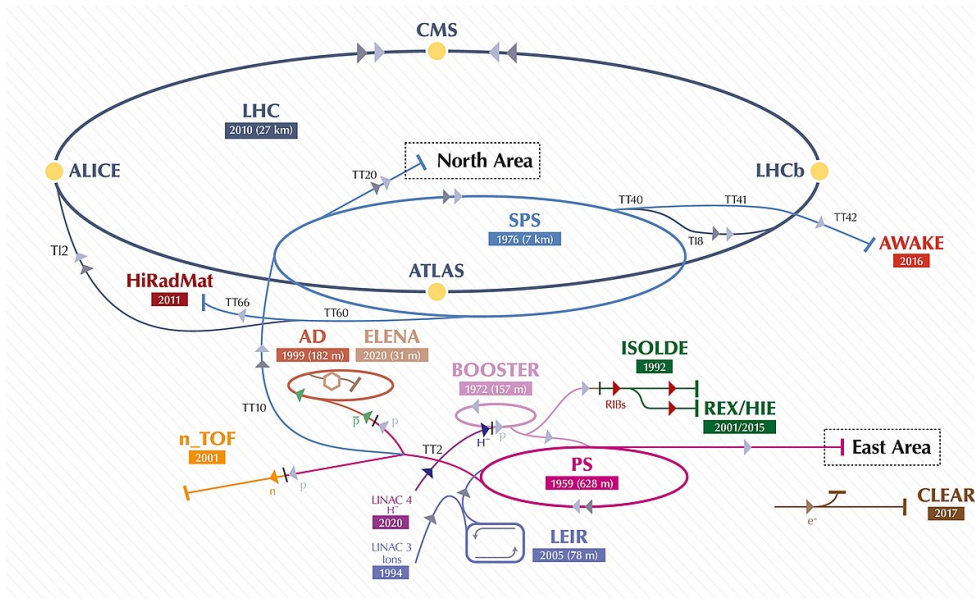


FIGURE 2.1: The accelerator complex and the various experiments operating at CERN. Figure from Ref. [4].

Proton acceleration starts with atomic hydrogen gas, coming from a small cylinder, from which the electrons are removed. What remains are hydrogen nuclei, or protons, which are injected in the form of packets (bunches) into a linear accelerator (LINAC 4) that leads them to an energy of 160 MeV. These protons are then injected into a small proton synchrotron, the PS Booster (PSB), which accumulates them in four separate overlapping rings and accelerates them to 1.4 GeV. The beam is then transferred into the Proton Synchrotron (PS) which brings it to 25 GeV of energy and which is then injected in turn into the Super Proton Synchrotron (SPS), where it will be accelerated to 450 GeV. At this point, the beam is injected into the LHC. The cycle described above is repeated many times to fill the accelerator up to the desired number of bunches (typically about 2800 per ring at full capacity). The filling operation typically requires about 5 minutes for each of the two rings. Once the two rings containing the proton beams circulating in the opposite direction are filled, it takes about 20 minutes to simultaneously bring the energy of the beams to its maximum value. At the end of the acceleration phase, the beams are made to collide in the four interaction points with the aid of appropriate magnets. At the beginning of the collision in Run 2, bunches are com-

posed of about 1.2×10^{11} particles and the time between each bunch crossing is 25 ns. During collisions, the number of particles per bunch (and consequently instantaneous luminosity) decreases since a small fraction is lost in each collision. When instantaneous luminosity drops too much, beams are dumped and the cycle restarts from the beginning. Under normal conditions, protons can circulate for several hours allowing experiments to collect data for long periods.

The LHC system can also accelerate lead nuclei. A highly purified sample of this metal is heated to a temperature of about 500 degrees Celsius and the resulting vapour is passed through by an electric current. Some of the 82 electrons forming the lead atoms are ripped off and the ionized atoms are pre-accelerated (LINAC 3) and passed through a sheet of graphite. In this passage, many other electrons are taken away from the atoms which, as a result, acquire an even greater positive electrical charge (typically 54+). The ionized lead atoms are then accumulated in the Low Energy Ion Ring (LEIR) which transfers them to the PS. Here they are further accelerated to about 7 GeV, passed through a second sheet of graphite that rips off all the remaining electrons and finally injected into the SPS. At this point, the lead nuclei are accelerated to 180 GeV and injected into the LHC which will bring them to 2.8 TeV.

2.1.2 Operations

An outline of the LHC program is illustrated in Fig. 2.2. The collider achieved full operational status in 2009 and is expected to continue operation until at least 2037. The operation timeline is divided into distinct phases: Runs, dedicated to collision events and data acquisition, and Shutdowns, allocated for essential maintenance and upgrades of the collider complex and of the experiments. We are currently in the middle of Run 3. The work described in this thesis touches three different CMS time phases: an upgrade that was made to the Electromagnetic Calorimeter calibration system during Long-Shutdown 2 (LS2); an analysis work which aims at publication at the end of Run 3 data taking; the development of a new algorithm for the first level of the trigger system that will hopefully be implemented at the High-Luminosity LHC, which is foreseen to start in 2029.

The instantaneous luminosity is a critical parameter in the collider performance and is defined as:

$$L = \gamma \frac{n_b N^2 f_{rev}}{4\pi\beta^* \epsilon_n} R. \quad (2.1)$$

Here, γ represents the relativistic gamma factor, n_b denotes the count of colliding bunches at the interaction point, N stands for the number of protons per bunch, f_{rev} is the frequency of revolution of the bunches within the ring, β^* is the focal length of the beam, and ϵ_n corresponds to the normalized transverse emittance of the beam. Additionally, R represents a geometrical factor contributing to luminosity reduction. The increase



FIGURE 2.2: The LHC plan spanning from 2011 to 2029, with luminosity and energy goals, reported in red at the bottom and at the top, respectively. At the end of Run 3 in 2025, CMS will undergo the Phase-2 upgrade in preparation for the High-Luminosity LHC (HL-LHC). Figure from Ref. [5].

of instantaneous peak luminosity has been achieved gradually throughout the operational life of the LHC. This gain has been accomplished by meticulously adjusting the focus of the beams in proximity to the interaction regions.

Another fundamental factor to consider is the integrated luminosity denoted as $\int L(t)dt$. The luminosity experiences fluctuations over time as a result of ongoing beam focusing and the gradual reduction in beam quality caused by successive interactions. The integrated luminosity is the metric that approximates the quantity of collisions occurring at the LHC, by $N = \sigma_{pp} \int L(t)dt$. Here, N represents the number of collisions, σ_{pp} is the cross-section of proton-proton interactions, and the integral corresponds to the integrated luminosity.

Figure 2.3 reports the integrated luminosity recorded by the CMS experiment throughout its operations. The first operational phase of the LHC, known as Run 1, spanned a period of two years from 2009 to 2011. During this time, the beam energies were gradually increased. In its initial months of operation, the LHC surpassed the Tevatron's energy record of 0.98 TeV per beam, establishing itself as the world's highest-energy particle accelerator. Subsequently, the beam energies were raised further, reaching 3.5 TeV in 2011 and 4 TeV in 2012. The first long shutdown of the collider (LS1) occurred upon reaching an integrated luminosity of 30 fb^{-1} , with the peak luminosity reaching approximately 75% of the design value ($1 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$). During this shutdown period, optimizations were carried out on the acceleration cycle and adjustments were made to the ring magnets to accommodate 6.5 TeV beams. The subsequent phase, Run 2, started in 2015 and featured a center-of-mass energy of 13 TeV. This Run also achieved and doubled the design luminosity of the collider. Run 2 resulted in five

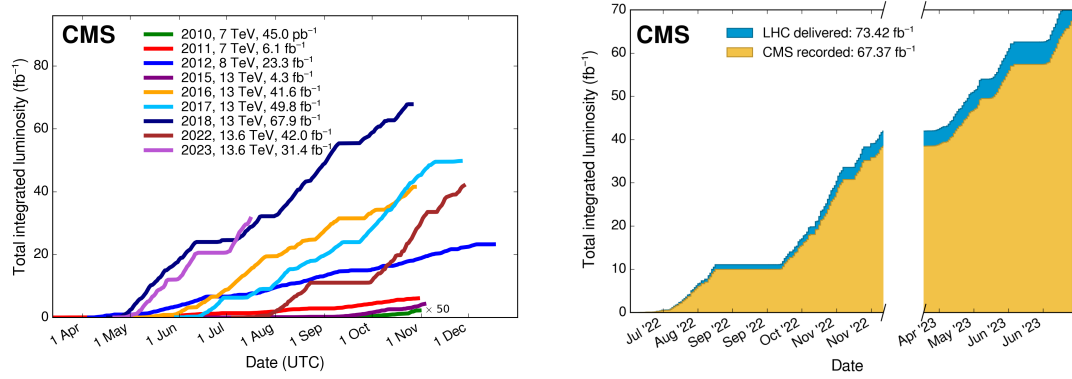


FIGURE 2.3: On the left, the CMS integrated luminosity data per year from 2010 to July 2023. On the right, focus on Run 3 integrated luminosity until July 2023. The CMS experiment recorded a total of about 67 fb^{-1} with a peak instantaneous luminosity of $2.22 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$, more than doubling the LHC design luminosity. Figure from Ref. [6].

times the integrated luminosity collected during Run 1 and ended in December 2018, with a peak luminosity of $2.14 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ at a center-of-mass energy of 13 TeV. The data-taking activities of the collider restarted in July 2022 for Run 3. Looking ahead, the LHC is scheduled for an upgrade during an extended shutdown known as LS3, as part of the High Luminosity program. This upgrade aims to achieve peak luminosities ranging from 5 to 7.5 times the nominal LHC luminosity. The HL-LHC upgrade is further described in Section 8.1.

2.2 The CMS Experiment

This section's purpose is the description of the currently used experimental apparatus of the CMS detector. A future upgrade of the detector planned for High-Luminosity LHC is described in Section 8.1.1.

The CMS experiment was designed and built to investigate physics phenomena occurring at the TeV scale across a variety of signatures and final states. As a result, it was intentionally made as a versatile apparatus, equipped with multiple subsystems specialized in identifying and measuring different particle categories.

The detector is constituted by a cylindrical structure measuring 15 m in diameter and 21.5 m in length, with a total weight of approximately 12,500 t. Collisions happen in the center of the CMS experiment every 25 ns, which implies that the particles are produced in the interaction point before those generated in the previous bunch crossing have even escaped the detector's external layers. Furthermore, each bunch crossing can involve numerous proton interactions. These two effects are collectively referred to as "out-of-time" and "in-time" pileup (PU), and they overlap with the signal of interest.

Due to the demanding conditions in which the detector operates, its design needs exceptional granularity, rapidity of response, and resilience to radiation. Simultaneously, it must precisely estimate the energy and momentum of the particles and be able to identify them.

2.2.1 Coordinate system

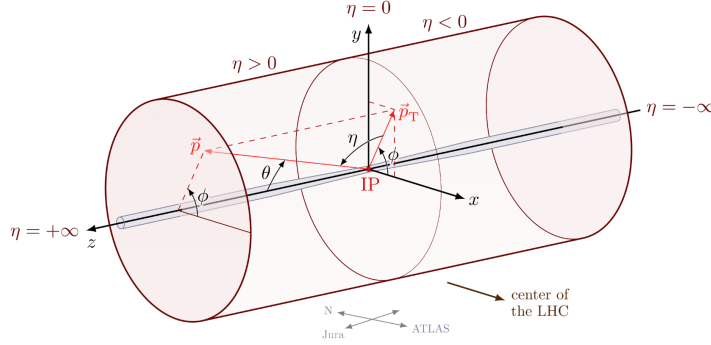


FIGURE 2.4: The CMS coordinate system. Figure from Ref. [1].

The standard coordinate system used at the CMS experiment is shown in Fig. 2.4. It is centred on the nominal interaction point and uses a right-handed orientation. The x-axis aligns with the centre of the LHC ring, the y-axis points upward, and the z-axis extends in the counterclockwise direction of the proton beam. Given the cylindrical nature of the detector, a polar coordinate system is also often employed. The azimuthal angle ϕ is defined within the (x,y) plane as the angle formed with respect to the positive x-axis. In this plane, the radial coordinate is labelled as "r". The polar angle θ is defined within the (r,z) plane as the angle formed with the z-axis and is commonly substituted with the equivalent pseudorapidity η using

$$\eta = -\ln \left(\tan \frac{\theta}{2} \right). \quad (2.2)$$

As θ approaches 90° , the pseudorapidity tends to zero, and when θ approaches 0° or 180° , it tends towards $+\infty$ or $-\infty$, respectively. The preference for η over θ arises from the fact that, under the assumption of massless particles, which is a good approximation at the LHC's high-energy interactions, η aligns with the rapidity,

$$y = \frac{1}{2} \ln \frac{E + p_z}{E - p_z}. \quad (2.3)$$

Differences in rapidity, Δy , are invariant under Lorentz boosts along the beam axis. Consequently for a pair of massless particles $\Delta \eta$ does not depend on the longitudinal boost of the reference frame. This characteristic is highly significant in hadron collider physics since the partons contained within the hadrons carry varying proportions of

the initial longitudinal momentum. This leads to collisions exhibiting varying boosts along the beam axis. The spatial separation between two particles can be expressed with respect to their angular separation as:

$$\Delta R^2 = \Delta\phi^2 + \Delta\eta^2. \quad (2.4)$$

The component of a particle's momentum projected onto the transverse plane is known as transverse momentum, denoted as p_T . Notably, this value remains unaffected by the Lorentz boost resulting from the initial momentum of the interacting partons along the z-axis.

2.2.2 Structure of the detector

The CMS detector, outlined in Ref. [11], is divided into distinct sections: a central "barrel" area and two "endcaps" in the forward regions. The demarcation points in terms of η that differentiate these two regions depend on the specific subsystem under consideration and are shown in Fig. 2.5. CMS experiment is arranged in concentric layers

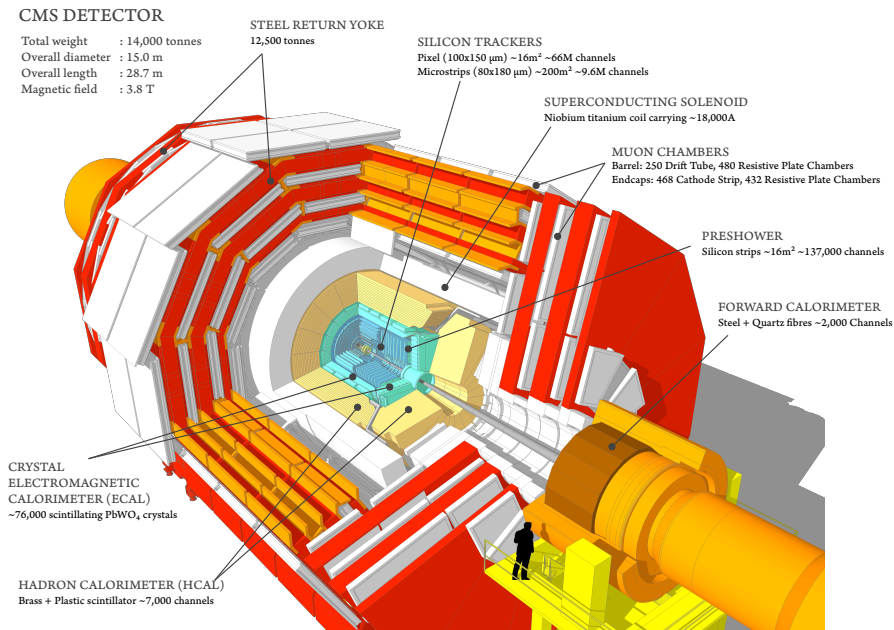


FIGURE 2.5: The scheme of the CMS detector. Key subsystems are highlighted. Figure from Ref. [7].

of detectors for the identification and measurement of particles produced by collisions. The interaction point is surrounded by the tracker detector, serving the purpose of accurately determining the positions of the primary interaction vertices, as well as tracking the trajectory and measuring the momentum of charged particles.

Around the tracking system, there are the electromagnetic and hadronic calorimeters, designed to absorb the energy deposited by electrons, photons and hadrons. Muons,

however, can cross the calorimeters and are measured via the muon tracking systems situated in the outermost section of the detector.

A key element of the experimental setup is the niobium-titanium superconducting solenoid, with a diameter of 6 m. Operating at a temperature of 4.5 K, this solenoid generates a strong 3.8 T magnetic field along the z -axis. This intense magnetic field serves to bend charged particles' trajectory, allowing for the measurement of their transverse momentum through the tracking subdetectors. Both the tracker and calorimeters are housed within the solenoid, imposing stringent spatial constraints and necessitating dense materials for the calorimeters to contain incoming particles and their secondary products of interaction.

The magnet's return field of about 2 T is exploited to measure transverse momentum through muon detectors placed inside the iron structure surrounding the solenoid. This results in muon trajectories curving in opposing directions within the inner tracker and muon systems. The information collected by individual subdetectors frequently exhibits redundancy and can be combined to allow the reconstruction of final-state objects.

In the subsequent sections, a comprehensive description of the subdetectors will be provided.

2.2.3 Solenoid Magnet

Constituted by a coil of superconducting wires, this component generates a magnetic field when an electric current flows through it [44]. It is characterized by an overall length of 12.5 meters and a diameter of 6 m, featuring a return yoke designed to manage the field and curve particles in the external parts of the detector. The magnet generates a magnetic field of 4 T, aligned along the beam axis and uniform inside its volume. When charged particles are produced inside the magnet their paths curve, forming a helix due to the uniform magnetic field strength. The part of the track aligned with the beamline remains unaffected by the magnetic field since charged particles only experience curvature in the plane perpendicular to the magnetic field direction. The measurement of the curvature allows the determination of the particles' charge and the precise measurement of their transverse momentum. The projection of the track onto the $r - \phi$ plane forms a circular pattern, and the helix's radius R (in centimetres) can be expressed as:

$$R = \frac{p_T}{0.003 \times q \times B} \quad (2.5)$$

Here, p_T is the transverse momentum in GeV, q is the particle's charge, and B stands for the magnetic flux density within the CMS solenoid expressed in T. The radius is influenced by both the particle's p_T and its charge. Consequently, particles with greater momentum exhibit less curvature. The direction of curvature is determined by the par-

ticle's charge. To enhance accuracy, subdetectors like the tracker and calorimeters are positioned inside the solenoid, avoiding scattering effects that occur when particles interact with the dense matter of the solenoid. Beyond the solenoid magnet, steel-made return yokes regulate the magnetic field and prevent all Standard Model particles, except muons and neutrinos, from entering the muon chambers. Additionally, these return yokes serve as the primary structural support for the entire experiment.

2.2.4 Tracker

The tracker is the subdetector closest to the interaction point within the CMS detector setup [45, 46]. Its primary function is the precise determination of the trajectories of charged particles in three dimensions, alongside the measurement of their charge and momentum. Notably, particles with lower momentum exhibit more pronounced bending in the magnetic field, facilitating their momentum measurements with higher accuracy. Thanks to the tracker's close proximity to the interaction point and its high spatial resolution, the measured particles' trajectories are used to reconstruct the collision's primary vertex. It comprises pixel and strip silicon detectors, with the former positioned closest to the beamline, offering superior granularity compared to the latter. The layout of the tracker which will be operational until the end of 2025, during the so-called Phase-1, is illustrated in Fig. 2.6. It covers the region $|\eta| < 2.5$, and was originally designed to function with an average pileup of 20-30 collisions per bunch-crossing, but it exhibits robust performance even at higher luminosities. As charged particles pass through the tracker, they ionize the silicon sensors, generating a small measurable charge. However, electrically neutral particles pass through the tracker undetected. The tracker is designed to be relatively lightweight, allowing particles to cross it with minimal multiple scattering.

The track transverse impact parameter (d_0) is the shortest distance between the trajectory of a charged particle and the primary vertex in the plane perpendicular to the beam axis. The track longitudinal impact parameter (z_0) corresponds to the z coordinate of the point of closest approach of the track to the z -axis. The resolution of d_0 is better than $35 \mu\text{m}$ over the full $|\eta| \leq 2.5$ range for particles with p_T above 10 GeV, while the z_0 resolution is significantly better than $75 \mu\text{m}$ over most of the rapidity range [45].

Pixel Detector

The pixel tracker has the highest granularity among all CMS subdetectors, essential for accurately resolving the small particle separations near the interaction point [47]. Its precision enables the measurement of secondary vertices originating from short-lived particles, including hadrons containing b quarks, which decay within the tracker volume. Each layer of the tracker comprises radiation-resistant silicon sensors, with the

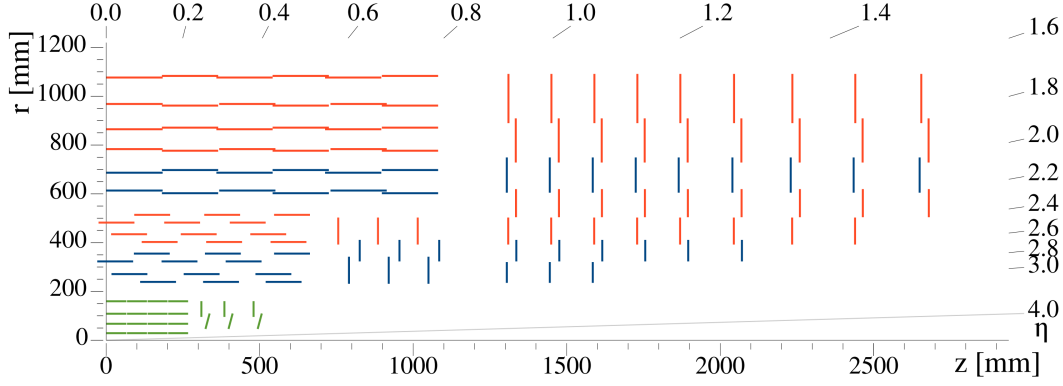


FIGURE 2.6: The Phase-1 tracker configuration is shown, illustrating one quadrant of the tracker in the r - z plane, centred around the interaction point. The pixel detector is represented in green, while the single-sided and double-sided strip detectors are displayed in red and blue, respectively. Figure from Ref. [8].

four-barrel layers positioned between 30 and 160 mm from the beamline, as depicted in Fig. 2.6. In total, the tracker comprises 1856 sensor modules housing over 100 million pixels. The layout is strategically designed to ensure four hits per charged particle within the region $|\eta| < 2.5$. A minimum of three hits is required for track reconstruction (two hits if the particle's origin is assumed to be at the detector's centre). With a pixel size of $100\mu\text{m} \times 150\mu\text{m}$, the tracker achieves an excellent resolution in both the z and ϕ directions, with an average occupancy of around 10^{-4} hits per pixel per bunch-crossing at design luminosity. The overall spatial resolution of the pixel tracker is approximately $10\mu\text{m}$. Each silicon sensor module is equipped with readout chips that amplify sensor signals and temporarily store recorded events in buffers, allowing storage of nearly 200 events simultaneously.

Strip detector

The strip detector, positioned at radii greater than 200 mm in the barrel region surrounding the pixel tracker, expands the tracking volume for particles. It comprises 15148 sensor modules distributed across ten barrel layers and twelve disks on each side of the barrels. Being farther from the interaction point, the strip detector does not necessitate the same level of resolution and granularity. Utilizing silicon strips, it achieves a coarser resolution in the z -direction (radial direction) in the barrel (disk) regions, while maintaining excellent resolution in the $(r - \phi)$ plane, where p_T is measured. Some layers of the strip detector incorporate double-sided modules, achieved by bonding two single-sided strip modules together at an angle, enhancing resolution in the $(r - z)$ plane. Overall, the strip detector includes approximately 10 million strips, ranging in size from $10\text{ cm} \times 80\mu\text{m}$ to a maximum of $25\text{ cm} \times 180\mu\text{m}$. With the double-sided modules, single-hit resolution can reach as low as $23\mu\text{m}$, contingent on the sensor module and its position. Despite the larger sensor cells, the occupancy re-

mains at most a few percent. Similar to the pixel tracker, the strip sensor modules also integrate readout chips to amplify signals and manage event data.

2.2.5 Electromagnetic calorimeter

The electromagnetic calorimeter (ECAL) is a homogeneous calorimeter that surrounds the tracker, with the main goal of absorbing and accurately estimating the energy of particles that interact predominantly through electromagnetic interaction, such as electrons/positrons and photons [9].

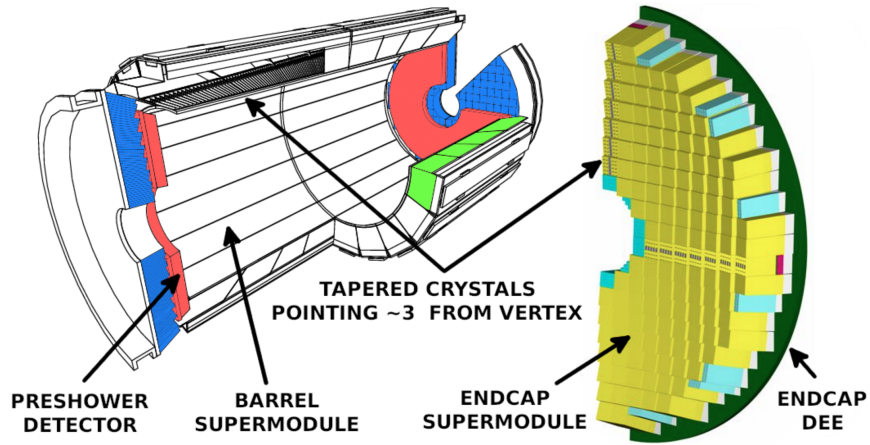


FIGURE 2.7: Layout of the CMS electromagnetic calorimeter showing the arrangement of crystal modules (in green), supermodules and endcaps. Figure from Ref. [9].

Constituted of lead tungstate (PbWO_4) crystals, the ECAL is engineered to produce electromagnetic showers when traversed by electrons or photons, starting a cascade of processes including pair production, bremsstrahlung, and Compton scattering. The absorbed energy manifests as scintillation light, which is then detected by digital photomultipliers, subsequently amplified and digitizing the analogue signal. The digitized output is directly proportional to the energy of the particle initiating the electromagnetic shower within the crystal.

The radiation length and Moliere radius describe the longitudinal and transverse profiles of the EM shower within a specific material medium. The dense PbWO_4 crystals are characterized by short radiation lengths and Moliere radius, allowing the creation of a compact yet highly granular ECAL design.

The ECAL, as shown in Fig. 2.7, is structured into three sections: the barrel (EB) and two endcaps ($\text{EE}^\pm$). The crystals are arranged cylindrically around the tracker and closed by two disk-shaped crystal arrangements on both sides of the cylinder. In the barrel region, there are 61200 crystals organized in a grid of 170×360 within $|\eta| < 1.48$ at a radial distance of 129 cm. To prevent particle leakage through gaps between the crystals, the central axis of the crystals is tilted by 3° from the origin. The endcaps cover

the pseudo-rapidity range $1.48 < |\eta| < 3.0$ and begin at $z = \pm 315.4$ cm. Each endcap comprises 7324 crystals in two halves arranged in the transverse plane with the shape of an annular disk.

ECAL crystals have a length of about 22 cm, corresponding to a radiation length of 25.8, and are oriented such that the shorter face points towards the collision vertex. The crystals have a trapezoidal geometry. The dimensions of the faces vary, with one face measuring between $22 \times 22 \text{ mm}^2$ and $28.6 \times 28.6 \text{ mm}^2$, while the larger surface area face points radially outward and ranges in size from $26 \times 26 \text{ mm}^2$ to $30 \times 30 \text{ mm}^2$.

Pre-shower (ES) detectors, in front of the EE disks, increasing by three units the radiation length in the region $1.653 < |\eta| < 2.6$ and allowing the discrimination of a single high-energy photon from two angularly close photons. Each ES is made of a 20 cm thick sampling calorimeter constructed from two layers of alternating lead absorbers and silicon sensors. These silicon sensors measure the energy deposition and transverse shower profile.

The energy resolution of the CMS ECAL can be expressed by the following equation:

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2. \quad (2.6)$$

In this context, the energy is expressed in GeV, S is expressed in $\text{GeV}^{-1/2}$, N is GeV and C is adimensional. $S = 2.8\%$ represents a stochastic factor attributable to fluctuations in the count of photo-electrons, $N = 12\%$ accounts for noise in the ECAL electronics, and $C = 0.3\%$ includes any constant effect on the ECAL resolution, such as calibration effects [48].

2.2.6 Hadronic calorimeter

The CMS hadronic calorimeter (HCAL), is designed for the measurements of hadrons. It is situated within the solenoid magnet and enclosing the ECAL, is designed to stop or halt hadrons by absorbing their energy as they pass through its dense material layers, allowing for accurate measurement and identification of these particles [49]. It allows the measurement of the energy of all particles present in an event. The geometrical design of the HCAL is constrained by the radial dimensions of the solenoid (3.15 m) and the outer boundary of the ECAL (1.77 m), combined with the minimum required interaction length to effectively capture the characteristics of hadronic showers. The interaction length (λ_{int}) is the parameter that characterizes the longitudinal spread of the hadronic shower shape. This detector must exhibit high resilience to radiation. Thus, a sampling calorimeter is chosen as the preferred design, constituted of layers of brass plates (70% Cu and 30% Zn) and tiles of scintillating plastic.

The total absorber thickness in the barrel varies from 5.4 interaction length at $\eta = 0$ to 10.3 at $\eta = 1.3$, while in the endcaps it is about $10 \lambda_{int}$. The measured energy resolution is

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + C^2, \quad (2.7)$$

where the stochastic coefficient $S = 84.7\%$ is due to the fluctuation of the hadronic shower in the active HCAL material and the term $C = 7.4\%$, comes from constant effects, such as calibration.

2.2.7 Muon system

Muons, alongside neutrinos, have the unique feature among standard model particles of not being stopped by either the ECAL or the HCAL. Positioned beyond the solenoid, the muon detectors have the task of identifying and precisely measuring the momenta of muons [50]. The muon detector is constituted of four subsystems, all employing gaseous detector technologies. The four subsystems are: drift tubes (DTs), cathode strip chambers (CSCs), resistive plate chambers (RPCs), and gas electron multipliers (GEMs). These detectors are strategically positioned throughout various layers of the detector. Each subsystem utilizes distinct detector techniques to yield accurate and reliable measurements of muons. Collectively, the four muon subsystems cover a range of $|\eta| < 2.4$. Despite crossing the muon detectors, high-energy muons continue their path unstopped, ultimately escaping the CMS detector.

2.2.8 Trigger system

In a single LHC bunch crossing, the CMS subdetectors produce on average 1 MB of data, which multiplied by the 40 MHz bunch crossing frequency gives an output rate of 40 TB/s. This amount of data is too large to be stored and it would be too computationally expensive to reconstruct and analyse all this data. Also, the electronics which read data from the detectors can't work at this high frequency and can't handle a transmitting rate of 40 MHz. Moreover, most of this data is not interesting for physics analysis purposes. It is therefore fundamental to select only events containing interesting processes. This is done at CMS with a trigger system divided into two stages: the Level-1 (L1) Trigger and the High-Level Trigger (HLT).

The Level-1 Trigger

The L1 trigger is the starting point of the CMS trigger system and reduces the rate from the 40 MHz of the bunch crossing to 100 kHz, producing in output a data flow of 100 GB/s. To ensure fast processing, algorithms run on special processor boards equipped with field programmable gate arrays (FPGA) close to the detectors, in an

adjacent cavern. By design, the latency of the L1 trigger, i.e. how quickly the system provides a trigger decision, is about $4\ \mu\text{s}$. However, the time needed for the L1 trigger calculations is only about $1\ \mu\text{s}$. The rest of the time is spent on receiving and sending data between the two adjacent caverns.

In order to decide whether an event should be kept for further analysis, the L1 trigger uses simplified data from the fastest subdetectors, such as DT and CSC segments from muon subdetectors and low granularity information from the calorimeters. In this way the L1 trigger gets a basic understanding of the event, identifying particles like photons, electrons, muons and jets. Currently, tracker information is not used at the L1 trigger because reading and processing all the tracker channel within the tight time constraints of the L1 trigger is impossible within the current detector design. However, tracker data will be included in the future upgrade of the detector, as detailed in Chapter 8. All this information is combined in the Global Trigger and decisions are taken depending on a trigger menu, i.e. a list of possible trigger conditions, which can change reflecting the desired physics needs. Events that pass the L1 trigger are sent from the detector to the Data Acquisition (DAQ) system, where the HLT makes the final decision on whether to keep or reject the events.

The High Level Trigger

The target of the HLT is reducing the rate from 100 kHz of the L1 trigger output to 1 kHz. In order to meet this goal it exploits the computational power of a farm of regular CPUs. On average, it takes about 100 ms to reconstruct one event, combining (full granularity) raw data from all the sub-detectors. The HLT starts exploiting information from subdetectors that need less computing power, such as calorimeters and muon detectors. Once most events are filtered, the track reconstruction, which is more computationally expensive, can take place. Tracking information in the HLT allows us to perform simplified versions of the offline particle reconstruction. Tracker data also allows for precise localization of primary and secondary vertices, which is crucial for identifying jets from b quarks (b -tagging).

2.3 Physics Object Identification and Reconstruction

2.3.1 The Particle Flow algorithm

The offline event reconstruction exploits the Particle Flow (PF) algorithm which integrates information from all the subdetectors in order to measure the properties and identify all the final-state particles, such as muons, photons, electrons, charged and neutral hadrons. The PF algorithm is composed of three main stages subsequently described. At the beginning, the basic PF elements are reconstructed in each subdetector.

Secondly, multiple PF elements which are recognized to have a common origin are linked together to form PF blocks. Finally, particles are identified from PF blocks.

Basic PF elements which are built are: tracks from charged particles in the inner tracker, clusters of energy deposits in the ECAL and HCAL and tracks in the muons detector (DT, CSC and RPC). This first phase needs to account for additional complexity due to electrons which lose energy through bremsstrahlung, photons that can convert electron pairs and hadrons which can have nuclear interaction with the detector material producing secondary particles. In the next step, inner tracks are linked with energy clusters along their trajectory, with tracks in the muon detector and together when compatible with coming from a common secondary vertex. The latter may arise due to photon conversion, bremsstrahlung emission or nuclear interaction with detector material. When a ECAL cluster is contained in a HCAL cluster, these are also linked together. Particles are identified from the PF blocks thus constructed and the used elements cannot be used again. Muons are identified when an inner track is matched with a track piece in the muon detectors. If the muons are not isolated or are linked to calorimeter clusters, further requirements are applied since a charged hadron could penetrate the HCAL, producing such a signature. If an inner track is matched with a ECAL or HCAL cluster, an electron or a charged hadron is identified, respectively. All the remaining ECAL and HCAL clusters are identified as photons and neutral hadrons, respectively. The reconstructed particles are used to find interaction vertices and identify hard-scattering collisions. PileUp mitigation algorithms are used to remove tracks that are not compatible with coming from the primary vertex. Instead neutral objects, lacking tracking information, are weighted depending on the proximity of non-pileup tracks, which allows estimating the probability of coming from the primary vertex.

2.3.2 Jets

In high-energy interactions, partons like quarks or gluons rapidly undergo fragmentation and hadronization. This process results in the formation of a clustered group of hadrons known as a jet. Essentially, a jet is a complex object resulting from a clustering algorithm. The measurement of a jet's energy and trajectory is an attempt to reconstruct the energy and trajectory of the initial parton.

Jet reconstruction

In the offline reconstruction jet reconstruction is crucial. There are three main sequential clustering algorithms: the k_T [51], Cambridge/Aachen [52], and anti- k_T [53]. They utilize different exponents (k) in the expressions for inter-particle distance (d_{ij}) and in-

dividual particle distance (d_{iB}):

$$d_{ij} = \min(p_{Ti}^{2k}, p_{Tj}^{2k}) \frac{\Delta R_{ij}^2}{R^2}, \quad (2.8)$$

$$d_{iB} = p_{Ti}^{2k}, \quad (2.9)$$

where R is the jet radius, ΔR_{ij} is the distance in the $\eta - \phi$ space between the particle i and j , defined as $\Delta R_{ij} = \sqrt{(\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2}$, and p_{Ti} is the transverse momentum of particle i . The clustering algorithms sequentially run over all the possible pairs of particles and evaluate the minimum between the inter-particle distance, d_{ij} (in Eq. 2.8), and the two individual particle distances, d_{iB} (in Eq. 2.9). If the first is the minimum then the two particles are merged to form a new single particle and the algorithm continues evaluating new pairs. Instead, if one of the individual particle distances is the minimum, the associated particle is identified as a jet and removed from the collection. The kt algorithm employs $k = 1$ to prioritize combining soft and collinear particles, the Cambridge/Aachen algorithm uses $k = 0$ to recombine particles based solely on their angular distance, and the anti- kt algorithm uses $k = -1$ to combine the hardest particles first. The results of these three clustering algorithms applied to an example dataset are shown in Fig. 2.8. At the CMS experiment, the most widely used definition of jet uses the anti- k_T algorithm (AK) with a jet radius R which can have a value of 0.4 (AK4) or 0.8 (AK8). The jets reconstructed with the AK8 algorithm are also referred to as fat jets. The scalar sum of jets p_T in an event, also known as H_T , is a typical jet-related quantity used at the CMS experiment both at trigger level and analysis offline selections (including the one described in Chapter 6).

Jet b -tagging

Identifying jets that originate from b quarks, known as b -jets, is of significant importance for probing both SM and BSM physics. This is due to the frequent presence of b quarks in various physics processes, such as those involving the decay of the Higgs boson or top quark. When b -quarks are generated in high-energy interactions, the result of the hadronization process often includes a b -hadron. The b -jets, illustrated in Fig. 2.9, have distinctive features that can be used for their identification.

The average lifetime of hadrons containing b quarks, is typically around 1.5 ps. This allows them to travel a few centimetres before decaying leading to the formation of a secondary vertex, which is a key feature exploited in flavour tagging techniques. Due to the mass of the b quark (about 5 GeV), the decay products of a b -hadron tend to have a higher transverse momentum than other constituents within the jet. Consequently, b -jets often exhibit wider profiles. Additionally, b -jets include an electron or a muon about 20% of the time. Algorithms for b -tagging are employed to distinguish b -jets from those originating from lighter flavour quarks. These algorithms assign a

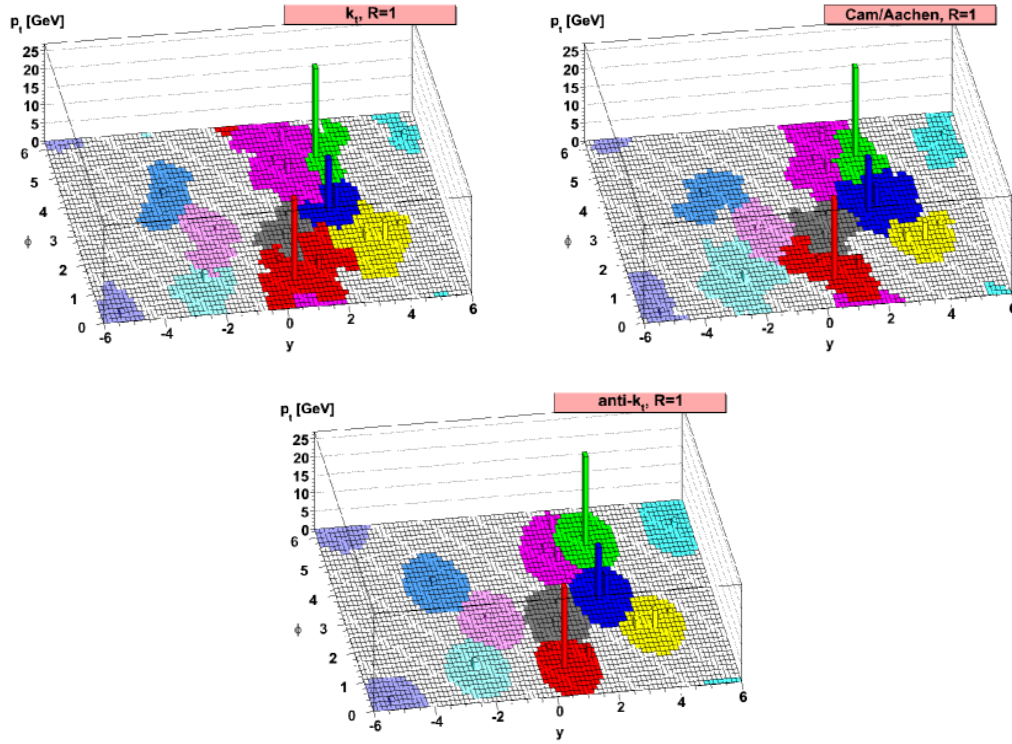


FIGURE 2.8: The results of the three main jet reconstruction algorithms, performed on the same data and with the same radius R , is shown. Different colours are used to represent the different jets. The k_T algorithm produces highly irregular clusters since it prioritizes combining soft and collinear particles. The Cambridge/Aachen algorithm neglects momentum information and looks only at angular distances. The anti- k_T algorithm produces regular clusters, conically shaped, centred around the hardest particle. Figure from Ref. [10].

numerical discriminator to each jet, indicating the likelihood of it being a b -jet. CMS employs several b -tagging algorithms, with the most effective ones utilizing multivariate techniques that consider information about displaced tracks and secondary vertices associated with the jet axis.

Heavy neutral particles sometimes decay into a $b\bar{b}$ pair. When the momentum of the original particle is high, this results in a boosted $b\bar{b}$ pair, appearing as a single wider jet. While a single quark's hadronization products are typically clustered by an AK4 jet, a boosted $b\bar{b}$ pair is usually represented by an AK8 jet. To identify a boosted double- b jet, jet substructure techniques are applied. These techniques involve analyzing the subjects within the jet.

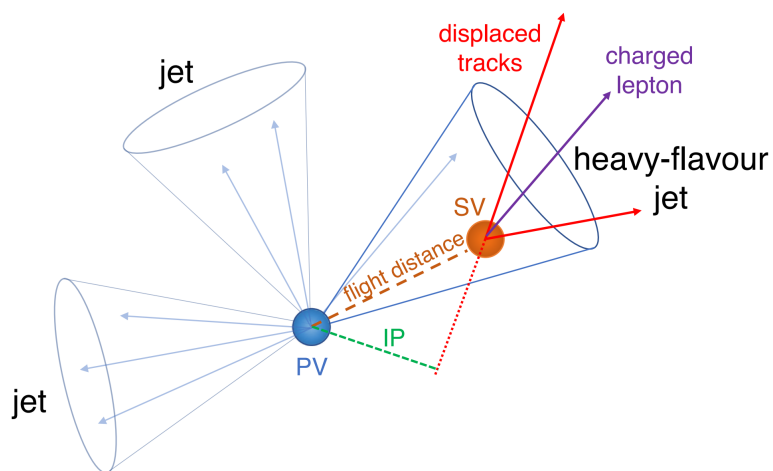


FIGURE 2.9: Sketch of a heavy-flavour jet and its main features. The jet results in a cluster of charged particles (often including a soft lepton) coming from a secondary vertex. Figure from Ref. [10].

Part II

The CMS Electromagnetic Calorimeter: transparency corrections updates

Chapter 3

Transparency corrections updates in the Electromagnetic Calorimeter

The electromagnetic calorimeter (ECAL) crystals undergo radiation damage with time causing a loss of light transmission. In this chapter, a correction strategy for the transparency of the ECAL crystals is presented.

As already described in 2.2.5, the ECAL of CMS, is a hermetic homogeneous calorimeter made of 61200 lead tungstate (PbWO_4) crystals mounted in the central barrel part (EB), closed by 7324 crystals in each of the two endcaps (EE). A pre-shower detector is placed in front of the endcap crystals. Avalanche photo-diodes (APDs) are used as photo-detectors in the barrel (shown in Fig. 3.1) and vacuum photo-triodes (VPTs) in the endcaps. The use of high-density crystals has allowed the design of a compact and granular calorimeter. The fast response and the radiation resistance of PbWO_4 crystals are important characteristics of ECAL to withstand the harsh LHC environment.



FIGURE 3.1: PbWO_4 crystal with APD photodetectors attached. The inset shows a capsule with two APDs. Figure from Ref. [11].

3.1 Crystal transparency loss mechanisms

The ECAL crystals have to withstand the radiation levels and particle fluxes throughout the operation of the LHC machine. It has been shown that irradiation does not affect the scintillation mechanism and that the intrinsic energy resolution is not degraded. Nevertheless, radiation damage affects the crystal transparency through the formation of colour centres due to oxygen vacancies and impurities in the lattice. There are two components associated with this damage:

- **Electromagnetic damage** is a dose-rate effect (depends on LHC instantaneous luminosity) and partially anneals at room temperature during beam-off periods
- **Hadron damage** depends on the total dose and does not anneal at room temperature

The practical consequence of this damage is a wavelength-dependent loss of light transmission, causing a decrease in the amount of collected light. The reason why electromagnetic damage anneals at room temperature, whereas hadron damage does not, lies in the nature and energy of the interactions causing the damage. Electromagnetic damage primarily results from ionizing radiation, such as photons and electrons, which create electron-hole pairs in the crystal lattice. These pairs can recombine over time, and the process is facilitated by thermal energy at room temperature. The mobility of these defects allows them to move and recombine, thus partially restoring the crystal's transparency. Hadron damage, on the other hand, is caused by highly energetic hadrons (protons, neutrons, etc.) that induce displacement damage in the crystal lattice. These interactions create more stable and complex defect structures. The energy barriers for the movement and recombination of these defects are significantly higher, making thermal annealing at room temperature ineffective.

3.1.1 Laser monitoring system

At the ECAL operating temperature 18°C, the balance between damage and annealing results in a dose-rate dependent equilibrium of the optical transmission. There is a cyclic transparency behaviour between LHC collision runs and machine refills. The magnitude of transparency changes is dose-rate dependent, and over a period of two weeks, it is expected to range from 1–2% in the barrel, to about 10% in the high η regions of the endcap where the radiation dose is higher. The performance of the calorimeter would be unacceptably degraded by these radiation-induced transparency changes if they are not measured and corrected.

A dedicated laser system injects laser light to the crystals and monitors the transparency of the crystals over time [9].

Figure 3.2 illustrates the primary components of the laser-monitoring system [11] [54]. Two laser wavelengths serve as the fundamental source: one is blue at $\lambda = 440$ nm, closely matching the scintillation emission peak, which is utilized to monitor transparency variations caused by radiation. The other, in the near-infrared range at $\lambda = 796$ nm, is distant from the emission peak and minimally influenced by transparency changes, making it suitable for assessing the stability of other system components. The laser pulses typically last around 30 ns and can be operated at a frequency of approximately 80 Hz. The blue laser pulse energy corresponds to ~ 1.3 TeV. The system includes two blue light sources. The redundancy ensures fault tolerance and allows for maintenance on one source while the other remains operational. The monitoring light pulses are transmitted through an optical fiber network. A fiber optic switch located at the laser routes the pulses to one of 88 calorimeter regions. Each calorimeter region is equipped with a two-stage distribution system that directs the light to individual crystals. The relative calibration of the light injected into each group of crystals is accomplished by simultaneously injecting light onto reference PN silicon photodiodes that are part of the distribution system. The design includes a redundancy of these reference photodiodes.

The transparency corrections are calculated in the following way. The laser system sends light pulses through the crystals. APDs photodiodes measure the intensity of light passing through the crystals. The detected signals are compared with reference measurements taken by silicon PN photodiodes, which indicate the expected signal intensity without transparency effects. Thus $R(t) = APD(t)/PN(t)$ is used as the measure of the crystal transparency. The relationship between the changes can be expressed by a power law,

$$\frac{S(t)}{S(t_0)} = \left[\frac{R(t)}{R(t_0)} \right]^\alpha$$

where $S(t)$ represents the response to scintillation light and α is characteristic of the crystal and t_0 refers to early 2011 before any substantial losses occurred.

3.1.2 Response losses measured in LHC Run 1 and Run 2

The transparency loss measured during Run 1 and Run 2 is shown in Fig. 3.3. The bottom panel of the figure shows how the LHC instantaneous luminosity has gradually increased between Run 1 and Run 2. Thus we can see the gradual decrease in the response over time due to radiation effects. For the barrel region, where the radiation levels are comparatively lower, the loss of ECAL response is about 10% over the last 10 years. However, at the high η regions of the endcaps, the loss of response is as large as 95%. In order to account for this effect, a set of corrections based on the laser data is applied during data-taking to restore the correct energy scale for electromagnetic objects (photons and electrons) and jets.

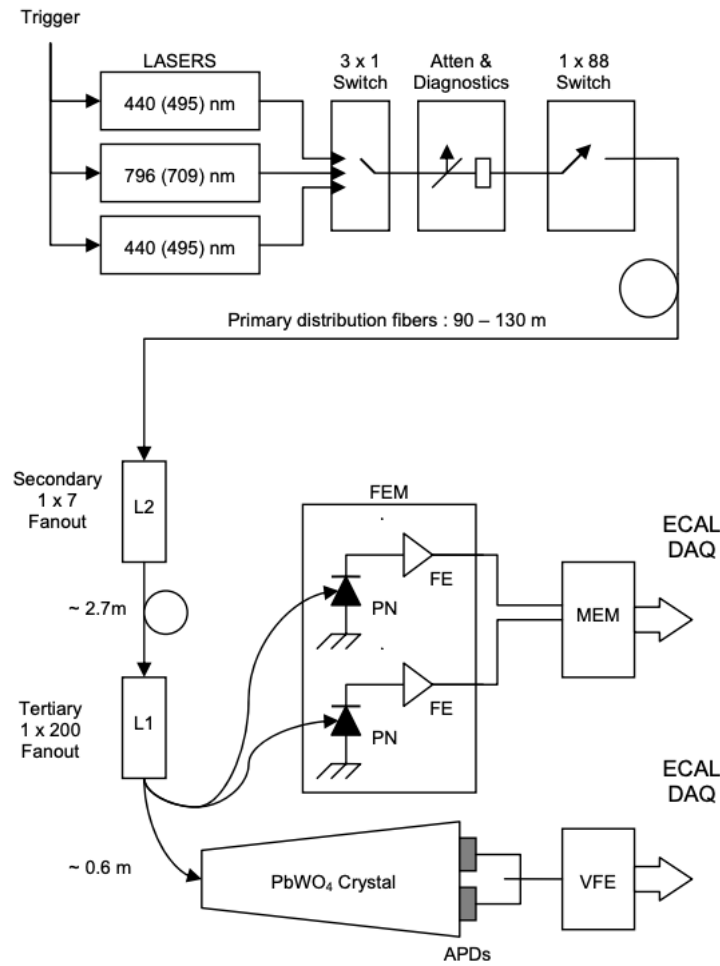


FIGURE 3.2: Schematic of the ECAL laser monitoring system. Figure from Ref. [11].

3.2 Transparency corrections

The laser monitoring system provides a measurement of the transparency of all 75848 crystals every 45 minutes during data taking. These measurements are processed and a series of corrections are derived to account for the transparency losses and stabilise the response of the crystals. The corrections are validated and inserted into the CMS offline database (DB), where they are used for the prompt reconstruction of ECAL data which occurs 48 h after the data are recorded.

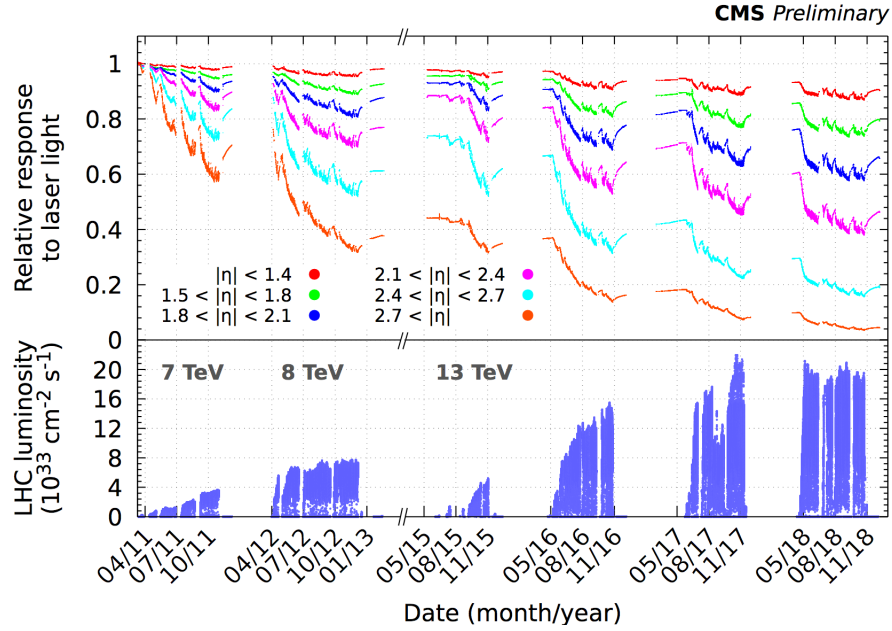


FIGURE 3.3: Top: Relative ECAL response to laser light over the period 2011-2018, averaged in bins of pseudorapidity η . Bottom: LHC instantaneous luminosity which has increased during the years. Figure from Ref. [12].

3.2.1 L1/HLT corrections history in Run 1 and Run 2

The laser corrections have been applied at L1 and HLT since 2012. The corrections that are used at L1/HLT take values from previous runs¹, whereas the offline reconstruction can use data from the existing run, due to the deferred processing of prompt reconstruction data. Hence, the L1 and HLT corrections can only be applied by starting a new run, and they are updated during an interfill period.

The ECAL Trigger Primitives (TPs) are the energy and timing information of the ECAL energy deposits, in regions of 5–25 crystals called “trigger towers” recorded during LHC collisions, that are transmitted to the Level-1 trigger system at a rate of 40 MHz. The L1 corrections are applied to the ECAL TPs via linearization constants, which are computed per crystal and applied in the front-end electronics while computing the trigger primitive energies [55]. At HLT, the transparency corrections are applied to each hit used in the reconstruction of the final object. The update frequency at L1/HLT is limited by the time of availability of the laser data in the DB, which, in Run 2, was approximately 6-7 hrs on average with long tails.

The frequency of the updates of the transparency corrections at L1/HLT over the years is as follows:

¹The term run refers to a session of the data acquisition software. Typically a run is started when beams reach stability after a LHC fill (i.e. the procedure of filling the LHC rings with beam bunches), and is ended when beams are dumped.

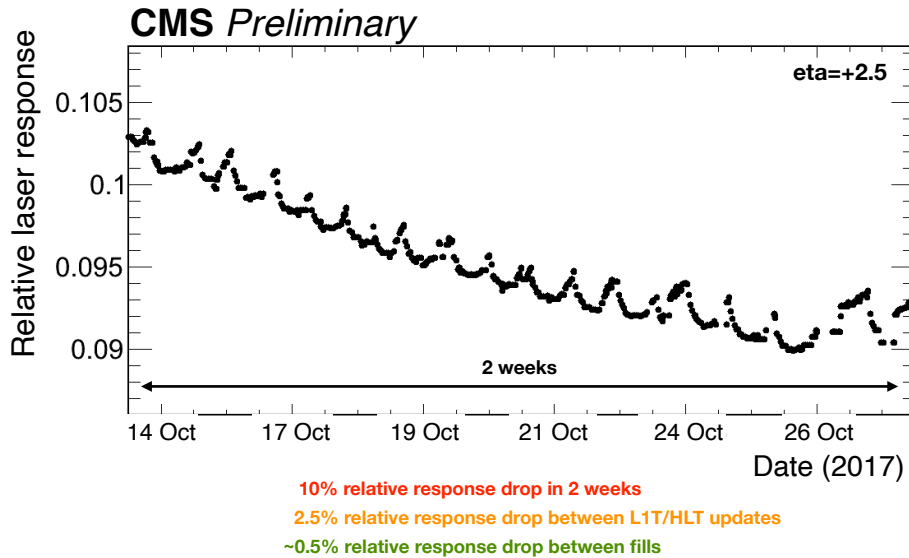


FIGURE 3.4: Laser response over two weeks from October 14th 2017 to October 26th 2017, for a single crystal in the endcap (EE) at $\eta=+2.5$.

- Run 1 (2012): the corrections were applied to 22 individual η rings in the endcaps only, once a week;
- Run 2 (2015-2016): the corrections were extended to the barrel part of the ECAL and were applied with increased (per-crystal) granularity, once a week;
- Run 2 (2017-2018): the correction frequency was increased to twice a week.

3.2.2 Planned L1/HLT update frequency in Run 3

This work supported the proposal for Run 3 to perform a once-per-LHC-fill update of the L1/HLT corrections. Updating the corrections more frequently should make them closer to those used in the offline reconstruction. Figure 3.4 shows the response of one endcap (EE) crystal ($\eta = +2.5$) over a two week period. We can see a drop of 10% for the full period which corresponds to a 2.5% drop between corrections applied twice-per-week, as compared to a drop of 0.5% between each fill. This clearly indicates the benefit of updating the transparency corrections more frequently. In order to verify the stability of the L1 and HLT rates, the ECAL TPs are re-emulated with the twice-per-week and once-per-LHC-fill transparency corrections. Re-emulation is performed by exploiting raw 2018 data, collected with zero bias trigger lines, and recomputing offline the output of L1 and HLT trigger algorithms.

3.3 Impact on ECAL energy scale

In order to gain better insights into how the more frequent transparency corrections can improve the stability of L1/HLT rates, the impact on the ECAL energy scale was first investigated, by comparing the median of the di-electron invariant mass ratio for $Z \rightarrow ee$ decays for the two laser transparency update frequencies: twice-per-week (Run 2) and once-per-LHC-fill (proposed for Run 3) compared to the results obtained with the full refined offline calibration. These are plotted in Fig. 3.5 for several LHC fills, between the 15th of July and the 22nd of July 2018. The once-per-LHC-fill corrections provide a more stable energy scale compared to the twice-per-week corrections and this is visible especially in the last four fills. What is therefore expected by applying the corrections more frequently is a more stable energy scale that will result in more stable and more representative L1/HLT rates.

3.4 Impact on L1 and HLT rates

3.4.1 L1 rates

First, the impact of more frequent transparency corrections on L1 rates is studied. In order to do this the ECAL TPs are re-emulated with the once-per-LHC-fill and twice-per-week laser corrections and the L1 seeds are re-emulated using these updated ECAL TPs.

The same data-taking period as shown in Fig. 3.5 is used for a period of one week in July 2018, corresponding to the LHC fills: 6923, 6931, 6040, 6942, 6956, 6957, 6960, 6961. The Level-1 seed that has been studied is `L1.SingleIsoEG28er2p5`, corresponding to isolated electron/photon candidate with transverse energy 28 GeV and with $|\eta| < 2.5$. The effect on the total rate seeds is checked on `L1APhysics`, which is the total accepted Level-1 rate from LHC collisions. To study the impact on Level-1 rates for two different transparency conditions, the absolute Level-1 rates for each fill is plotted. Figure 3.6 shows the variation of rates for the Level-1 single electron/photon seed. There are visible differences between the two sets of rates. However, these plots are not ideal to quantify any improvement in rate stability due to the large fill-to-fill variations in instantaneous luminosity and pileup (PU). The last four fills are the most representative ones as they have a similar beam intensity and fill duration.

In order to better understand the stability of the rates between different fills, the variation of rates vs. PU is studied. In Fig. 3.7, parabolic fits to the rate vs. PU plots are shown for 6956-6961 fills for L1 electron/photon candidates. The fit lines seem more tightly bunched for once-per-LHC-fill than for twice-per-week updates. In both cases, the fits disperse at higher PU values where there are larger fluctuations in rates. However, a slight improvement for the once-per-LHC-fill corrections can be observed. These

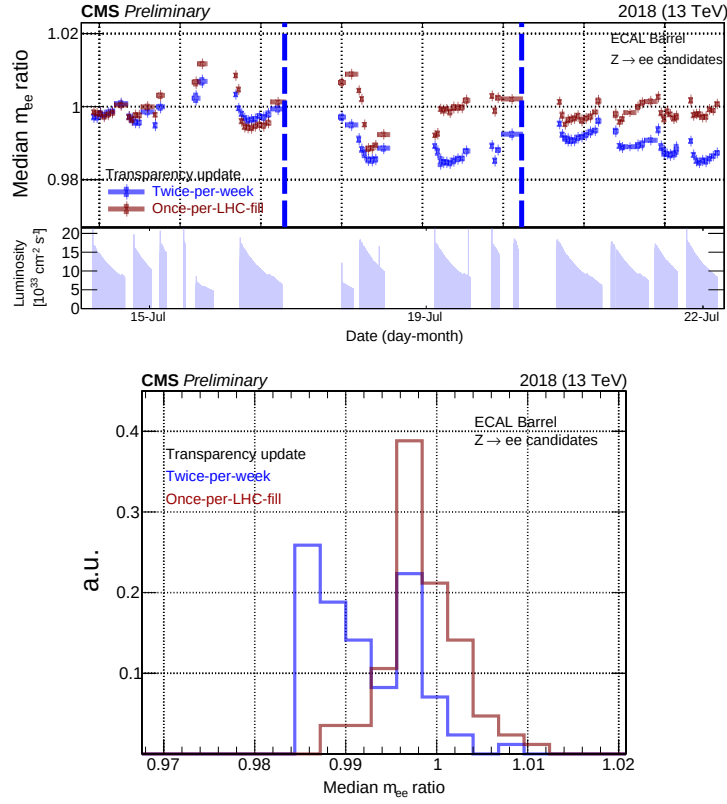


FIGURE 3.5: Impact of more frequent ECAL transparency updates on the time stability of the di-electron invariant mass m_{ee} distribution for CMS data taken during a period of one week in July 2018 using $Z \rightarrow ee$ events. Both electrons are required to be in the ECAL barrel. The top plot displays the ratio of the median m_{ee} calculated with the twice-per-week (blue) transparency corrections and once-per-LHC-fill (red) corrections, to the median m_{ee} calculated with the full refined calibration (corrections applied every 40 minutes). The two blue vertical dashed lines indicate when the twice-per-week update was applied. The panel at the bottom shows the corresponding history of the LHC instantaneous luminosity during this period, with a clear LHC fill structure. The bottom plot displays the distribution of the ratio of the median m_{ee} calculated with the twice-per-week (blue) corrections and the once-per-LHC-fill (red) to the median m_{ee} calculated with the full refined calibration for this data-taking period.

results give us a hint of the better performance of the more frequent updates. In order to provide a more representative comparison, the rates for each fill are normalised to an average PU value in the following way assuming that the rate is linear with PU (in the range 30-45). An average rate is defined as:

$$R = \frac{1}{N} \left(\sum_{i=30}^{45} \frac{R_i}{P_i} \right) \times \langle \text{PU} \rangle \quad (3.1)$$

where R_i and P_i are the rate and PU values for the i^{th} PU bin, and $\langle \text{PU} \rangle$ is the mean of the PU in the range 30-45 which is equal to 37.5. Therefore, there are 16 bins of PU with a width of 1. The errors on this average rate are calculated by taking the standard

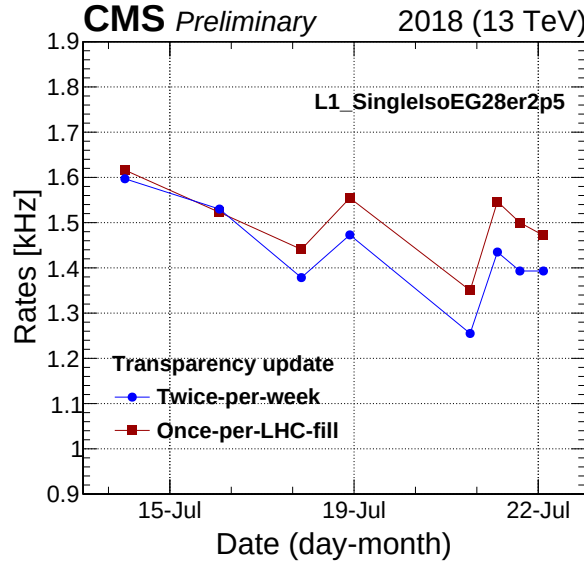


FIGURE 3.6: Variation of the L1_SingleIsoEG28er2p5 trigger rate for a one-week period in July 2018. The rates are emulated for two transparency update frequencies (once-per-LHC-fill shown in red and twice-per-week shown in blue). The L1 electron/photon candidate are chosen to have a transverse energy greater than 28 GeV, pass an isolation requirement, and be within $|\eta| < 2.5$.

error on the mean given by $\sigma/\sqrt{N}$, where σ is the standard deviation of the rates in the PU range 30-45. Thus $\sigma = \sqrt{\frac{1}{N} \sum_{i=30}^{45} (\frac{R_i}{P_i} \times \text{PU} - \bar{R})^2}$, where $\bar{R} = \frac{1}{N} \sum_{i=30}^{45} R_i$ is the mean rate and $N=16$.

Figure 3.8 verifies the behaviour of the average rate as defined in Eq. 3.1 for L1 electron/photon candidates for a one week period from the 15th to the 22nd of July 2018. The rates are consistently higher and more stable for once-per-LHC-fill updates which is in good agreement with the $Z \rightarrow ee$ plot in Fig. 3.5. Considering the last four LHC fills, the transparency is under-corrected when applying twice-per-week updates, causing a drop in the L1 rates, and the mean rate is 18.0 kHz for this time period, whereas; for the Once-per-LHC-fill updates, the L1 rate is more stable, and the mean rate is 19.8 kHz.

Figure 3.9 shows the behaviour of the average total Level-1 rate over the same time period. For the last four LHC fills, the Twice-per-week updates result in an underestimate of the transparency corrections, causing a drop in the total L1 rate, and the mean rate is 65.4 kHz. However, for the Once-per-LHC-fill corrections, the mean rate is higher, and has an average rate of 69.1 kHz in this time period.

Figure 3.10 shows the 1D profile distributions of the average rate for electron/photon candidates and the total L1 rate. These plots confirm a higher and more stable L1 rate with Once-per-LHC-fill corrections as compared to the previously used Twice-per-week updates.

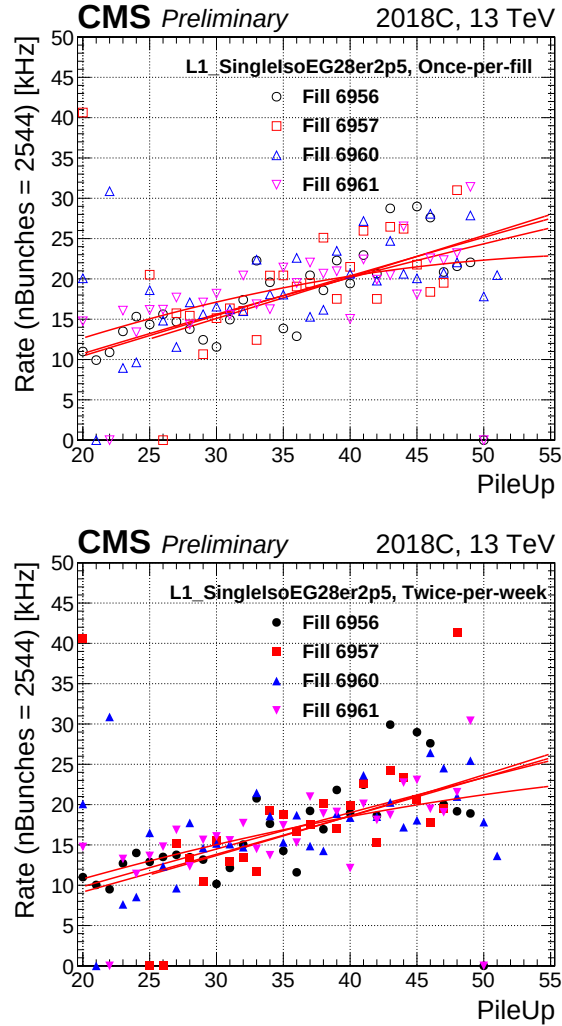


FIGURE 3.7: Rate vs PU plot for fill numbers 6956-6961. A fit for all the fills for seed L1_SingleIsoEG28er2p5 is presented for the once-per-LHC-fill (top) and the twice-per-week (bottom) updates.

3.4.2 HLT rates

Similarly, the stability of HLT rates is verified. In order to check the performance of more frequent transparency corrections on HLT rates a similar prescription to the one used for the Level-1 rate study described above was followed. HLT electron objects are re-emulated with the two transparency conditions and the rates for HLT electrons with $p_T > 30$ GeV are computed.

Figure 3.11 shows the ratio of the rates for once-per-LHC-fill (red) and twice-per-week corrections (blue) with respect to the full refined calibration. It can be clearly seen that there are significant improvements in HLT rates with the more frequent transparency corrections shown in red. The once-per-LHC-fill corrections yield a ratio that is closer to 1 and more stable as compared to twice-per-week corrections shown in blue.

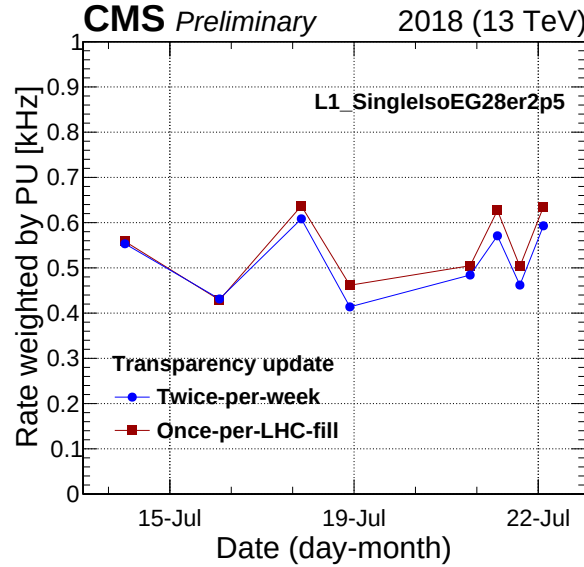


FIGURE 3.8: The plot displays the average L1 electron/photon rate calculated with the twice-per-week (blue) corrections and the once-per-LHC-fill (red) corrections for CMS data taken during a period of one week in July 2018. The electron/photon candidates are required to have a transverse energy greater than 28 GeV, pass an isolation requirement, and be within $|\eta| < 2.5$. The same dataset was used to produce the history of the L1 electron/photon rate using two different sets of ECAL transparency corrections to demonstrate the impact of more frequent transparency updates on the stability of the L1 rate. The ECAL transparency corrections are updated either twice-per-week, emulating the update frequency that was applied at L1/HLT in 2018, or once-per-LHC-fill, emulating the L1/HLT update frequency that is applied from Run 3 onwards.

3.5 Impact on L1/HLT resolutions

3.5.1 L1 resolution

We have also re-emulated L1 electron/photon candidates to study the impact of more frequent corrections on L1 energy scale and resolution. Figure 3.12 shows the L1 electron/photon energy scale which is defined as:

$$\frac{E_T^{e\gamma, L1}}{E_T^{e\gamma, offline}} \quad (3.2)$$

where E_T of the electron/photon candidate reconstructed at Level-1 and the denominator is the E_T of the matched offline electron.

The energy scale is obtained from $Z \rightarrow ee$ events where the L1 object is matched to an offline electron and the comparison of the two laser transparency update frequencies for the last four fills is made. In Fig. 3.12, the left plot shows the resolution obtained with once-per-LHC-fill corrections and the right plot shows the same result for twice-per-week corrections for the four consecutive LHC fills (6956-6961) as mentioned be-

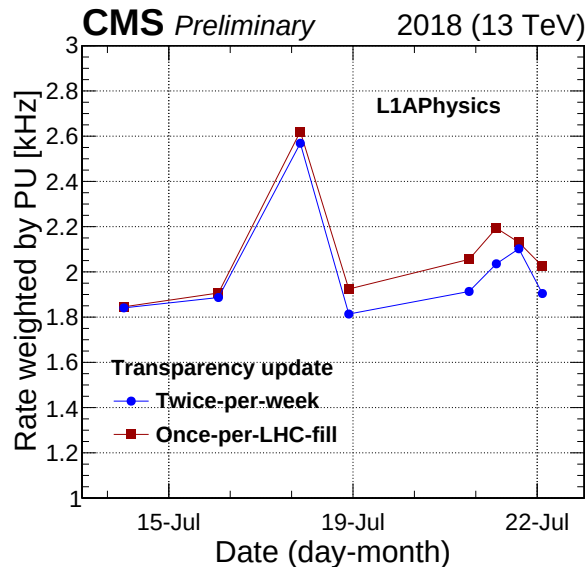


FIGURE 3.9: The plot displays the average total L1 rate calculated with the Twice-per-week (blue) corrections and the Once-per-LHC-fill (red) corrections for CMS data taken during a period of one week in July 2018. The ECAL transparency corrections are updated either twice-per-week, emulating the update frequency that was applied at L1/HLT in 2018, or once-per-LHC-fill, emulating the L1/HLT update frequency that is applied from Run 3 onwards.

fore. The mean values of these plots are quoted in Tab. 3.1, for both update frequencies. The mean value does not vary for once-per-LHC-fill updates, whereas a visible drift in the mean is observed for twice-per-week updates. This indicates that the L1 electron/photon energy scale will be more accurate and more stable if once-per-LHC-fill updates are applied.

TABLE 3.1: The mean and RMS of each of the gaussian distributions from Fig. 3.12.

Fill Number	Once-per-fill		Twice-per-week	
	Mean	RMS	Mean	RMS
6956	0.997	0.0343	0.998	0.0342
6957	0.997	0.0343	0.998	0.0341
6959	0.997	0.0345	0.997	0.0343
6961	0.997	0.0344	0.991	0.0338

3.5.2 HLT resolution

We define the resolution of HLT electron objects as the fractional difference $\left(\frac{\Delta p_T}{p_T}\right)$ between the HLT electron p_T calculated with the twice-per-week (blue) conditions or the once-per-LHC-fill (red) conditions, compared to the p_T calculated with the full refined calibration (Ultra-Legacy or UL). It is defined as:

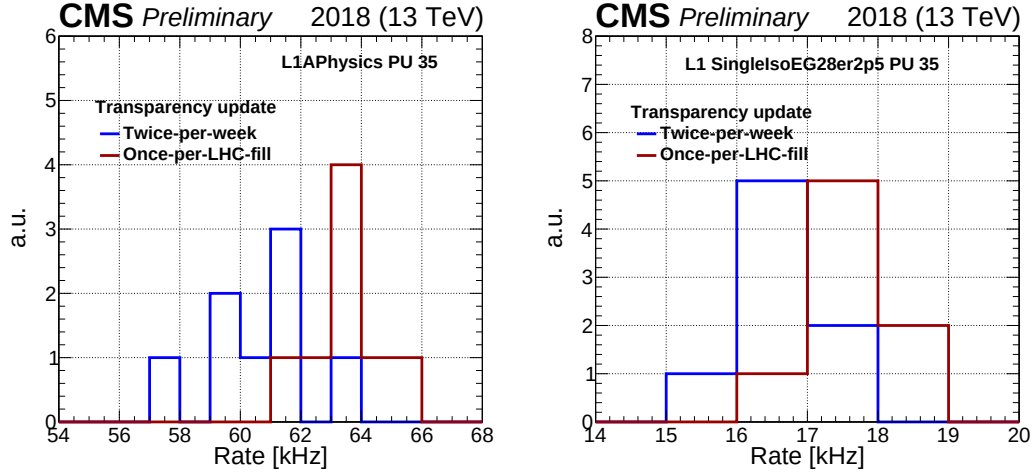


FIGURE 3.10: The same dataset as used in Fig. 3.9 was also used to produce the profile distribution of the L1 rate using two different sets of ECAL transparency corrections to demonstrate the impact of more frequent transparency updates on the stability of the ECAL energy scale and resolution. The plot displays the distribution of the median of average rate calculated with the twice-per-week (blue) conditions and the once-per-LHC-fill (red). The left plot shows the profile distribution of the total rate (L1APhysics) and the right plot shows the profile distribution of electron/photon rate (L1_SingleIsoEG28er2p5).

$$\frac{\Delta p_T}{p_T} = \frac{\text{HLT}_{p_T} - \text{UL}_{p_T}}{\text{UL}_{p_T}} \quad (3.3)$$

Figure 3.13 shows the resolution plot for once-per-LHC-fill and twice-per-week corrections. A substantial improvement in resolution is observed. The more frequent transparency corrections yield a distribution that is much closer to 0 with a narrower Full Width at Half Maximum (FWHM) as compared to the twice-per-week corrections which is further from 0 and has a larger FWHM.

3.6 Workflow of the transparency updates

In Run 3 we apply the transparency updates at the beginning of each LHC fill. This is made possible by two improvements: firstly a new laser data processing framework was developed during Long Shutdown 2 (LS2), which has significantly reduced the time delay between the time of the laser measurement and when the laser corrections are written to the DB. This latency has been reduced from an average of 6-7 hrs in Run 2 to less than 1.5 hrs in Run 3. This can be seen in Fig. 3.14, where the average latency of the “last” tag (processing of the corrections starts as soon as a laser sequence is complete), is measured to be 1h 15 mins. In the same figure, the average latency of the “prompt” tag (the processing waits until one more laser sequence has completed, one laser sequence takes ~ 45 mins to complete) is ~ 2 hrs. The second improvement is that the three-step validation procedure, which took 3-4 hrs in Run 2 and was triggered

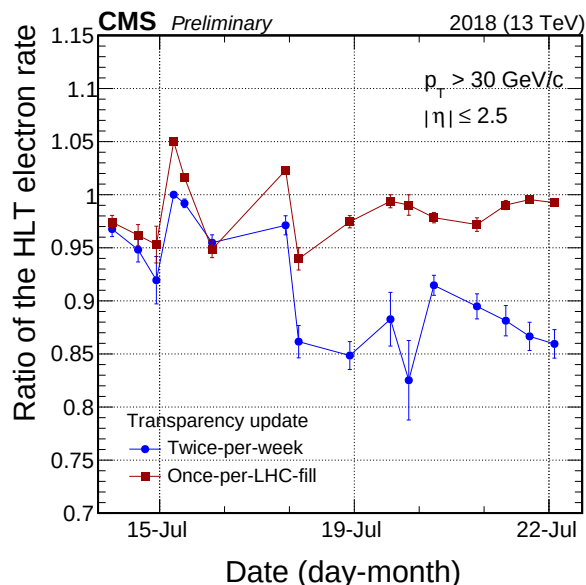


FIGURE 3.11: Time stability of the rate of electrons reconstructed at HLT with $p_T > 30$ GeV per LHC fill, for CMS data taken during a period of one week in July 2018. The first set of corrections corresponds to the refined calibration performed in 2019, where the ECAL transparency corrections are updated every 40 minutes. For the other two sets, the ECAL transparency corrections are updated either twice-per-week, emulating the update frequency that was applied at L1/HLT in 2018, or once-per-LHC-fill, emulating the L1/HLT update frequency that is applied in Run 3 (2022 onwards). The plot displays the ratio of the HLT rates calculated with the twice-per-week (blue) conditions and the once-per-LHC-fill (red) to the rate calculated with the full refined calibration for this data-taking period.

manually, is now fully automatic. Since the validation jobs now run on dedicated 32 and 64 core machines, they run in less than 1 hr. Both of these improvements now make it possible to update each LHC fill, which typically lasts 10-12 hrs on average.

The validation procedure for Run 3 is explicitly governed by the start and end of LHC fills. The DB is a collection of tags used to point to stored detector conditions evolving with time and valid for an interval defined either with timestamps or run numbers. A payload which is a set of values of the transparency corrections is stored in the DB at the “last” or prompt “tag”. It’s (payload=conditions=values of corrections) Interval of Validity (IoV) is defined from the time of its insertion (rather measurement) until a new payload is inserted. The IOV that is used for the validation is the first available IOV at least 2 hrs from the start of stable beams. The reason behind choosing the 2 hrs buffer (corresponding to 270 Lumisections) is due to the fact that at this point the transparency of the crystals has stabilized after the fast initial drop at the start of the fill, and is, therefore, a more representative measurement of the transparency within the fill. The validation process starts once the chosen IOV has been written to the DB, which should occur around 3h 15m after the declaration of stable beams.

This period of 3h 15m is unavoidable, thus we need to perform the validation process as quickly as possible, ideally in less than 1 hr, so that after $\sim 4 - 5$ hrs from the

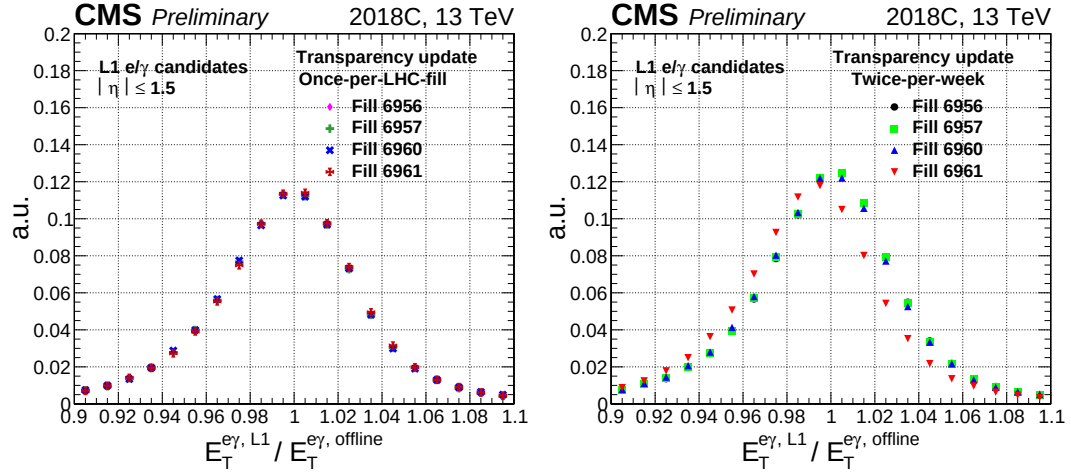


FIGURE 3.12: The plots display the ratio between the L1 electron/photon E_T and the offline electron/photon E_T for the once-per-LHC-fill conditions (left) or the twice-per-week conditions (right) for CMS data taken during the 21st and the 22nd of July 2018 corresponding to four consecutive LHC fills 6956-6961. The same dataset was used to reconstruct L1 electron/photon candidates using two different sets of ECAL transparency corrections, to demonstrate the impact of more frequent transparency updates on the stability of the ECAL energy scale and resolution. The electrons are required to be within pseudo rapidity $|\eta| \leq 1.5$.

beginning of the fill, we will be prepared with the new condition for the next fill.

The three validation processes that are required to run are:

- Validation of trigger primitives (TPG validation)
- Validation of Level-1 rates (L1 validation)
- Validation of high level trigger rates (HLT validation)

The tool that allows us to run the validation steps automatically is called Jenkins [56], and dedicated processes run to:

- detect when a suitable IOV is available within a LHC fill
- run the validation jobs
- collate the output in a set of slides containing L1/HLT rates and TP occupancy and transverse energy distributions

These slides are checked by the ECAL trigger shifters, and automatic sanity checks of the difference in HLT rates between one set of corrections and the previous validated set are made. If the fill is sufficiently long such that by the end of the fill a new set of validated conditions are already available, these conditions are then uploaded in the configuration and condition DBs.

Thus the whole validation process can be listed in the following steps:

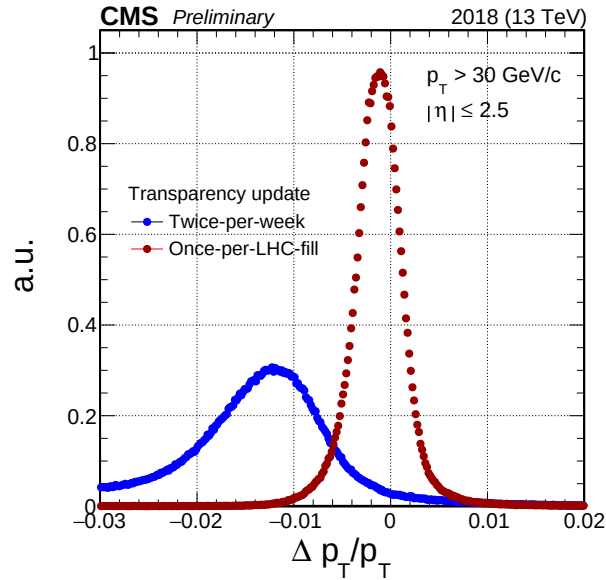


FIGURE 3.13: The p_T resolution obtained for electrons reconstructed at HLT with $p_T > 30$ GeV, for CMS data taken during a period of one week in July 2018. The plot displays the fractional difference between the HLT electron p_T calculated with the twice-per-week (blue) conditions or the once-per-LHC-fill (red) conditions, compared to the p_T calculated with the full refined calibration for this data-taking period.

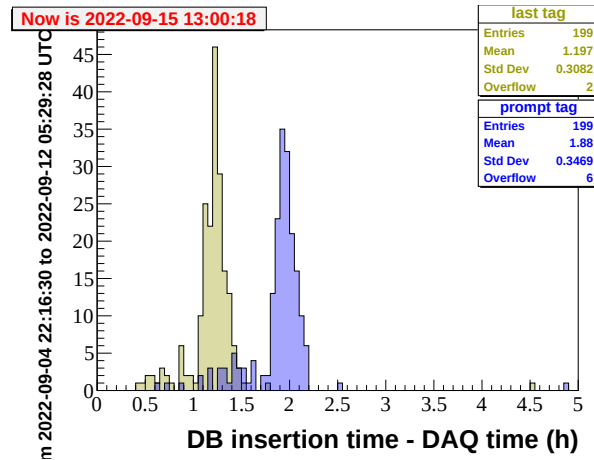


FIGURE 3.14: The plot shows the latency of the insertion time of last tag (the processing of the corrections starts right after the completion of each laser sequence) and the prompt tag (the processing waits until one more laser sequence has been completed) in the DB for a one week period in September 2022.

- Beginning of a fill
- Stable beams declaration
- Laser condition taken after 2 hrs from the start of stable beams
- Laser condition arrives at the DB after another 1h 15m
- Three step validation begins in parallel, lasts for 20-30m

- New validation condition available after ~ 4 hrs from the start of stable beams
- Upload in p5 of conditions DB (serving HLT) at beam dump (end of the fill), and uploaded configuration DB (ECAL TPs). ECAL front-end electronics is configured with new corrections following GREEN recycle, and conditions DB is updated.

3.7 Procedure implementation in Run 3

The commissioning of the once-per-LHC-fill update procedure started at the beginning of 2022 and the validation chain has been successfully commissioned prior to the start of stable beams LHC collisions in July 2022. From March 2022, validation jobs were performed during many LHC test fills to test the performance of the procedure and measure the latencies. Figure 3.15 shows the results of the latencies for fill numbers from 7650 to 8020, taken between May and July 2022. The Jenkins jobs were all successfully finished within the target time of 5 hrs. A few peaks have been observed mainly due to some delays of the laser data to arrive or rare delays in the start of the Jenkins validation.

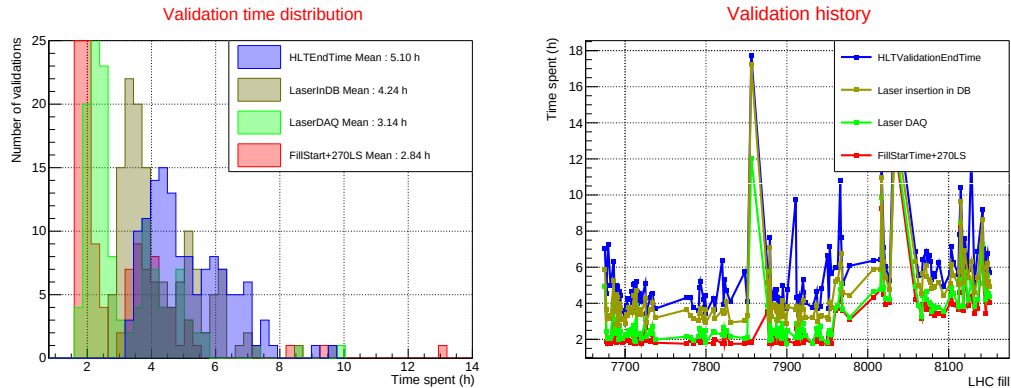


FIGURE 3.15: The left plot shows the time taken for each step of the validation process to run relative to the start of each LHC fill for the Run 3 commissioning and data taking period between fill numbers 7650 to 8020. The right plot shows the total number of validation process during this period and the average time of each step. Before fill 7960, the laser data was collected after 2 hours from the start of a fill, later it was modified to be the start of stable beams. Thus an additional latency is observed after fill 7960. The longer latency peaks are usually caused by either ECAL not participating in the run or due to a delay in the Jenkins job.

Each new validated IOV is uploaded to the front end of the detector as soon as there is a beam dump. These new values are used at L1/HLT for the subsequent fill and a new validation process can begin. Note that a fill has to be more than two hours long to have the availability of the laser data and longer than the time required to have a new validation (usually 4-5 hrs) for a successful upload by the end of that fill.

3.8 Conclusion of the study

The effects of more frequent transparency correction on L1 and HLT rates are studied using data from the 2018 Run period by emulating the L1 and HLT rates for two update frequencies, namely twice-per-week (Run 2) and once-per-LHC-fill (Run 3). Trigger rates from different Level-1 and HLT reconstructed objects were tested and rates vs. PU distributions and rate distributions for a one-week data were studied. A visible recovery in the EGamma rates was observed. The effect is consistent with the results from the offline study which shows a more stable ECAL energy scale with the once-per-LHC-fill updates. HLT electron rates also show substantial improvements in rate stability and energy resolution. These results also indicate that the L1.EG and HLT turn-on curves will be more stable over time because the ECAL energy scale is better corrected with once-per-LHC-fill updates. We tested the validation procedure in the early 2022 commissioning period, achieving an acceptable latency of the entire validation chain. Once-per-LHC-fill updates were routinely performed in 2022 with over 100 updates made during the year.

Part III

Search for Light Higgs Bosons from SUSY Decay Cascades

Chapter 4

Theoretical and experimental motivation

4.1 Introduction

A considerable portion of the Minimal Supersymmetric Standard Model (Section 1.2.1), parameter space has been explored and ruled out at the LHC, imposing limitations on the characteristics of potential SUSY particles. These findings motivate the consideration of alternatives beyond the MSSM, such as the NMSSM. In this thesis, we will describe a specific scenario of the NMSSM where many of the MSSM constraints do not apply [57].

Our analysis focuses on a particular benchmark described in Fig 4.1. The lightest particle of our benchmark is a singlino-like Lightest Supersymmetric Particle (LSP), which has only small couplings to other SUSY particles. For this reason, direct decay from the squark to the LSP particle is suppressed, leading to extended decay cascades. The absence of the direct decay allows signal events to exhibit properties significantly divergent from the anticipated characteristics in the MSSM. For instance, in this scenario, there may be very little missing transverse momentum p_T^{miss} . Moreover, within the framework of the NMSSM, the theory predicts the existence of seven Higgs bosons where, one of them, the CP-even singlet-like Higgs boson, is anticipated from theory to have possibly a mass lighter than that of the Standard Model Higgs, rendering it a compelling candidate for potential discovery at the LHC.

The analysis discussed in this chapter is searching for the light CP-even singlet-like Higgs boson. This study describes a negligible p_T^{miss} scenario that makes boosted Higgs bosons the primary signature of the decay. This analysis has developed from a previously published analysis [14], which utilized CMS Run 2 data, corresponding to an integrated luminosity of 138 fb^{-1} . The analysis concentrates on an all-hadronic final state scenario, where pairs of light Higgs bosons each decay into pairs of b-quarks. How-

ever, for the selected benchmark model, no evidence of an excess beyond the Standard Model predictions was observed in the published analysis. In particular, H_1 bosons with masses in the range 40–120 GeV arising from the decays of squarks or gluinos with a mass of 1200 to 2500 GeV are excluded at 95% confidence level in the published analysis. In this study, we introduce an extra scenario for the benchmark model beyond the published one, for which the exclusion limits will be different. Both the previously published benchmark scenario and the new one will be discussed in Section 5.1.

Data-taking for Run 3 has been ongoing since 2022, with an integrated luminosity of $\sim 300 \text{ fb}^{-1}$ anticipated by 2025. The larger data sample motivates an update to this search for light Higgs bosons via NMSSM cascade decays at the CMS experiment. Since the Run 2 analysis, there have been considerable developments in the reconstruction of boosted massive bosons, such as the Higgs, using machine learning techniques. The corresponding improvements to the analysis sensitivity could exceed those expected from the increased statistical power of the larger data sample providing a compelling reason for a new analysis.

In 2020, members of the CMS collaboration released an innovative tool called ParticleNet [58]. This highly effective tool probes multiple properties of jets by employing advanced neural network architecture. It fulfils various functions, including mass reconstruction and the implementation of the b-tagging algorithm. Particularly noteworthy is its proficiency in identifying the properties of fat jets, also known as AK8 jets, which are large-radius jets, discussed in section 2.3.2. These jets are instrumental in studying the decay products of highly boosted particles like the Higgs boson present in this search. ParticleNet double-b-tagger effectively identifies subjects likely originating from the decay of bottom quarks, a crucial aspect of the decay chain of this analysis. These functionalities will be elaborated upon in Section. 6.2.

The analysis encounters multiple systematic uncertainties that need to be studied to assess the impact on the analysis's result. In the published analysis [14], the predominant component of the total uncertainty was coming from statistical uncertainties in the Signal and Control regions. The most significant contributors to the systematic uncertainties were identified as follows: double-b-tag efficiency, jet mass scale, jet mass resolution, jet energy scale and jet energy resolution. These main sources are not expected to change in the future analysis, as studied in [59].

This study, at this stage, exclusively uses data from the year 2018 of Run 2, corresponding to an integrated luminosity of $59.7 \pm 1.5 \text{ fb}^{-1}$.

Table 4.1 shows the main differences between the published analysis and the ongoing one. The fat jet algorithms used in the published analysis were a boosted decision tree (BDT) based double-b-tagger and the Softdrop [60] algorithm. In this new analysis, we use ParticleNet double-b-tagger and ParticleNet mass algorithms [58]. Each of these algorithms will be described in future sections.

	Published Analysis	Ongoing Analysis
Double b-tagger	BDT double-b-tagger	ParticleNet double-b-tagger
Mass reconstruction	Softdrop	ParticleNet mass
Main backgrounds	QCD	QCD & $t\bar{t}$
Secondary backgrounds	$t\bar{t}$, W+Jets, Z+Jets	W+Jets, Z+Jets*
Mass range	$m_{H_1}=30-125$ GeV $m_{SUSY}=800-2800$ GeV	$m_{H_1}=30-125$ GeV $m_{SUSY}=800-3200$ GeV

TABLE 4.1: Comparison between published analysis and ongoing analysis.

Secondary backgrounds are estimated from MC samples.

*Plan to consider also rare backgrounds such as $t\bar{t}H$, $t\bar{t}Z$, ZZ .

4.2 The Signal Model

The decay cascade of the benchmark signal model presented in this search is shown in Fig. 4.1. Since gluinos ($\tilde{g}$) in this model decay almost always to a squark ($\tilde{q}$) and a quark, the decay arms in Fig. 4.1 can arise either from either squarks or gluinos. Assuming a slight mass difference between the gluino and the squark leads to comparable decay processes and comparable final-state particles for both squark and gluino production modes. In the gluino production mode, the additional quark from the gluino decay carries away only a small fraction of the total momentum. Each squark subsequently decays into a quark and a bino-like Next-to-Lightest-Supersymmetric-Particle (NLSP) ($\tilde{\chi}_2^0$). The $\tilde{\chi}_2^0$ subsequently decay into $\tilde{\chi}_1^0$ which is the Lightest Supersymmetric Particle (LSP) and a light Higgs boson. The latter decays into a boosted $b\bar{b}$ pair, resulting in an all-hadronic event.

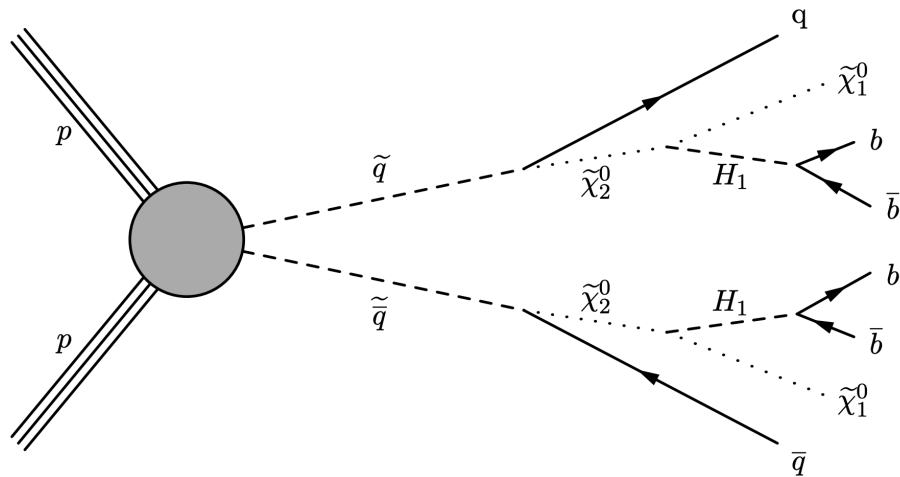


FIGURE 4.1: Feynman diagram of the signal model following squark pair production.

The decay arm can arise either from $\tilde{q}$ or from $\tilde{g} \rightarrow \tilde{q} + q$ states. $\tilde{\chi}_2^0$ is the Next-to-Lightest Supersymmetric Particle which decays to a singlino-like Lightest Supersymmetric Particle, $\tilde{\chi}_1^0$, and the light CP-even neutral Higgs boson H_1 .

In the NMSSM, a stable LSP, characterized by singlino-like features ($\tilde{\chi}_1^0$), does not carry any charge and for this reason does not couple directly to SM particles. Moreover, it has a small coupling with other SUSY particles because its mixing with the other neutralinos is suppressed [61]. Squarks and gluinos decay via the NLSP $\tilde{\chi}_2^0$ into a $\tilde{\chi}_1^0$ and, in general, a Higgs or Z boson. This study focuses on scenarios where the final-state particle is a Higgs boson called H_1 with SM-like couplings but either lighter than or equals the H_{SM} ($m_{H_1} \leq 125$ GeV). The direct production cross-section of the NMSSM light Higgs boson is notably suppressed compared to the H_{SM} , offering an explanation for its potential elusiveness in previous detection attempts.

The mass of the light-flavour squarks is denoted by m_{SUSY} . This is a degenerate mass such that $m_{SUSY} = m_{\tilde{q}} = m_{\tilde{u}}, m_{\tilde{d}}, m_{\tilde{s}}, m_{\tilde{c}}$. The masses of sbottom and stop are presumed to be greater than those obtainable through direct production.

The cross section for producing squark or gluino pairs ($\tilde{q}\tilde{q}^*, \tilde{q}\tilde{q}, \tilde{q}\tilde{g}, \tilde{g}\tilde{g}$) with mass m_{SUSY} in 13 TeV proton-proton collisions is detailed in Tab. 4.2. The cross sections are calculated using NNLL-fast 1.1¹ [20] which provides an estimate of the theoretical uncertainty due to scale variation, the parton distribution functions, and the value of the strong coupling α_s .

TABLE 4.2: Inclusive cross sections for $pp \rightarrow \tilde{q}\tilde{q}^*, \tilde{q}\tilde{q}, \tilde{q}\tilde{g}, \tilde{g}\tilde{g}$ as a function of m_{SUSY} with uncertainties. The cross sections are calculated using NNLL-fast 1.1 [20] for squark masses and gluino masses set to 1% larger than m_{SUSY} .

m_{SUSY} [GeV]	$\sigma(pp \rightarrow \tilde{q}\tilde{q}^*, \tilde{q}\tilde{q}, \tilde{q}\tilde{g}, \tilde{g}\tilde{g})$ [fb]	Uncertainty
800	7720	6.6%
1000	1952	7%
1200	582.3	8%
1400	194.1	9%
1600	69.71	9%
1800	26.29	10%
2000	10.29	11%
2200	4.114	13%
2400	1.668	14%
2600	0.6792	16%
2800	0.2765	18%
3000	0.1117	21%
3200	0.04512	24%

Apart from m_{SUSY} and m_{H_1} , the signal model also relies on other unknown parameters such as $m_{\tilde{\chi}_2^0}$ and $m_{\tilde{\chi}_1^0}$. These two extra degrees of freedom are described through the

¹The NNLL-fast 1.1 tool upper range evaluation extends only to 3000 GeV. To obtain the value at 3200 GeV, an extrapolation was performed.

TABLE 4.3: Branching fraction of $H_1 \rightarrow b\bar{b}$ for different values of m_{H_1} . The branching fractions were computed utilizing the HDECAY software [21] [22].

m_{H_1} [GeV]	Branching Fraction ($H_1 \rightarrow b\bar{b}$)
30	0.868
35	0.867
40	0.865
50	0.858
60	0.850
70	0.840
80	0.829
90	0.816
100	0.795
110	0.749
125	0.581

quantities:

$$R_M \equiv \frac{m_{H_1}}{m_{\tilde{\chi}_2^0}}, \quad \Delta_M \equiv m_{\tilde{\chi}_2^0} - m_{H_1} - m_{\tilde{\chi}_1^0} \quad (4.1)$$

Choosing $R_M = 0.99$ and $\Delta_M = 0.1$ GeV yields a light $\tilde{\chi}_1^0$ with minimal momentum in comparison to H_1 , which is what is needed to achieve a p_T^{miss} -suppressed signal. Values of $R_M > 0.9$ still exhibit a similar signal, but R_M values lower than 0.9 introduce non-negligible p_T^{miss} and in that case it would be more suitable to employ standard SUSY searches that incorporate the missing transverse momentum [62, 57]. With these chosen parameters, the squark and gluino decay cascade concludes with a soft $\tilde{\chi}_1^0$ and a boosted H_1 (for $m_{\text{SUSY}} > 1200$ GeV and $m_{\text{SUSY}} \gg m_{H_1}$). Typically, H_1 decays into a boosted $b\bar{b}$ pair, as indicated by the branching fractions in Tab. 4.3, assuming the $H_1 \rightarrow b\bar{b}$ branching fractions to be SM-like [61]. The branching fractions were calculated using HDECAY [21, 22], assuming a SM Higgs but adjusting the mass. The branching ratios of the SM Higgs are shown in Fig. 4.2.

For an event to be considered a signal, both H_1 bosons must decay into $b\bar{b}$ pairs. Under the specified values of R_M and Δ_M , the branching fractions of the decays $\tilde{q} \rightarrow q + \tilde{\chi}_2^0$ and $\tilde{\chi}_2^0 \rightarrow H_1 + \tilde{\chi}_1^0$ are assumed to be unity. In the benchmark model, this is true except where $m_{\tilde{\chi}_2^0} > m_Z + m_{\tilde{\chi}_1^0}$. This means that the $\tilde{\chi}_2^0 \rightarrow Z + \tilde{\chi}_1^0$ decay is allowed if the $\tilde{\chi}_2^0$ has a higgsino component [63]. However, the $\tilde{\chi}_2^0$ is expected to be mainly bino-like for values of its mass relevant to this study [61].

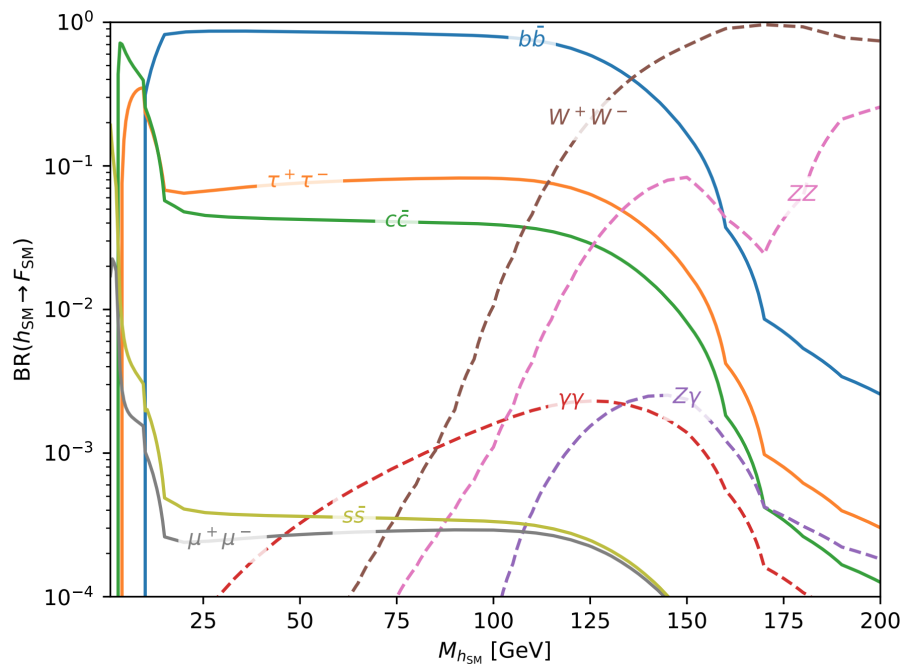


FIGURE 4.2: SM-like Higgs boson branching ratios for various SM particle final states as a function of the Higgs mass. Figure from Ref. [13]

Chapter 5

Monte Carlo production

To simulate and predict the outcomes of the data analysis, Monte Carlo (MC) simulation samples play a crucial role in developing and testing the signal event selection, as well as estimating background contributions. For this study, official CMS MC background samples were utilized, generated using MadGraph5_aMC@NLO [64] and Pythia [65] and using Geant4 [66] for simulating particle interactions with the CMS detector.

The MadGraph and Pythia tools will be extensively described in future sections. These generators, along with their specific settings, have been validated to provide the best description of CMS data to date. The same simulation procedure is employed in the generation of the signal MC sample, described in Section 5.5.

QCD multijet events and $t\bar{t}$ events represent the main background of this analysis. Both QCD and $t\bar{t}$ samples were produced in bins of H_T ¹ at the generator level, in order to ensure the large- H_T region relevant to the analysis is adequately populated. Additionally, Z+jets and W+jets processes are secondary background processes. They can generate background events through decays into quark-antiquark pairs, particularly $Z \rightarrow b\bar{b}$, which if sufficiently boosted is likely to produce a double-b-tagged fat jet.

5.1 Signal MC

For this analysis, we generated a range of signal MC samples at NLO using NNPDF3.1 NNLO, with the values of m_{H_1} and m_{SUSY} varied within the mass ranges of 30-125 GeV and 800-3200 GeV, respectively. These samples were validated against the old MC samples produced for the published analysis.

Two different categories of signal MC samples were produced:

¹As already mentioned in Section 2.3.2, H_T is defined as the scalar sum of the jets p_T . In this case, it is the p_T sum of the generated partons.

- A first category that comprises all possible pair-production modes involving squarks and gluinos: $\tilde{q}\tilde{q}^*$, $\tilde{q}\tilde{q}$, $\tilde{q}\tilde{g}$, $\tilde{g}\tilde{g}$. Here, the gluino mass is set to be 1% higher than the squark mass. This is chosen to ensure that the final state is essentially the same for squark and gluino production because most of the gluino momentum is transferred to the daughter squark in the $\tilde{g} \rightarrow \tilde{q}q$ decay and the only additional particle is a soft quark.
- A second category in which the decay chain arises exclusively from squarks. This channel is restricted to on-shell $\tilde{q}\tilde{q}^*$ production and the gluino is deliberately decoupled with a mass fixed at 10^6 GeV.

The main reason for these two scenarios is that these processes have different cross-sections. The first case represents a more optimistic scenario for the production cross section for the final state of interest, while the second case is more pessimistic (although it still assumes 8 degenerate light-flavoured squarks).

The published analysis exclusively examined the first scenario. The cross-sections for this case were previously discussed in Tab. 4.2. In contrast, the cross-sections for the second scenario are presented in Tab. 5.1. To distinguish between the two cases, we will denote the squark mass as m_{SUSY} in the first scenario and m_{Squark} in the second scenario.

TABLE 5.1: Cross sections for $pp \rightarrow \tilde{q}\tilde{q}^*$ as a function of m_{Squark} with uncertainties. The cross sections are calculated using NNLL-fast 1.1 [20].

m_{Squark} [GeV]	$\sigma(pp \rightarrow \tilde{q}\tilde{q}^*)[\text{fb}]$	Uncertainty
800	328	9%
1000	68.8	11%
1200	17.1	13%
1400	4.78	15%
1600	1.44	18%
1800	0.457	20%
2000	0.150	24%
2200	0.051	28%
2400	0.0174	33%
2600	0.00599	39%
2800	0.00207	46%
3000	0.00072	55%
3200	0.00025	65%

5.2 MadGraph and Pythia

MadGraph and Pythia are two essential tools for the MC production.

MadGraph5 is a Monte Carlo event generator for collider studies designed to automatically generate matrix elements for various particle physics processes. It can calculate cross-sections and produce parton-level events across a broad spectrum of processes. Outputs from MadGraph contain the events generated following the matrix element for that particular process. MadGraph describes the probabilities for the outgoing partons to have certain properties and it can generate randomly one event at a time, according to the probability set out from the matrix elements. These outputs are often forwarded to complementary programs, such as Pythia, to simulate subsequent stages like hadronization and particle decays.

Pythia is a general-purpose event generator for simulating high-energy particle collisions. It is widely used to model the development of parton showers, the hadronization process and the subsequent decay of particles produced in high-energy collisions. It provides a detailed description of the final state of a collision, which is essential for comparing theoretical predictions with experimental data obtained from particle accelerators like the Large Hadron Collider.

They often work in synergy in this way: MadGraph is used to generate parton-level events, and then these events are passed to Pythia for simulating the parton shower, decay and hadronization processes. This combined approach allows for a more comprehensive simulation of the entire collision process, from the initial interaction of partons to the final-state particles observed in experiments.

5.3 The merging problem

In the Monte Carlo production chain, the Initial State Radiation (ISR) refers to the emission of high-energy particles (typically gluons) from the incoming partons involved in a collision process.

In the Monte Carlo simulations, the initial state radiation is simulated to model the effects of these additional emissions accurately. Calculating the ISR is crucial because it can have a significant impact on the kinematics and observables of the final-state particles in the process.

Madgraph can handle the production of a custom number of additional partons. However, it is less accurate when additional produced particles are soft or collinear, so Pythia treatment of ISR is preferable in these cases.

To ensure good modelling of the hardest additional jet in our MC production, events are generated in Madgraph with zero or one additional jet at the Matrix Element level

and then decayed in Pythia (parton shower). However, sometimes Pythia can produce an additional jet as well and thus there is the potential for double-counting. To avoid this, we employed the MLM merging technique [67], which mitigates this issue.

This technique uses two variables called XQCUT and QCUT. For our study, we want to generate only up-to-one additional ISR-associated jet in Madgraph. In this context, **XQCUT** is the minimum p_T of the additional parton generated by Madgraph. **QCUT**, on the other hand, defines the minimum p_T required when counting jets clustered from the partons immediately after the Pythia parton shower, but before hadronization.

We configure Madgraph to produce two subsamples of events, with zero and one additional parton. For the Madgraph sample with zero additional partons, only events where the additional jets produced by Pythia have a p_T lower than QCUT threshold are retained. This is the region (where soft particles are produced) where Pythia is preferred over Madgraph.

In the second subsample, events are generated by Madgraph with one additional parton with $p_T > \text{XQCUT}$. Jets are clustered from the partons after the Pythia parton showering has been performed. These jets are matched with the original partons (produced by Madgraph). The jet matched with the additional parton is required to have a p_T higher than QCUT, which avoids double counting of events from the sample with zero additional partons. Furthermore, any events with non-matched jets (which represent the additional jets produced by Pythia) harder than the matched one are vetoed, because the jet generated by MadGraph is the hardest one by definition.

It is crucial to ensure that **QCUT is greater than or equal to XQCUT**, as failing to meet this condition would leave a region of phase space of the additional jet p_T uncovered. Beyond this requirement, QCUT effectively controls the transition point in p_T between jets produced by Pythia and MadGraph, but there is no absolute theoretical specification for the choice of the value of QCUT. It must therefore be determined empirically.

Differential Jet Rate (DJR) plots serve as a test that the resulting predictions are physical. DJR plots show the transition between Pythia and MadGraph jet p_T regimes, by illustrating the logarithm of the transverse momentum of the additional jet. A smooth curve in DJR plots implies a physical distribution and thus a good choice of QCUT. The values of XQCUT and QCUT thus have to be carefully selected to ensure stability in the DJR plot.

An example of this kind of plot is presented in Fig. 5.1. As this is a physical quantity, the distribution in the plot should be continuous and any discontinuity in the distribution indicates an issue with the merging processes.

In the pursuit of understanding the merging process within MC samples, our investigation focused on both signal scenarios: the one with gluino decoupled $m_{\text{gluino}} = 10^6$ GeV and the one with $m_{\text{gluino}} = 1.01 \times m_{\text{SUSY}}$. Values of XQCUT from 30–60 GeV were tested, and values around 60 GeV gave QCUT results that depended less on signal

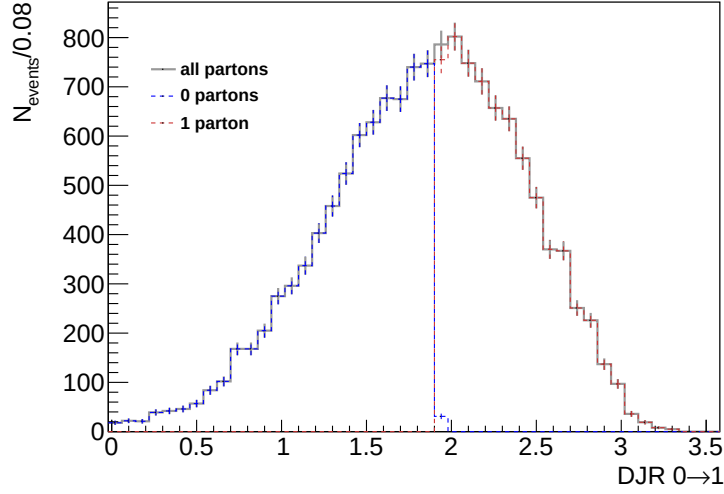


FIGURE 5.1: An example of a DJR plot, generated by adjusting XQCUT and QCUT values in MC simulations. The plot illustrates the logarithm of the transverse momentum of the first additional jet in our dataset. Since this is a physical quantity, the distribution on the plot should be continuous.

model and mass. While somewhat larger than the XQCUT value of 30 GeV often used for CMS SUSY production, the indicated value of 60 GeV is still small compared to the jet p_T selection requirement applied in this analysis that will be discussed in Section 6.4.

Fixing XQCUT to 60 GeV imposes a $QCUT \geq 60$ GeV. Appendix A shows some examples of all the DJR plots that were produced by adjusting QCUT values in order to obtain a smooth shape in the plot. As an illustration, in Appendix A, we report values for low, medium, and high m_{SUSY} mass points, although all mass points were examined.

The output of this study is that we have identified a range of QCUT that produce a reasonably smooth curve in the DJR plots, see Tab. 5.2. To ensure good performance of the MLM merging algorithm for both scenarios across the entire mass range of this analysis, we opted for a value of XQCUT equal to 60 GeV and QCUT equal to 90 GeV.

m_{SUSY}	QCUT 1° scenario	QCUT 2° scenario
800	90-120	80-120
1600	70-120	80-120
2400	70-120	70-120
2800	70-90	70-120
3200	90-100	90-100

TABLE 5.2: Comparison of QCUT values resulting in smooth DJR curves for two distinct scenarios. The first scenario involves $\tilde{q}\tilde{q}$, $\tilde{q}\tilde{q}$, $\tilde{q}\tilde{g}$, $\tilde{g}\tilde{g}$ processes, while the second involves the $\tilde{q}\tilde{q}$ production.

5.4 Comparison between the two MC scenarios

Due to the significant time and resource investment required for the MC production, the initial step focused on generating samples exclusively for the first scenario across the entire mass range of m_{SUSY} from 800 to 3200 GeV and m_{H_1} from 30 to 125 GeV. Conversely, only a limited number of samples were generated for the second scenario, specifically those with $m_{H_1} = 60$ GeV and $m_{SUSY} = 2800$ GeV, to allow comparisons across the full range of masses. This approach aimed to verify the similarity between the two sets of samples and test the possibility of a simple reweighting method to translate between the two predictions, potentially saving considerable time.

Given the 1% assumed mass separation between the squarks and gluinos in the first scenario, the differences between the two scenarios are expected in the kinematic distributions of the fat jets, not in their other properties such as mass and double-b-tag distribution. We test this using the H_T , the mass and the double-b-tag distributions in three distinct m_{SUSY} mass points: low mass (1200 GeV), middle mass (2400 GeV), and high mass (3200 GeV), spanning the range examined in our study. The mass of the light Higgs particle is set to $m_{H_1} = 60$ GeV.

The comparisons are illustrated in Figs. 5.2-5.4. In the plot legends, " m_{SUSY} " represents the first scenario where the decay chain possibly originates from the gluino and " m_{Squark} " denotes the gluino decoupled scenario.

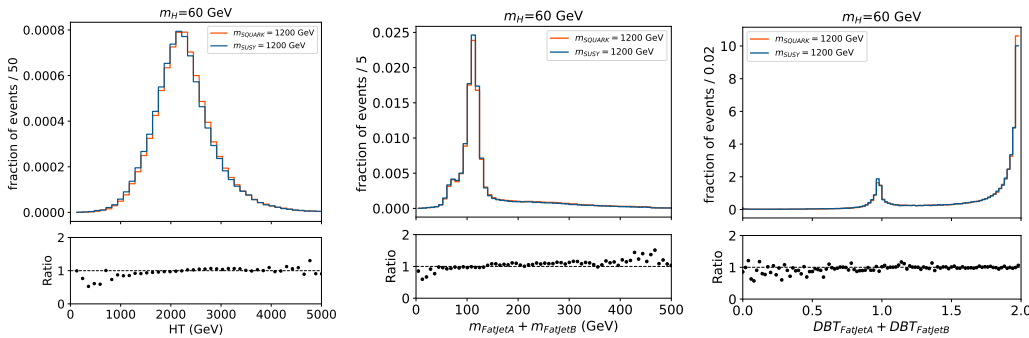


FIGURE 5.2: These three plots in a row represent from left to right: the H_T distribution, the mass distribution of the two fat jets and the double-b-tag distribution of the two fat jets for m_{Squark} or m_{SUSY} equal 1200 GeV, $m_{H_1} = 60$ GeV.

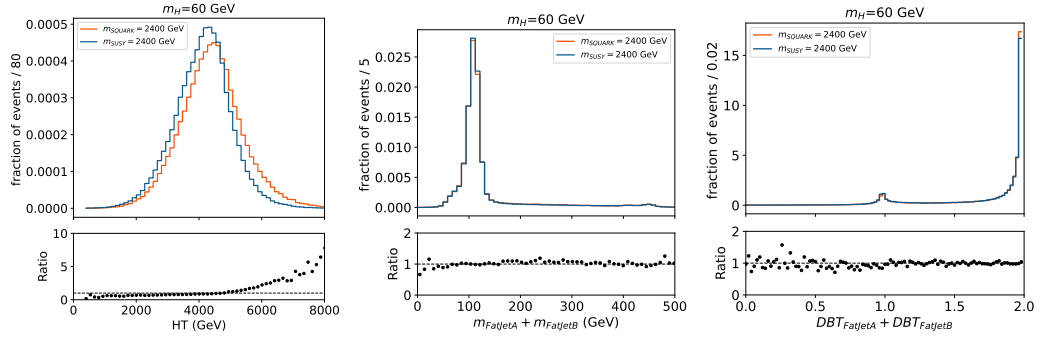


FIGURE 5.3: These three plots in a row represent from left to right: the H_T distribution, the mass distribution of the two fat jets and the double-b-tag distribution of the two fat jets for m_{Squark} or m_{SUSY} equal 2400 GeV, $m_{H_1} = 60$ GeV.

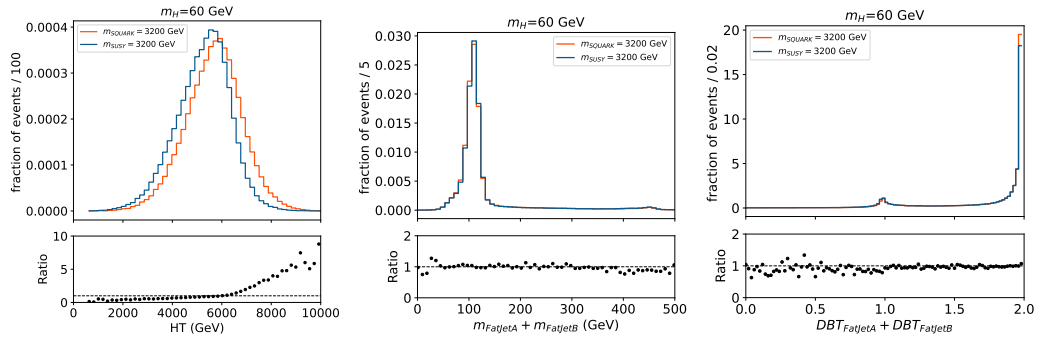


FIGURE 5.4: These three plots in a row represent from left to right: the H_T distribution, the mass distribution of the two fat jets and the double-b-tag distribution of the two fat jets for m_{Squark} or m_{SUSY} equal 3200 GeV, $m_{H_1} = 60$ GeV.

Analysing these plots, it is visible that the two cases differ in H_T distribution, especially at high m_{Squark}/m_{SUSY} mass. On the other hand, they share a similar distribution for the mass reconstruction and the double-b-tag score of the two fat jets, as expected.

We show in Fig. 5.5 the distributions at $m_{Squark}/m_{SUSY} = 3200$ GeV after reweighting the H_T distribution of the m_{SUSY} to the one of m_{Squark} . We do not suggest this H_T -based reweighting as a real method to use, but we show it as a proxy to test that a kinematic reweighting does not appreciably affect the fat jet mass and double-b-tag distributions.

A more precise reweighting could be performed on the squark/gluon momenta at the generator level, which should capture most of the difference between the two models, but this is left for future studies.

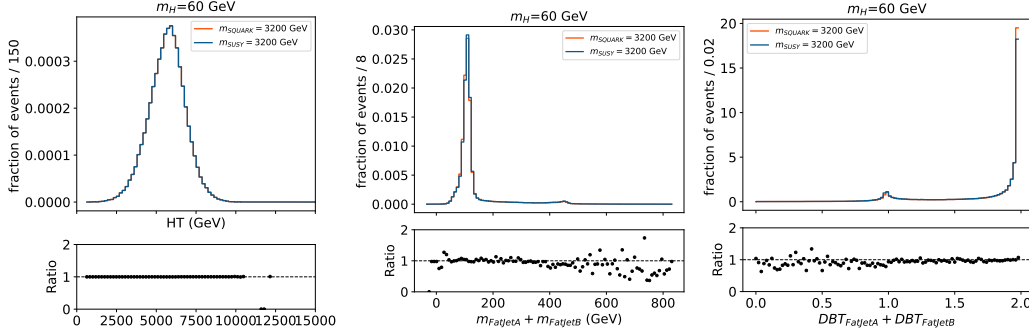


FIGURE 5.5: These three plots in a row represent from left to right: the H_T distribution, the mass distribution of the two fat jets and the double-b-tag distribution of the two fat jets for m_{Squark} or m_{SUSY} equal 3200 GeV after the reweight of m_{SUSY} to m_{Squark} following the H_T distribution ratio.

5.5 MC production chain

A private MC production was implemented in order to generate signal events for the benchmark model². At this stage of the study, the Monte Carlo simulations are tailored to replicate the conditions of Run 2. The Monte Carlo production process was designed to ensure systematicity, to facilitate adaptation of the methodology for Run 3 production.

The production process is designed to follow the CMS workflow composed of several steps which correspond to various stages of simulating and reconstructing particle interactions. The most important points of the production are as follows:

1. **LHEGEN (Les Houches Event files and Generator):** represents the Madgraph and Pythia steps that are already discussed in section 5.2
2. **SIM (Simulation):** provides full simulation of the particle interactions in the CMS detector. The simulated events from LHEGEN are processed through a Geant4-based simulation of the CMS detector, taking into account the materials, geometry, and response characteristics. This step generates a realistic representation of how particles would behave and interact within the detector.
3. **DIGIPremix (Digitization and Premixing):** it simulates electronic signals from particles interacting with the detector and the mixing of pileup events. The simulated data undergo digitization to replicate the response of the detector's electronics to particle signals. Additionally, the step involves premixing to account for pileup events, which are interactions from other proton-proton collisions occurring simultaneously.

²This necessity arose due to the absence of ParticleNet in the samples generated for the paper, as well as to facilitate testing of the new alternative signal model.

4. **RECO (Reconstruction):** it takes care of the reconstruction of particle trajectories and identification of physics objects (same reconstruction as in data explained in 2.3.1). The reconstructed data include information on tracks, calorimeter energy deposits, and other detector features. Algorithms are applied to identify and reconstruct particles like electrons, muons, photons, and jets.
5. **Mini (MiniAOD) and Nano (NanoAOD):** the purpose of these steps is the creation of a smaller and more condensed dataset for analysis. The data are processed to create a MiniAOD (Mini Analysis Object Data) format, which is a more compact representation compared to the full RECO data. Then they are further reduced to create NanoAOD which is a specialized format that provides even more condensed data, typically tailored to specific analysis requirements. It retains only the most relevant information needed for particular physics analyses. The same steps are applied to real data.

These steps collectively represent the sequential stages of simulating, reconstructing, and processing data in the CMS experiment's MC production chain. The final NanoAOD file contains all the ParticleNet information that we need for our analysis, such ParticleNet-tagger and ParticleNet-mass. These variables will be better discussed in future sections. Following these steps, we generated 150 000 events for 143 mass points for the first benchmark scenario and 23 for the second scenario.

Chapter 6

Event selection

6.1 The analysis framework

The analysis framework used in this study was based on NanoAODTools [68]. Coffea [69] was used for making histograms. *NanoAODTools* is a common Python-based CMS analysis tool. The framework incorporates tools to evaluate commonly used systematic uncertainties and provides corrections based on scale factors with calibration values recommended by CMS, which are applicable for Run 2 (2016-2018) data. These include corrections and systematic uncertainties for AK4 and AK8 jet energies and mass reconstructions, as well as pileup and scale factors for b-taggers to account for any discrepancies in b-tagging performance between data and simulated events. However, since ParticleNet is a relatively new double-b-tag discriminator, manual implementation of scale factors and uncertainties specific to ParticleNet is required.

Coffea stands for Columnar Object Framework For Effective Analysis and is an experimental package designed to streamline high-energy collider physics (HEP) analysis by providing tools with a familiar syntax. Exploiting uproot and awkward-array, it enables array-based manipulation of HEP event data efficiently.

In order to select the signal with high efficiency while strongly reducing the background sources, this analysis exploits the properties of AK8 jets. In particular, the selection relies on the output of double-b-tag-discriminant and jet mass reconstruction algorithms, which are going to be discussed in the following sections.

6.2 Double-b-Tagging

As already mentioned in Section 2.3.2, a double-b-tagger is an algorithm designed to identify AK8 jets originating from a $b\bar{b}$ pair. Typically, this tagger provides a discriminator value that indicates the likelihood of the jet containing a $b\bar{b}$ pair. Recent advance-

ments in machine learning have significantly enhanced the performance of jet flavour tagging algorithms.

Some taggers have been trained to classify jets from specific decays. BtagHbb¹ in particular is a double-b-tag discriminator developed in 2015 by the CMS collaboration [70] and specifically designed for boosted $H \rightarrow b\bar{b}$ decays [10]. This is the double-b-tagger employed in the Run 2 published analysis [14]. Other taggers are more versatile, and capable of classifying jets from a $b\bar{b}$ pair stemming from the decay of an unspecified particle. An illustration of a versatile classifier is provided by the ParticleNet double-b-taggers used in this new search.

6.2.1 The btagHbb double-b-tagger

The btagHbb double-b-tagger [70] has been in use for several years and demonstrated superior performance compared to earlier double-b-tagging techniques at the time of its creation. The btagHbb double-b-tagger utilizes a machine learning algorithm known as a Boosted Decision Tree (BDT) for the purpose of data classification.

Input features from a set of training data are used in order to construct a decision tree. Subsequently, this decision tree is employed to classify new data points based on similarities to the patterns observed in the training set. In essence, the decision tree algorithm learns from the provided training data, identifies underlying relationships, and applies this knowledge to categorize unseen data. The decision-making begins with a single "node" that divides the input data based on one or more characteristics. The data then progresses to one of the nodes in the next layer, determined by the values of the data. At each node, another input feature is selected, and this process continues until no additional nodes are available. The final node in this decision tree is also called a "leaf" and contains the output of the classifier. Boosting is a method of combining many weak classifiers (in this case the decision trees described above) into a strong classifier which we call BDT. Many decision trees are iteratively created and for each, a weight is assigned depending on its accuracy. The BDT output will be the sum of all weighted decision trees. At each iteration, each data sample is given a weight based on its misclassification. The more often a data sample is misclassified, the more important it becomes.

For btagHbb, the leaves contain values ranging from -1 to 1. A value of -1 suggests that the jet likely did not originate from a $b\bar{b}$ pair, while a value of 1 indicates a high likelihood that it did. A value of 0 means the algorithm couldn't determine if the jet contained a $b\bar{b}$ pair or not. The btagHbb BDT was trained using simulated $H \rightarrow b\bar{b}$ decays as the signal and QCD multijet events as the background. Inputs considered by the BDT included the impact parameters of tracks in the jet and the correlation between the directions of the two subjets from the b hadrons. To avoid any correlation between

¹Also known as "pfBoostedDoubleSecondaryVertexAK8BJetTags".

the double-b-tag discriminator value and the reconstructed mass of the AK8 jet, any dependency on the mass of the particle that decayed to form the $b\bar{b}$ pair is intentionally not taken into account.

6.2.2 ParticleNet-tagger

ParticleNet is an algorithm based on a neural network, which is inspired by human neuron connections. It can be seen as a network of artificial brain cells, each with adjustable weights. These brain cells, called nodes learn from simulated data. The first layer of nodes can handle multiple input features at the same time, processing information from different aspects simultaneously. Moving in depth through layers, each node is connected to multiple nodes in the previous layer. The last layer contains the output of this process. During training, the weights of these nodes change, helping the algorithm to improve its performance over time.

ParticleNet is a powerful flavour tagger designed in the CMS experiment. Its performances are high in jet tagging. It treats each jet as a collection of particles, and uses a graph neural network (GNN) approach to infer relationships between particles. The input information is taken from the properties of the jet's constituent Particle Flow (PF) candidates, such as their momentum, charge, and impact parameter, as well as properties of the primary vertex. ParticleNet has been trained to distinguish different types of jets from the background, especially those coming from QCD multijets. One of its roles is to act as a double-b-tagger, helping identify jets resulting from the decay of b-quark pairs.

ParticleNet-tagger outputs a value between 0 and 1, representing the probability that the AK8 jet originates from a b-quark pair. The closer the value is to 1, the more confident ParticleNet is about the jet originating from a b-quark pair.

6.3 Jet Mass Reconstruction

After the reconstruction and identification of a jet, another crucial step is determining the mass of the particle that gave rise to the jet. This involves employing mass reconstruction algorithms designed for AK8 jets. The following section outlines specific techniques for accomplishing this task, the one used in the Run 2 published analysis is the Softdrop algorithm and the one that will be used in this new search is the ParticleNet algorithm.

6.3.1 The Softdrop algorithm

The Softdrop algorithm [60] is employed for jet reconstruction, specifically for estimating the mass of jets with substructure, such as AK8 jets originating from $H \rightarrow b\bar{b}$ decays. This algorithm is designed to eliminate soft wide-angle radiation, including pileup effects, that might obscure the inherent properties of the jet.

The process initiates with a jet of radius R_0 , reconstructed through the anti-kt [53] algorithm. Subsequently, the jet undergoes reclustering using the Cambridge/Aachen algorithm [52] (described in Section 2.3.2) to restore the structure of the parton shower, which is compromised during the application of the anti-kt algorithm. Following this, the soft drop algorithm selectively eliminates soft particles from the jet, unless the following condition is satisfied:

$$\frac{\min(p_{T1}, p_{T2})}{p_{T1} + p_{T2}} > z_{\text{cut}} \left(\frac{\Delta R_{12}}{R_0} \right)^\beta$$

where p_{T1} and p_{T2} are the transverse momenta of the subjets, ΔR_{12} is the angular separation between the subjets, R_0 is the jet radius (i.e., $R_0 = 0.8$ for an AK8 jet).

The Softdrop algorithm relies on two parameters: z_{cut} removes soft radiation, while β removes wide-angle radiation. The parameters z_{cut} and β control the extent of the grooming with $z_{\text{cut}}=0$ or $\beta \rightarrow \infty$ returning an ungroomed jet. In the context of CMS, specific values are typically employed, with $z_{\text{cut}} = 0.1$ and $\beta = 0$. Consequently, the Softdrop criteria depend solely on momentum.

The Cambridge-Aachen algorithm clusters the radiation that is physically closest first. If you reverse the last step of this algorithm, you get the two subjets needed for the Softdrop criteria. The process involves breaking apart the jet, checking the Softdrop criteria, and removing the softest subjet. This cycle is repeated until both subjets satisfy the criteria. Subsequently, the mass is computed from the remaining constituents of the jet.

If neither subjet satisfies the Softdrop condition, the mass is set to 0 GeV. This scenario may arise, for instance, when the PF objects from one of the b quarks are not within the AK8 jet radius or if a light jet is erroneously identified as one of the two selected fat jets. Another circumstance leading to a jet mass close to 0 GeV is when the algorithm removes an excessive number of PF objects.

6.3.2 The ParticleNet mass

The ParticleNet mass [71] is determined utilizing the ParticleNet mass regression algorithm, which employs a neural network similar to that utilized in the ParticleNet jet flavour taggers. The input data for the neural network is the same as for the double- b -tagger, and includes the reconstructed properties of each PF particle within the AK8

jet. The output is the reconstructed mass of the jet. The training process involves simulated events with a flat mass distribution for particles decaying to boosted $b\bar{b}$ pairs. Therefore, the algorithm is not dependent on the mass of the original particle. Additionally, the training includes QCD multijet events to allow mass reconstruction also for jets arising from quarks or gluons. In this case, the true mass is taken from the Soft-drop mass evaluated on jets clustered at the generator level (after hadronisation). Since the neural network for the double-b-tagger is trained using the same simulated events (with a flat mass distribution), the double-b-tag score is not expected to be correlated with the mass.

6.4 Kinematic Event Selection

To distinguish signal events from the background, we examined various event characteristics. Since most of the energy from squarks and gluinos manifests as jets in the chosen $H \rightarrow b\bar{b}$ final state, a high value of H_T serves as an indicator of signal events. Consequently, events with high H_T were selectively chosen at the trigger level. Only events passing a specific HLT H_T trigger were considered for signal event selection. The H_T distribution for various signal and background samples is visualized in Fig. 6.1.

In the published analysis, an HLT trigger line was used with a minimum H_T threshold of 1050 GeV, since this was the lowest available for Run 2. The same cut is applied to the new analysis. This threshold effectively eliminated over 99% of QCD events and 92% of $t\bar{t}$ events. While m_{SUSY} and m_{H_1} had a minor impact on the percentage of rejected signal events, less than 1% of the events were eliminated on average.

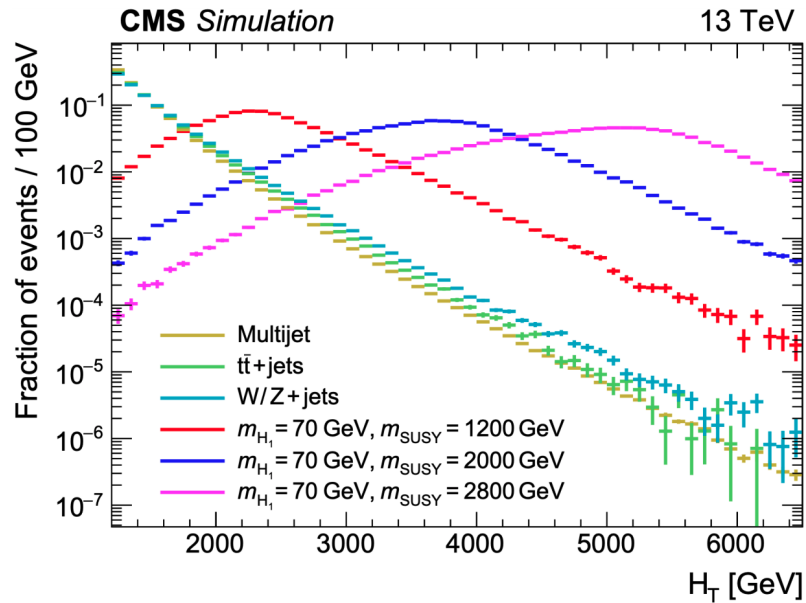


FIGURE 6.1: The H_T distribution analyzed across diverse background and signal events. Every event fulfils the H_T trigger selection criteria.

The H_T is calculated as the scalar sum of all AK4 jets with $p_T > 40 \text{ GeV}$ and $|\eta| < 2.4$. The $|\eta| < 2.4$ cut is employed to avoid efficiency losses in the ECAL endcaps due to radiation damage in the 2018 Run 2 dataset. Events are categorized into H_T bins to enhance sensitivity for larger values of m_{SUSY} , given that signal event H_T increases with m_{SUSY} , while all background sources have a distribution falling in H_T as clearly visible in Fig. 6.1. In the published analysis three H_T bins are studied, corresponding to $(1500 < H_T < 2500) \text{ GeV}$, $(2500 < H_T < 3500) \text{ GeV}$ and $H_T > 3500 \text{ GeV}$.

The two AK8 jets with the highest double-b-tag discriminator value are selected to represent the $b\bar{b}$ pairs resulting from the decay of the two boosted light Higgs bosons. The two selected fat jets are arbitrarily assigned the labels “FatJetA” and “FatJetB” to ensure symmetrical distributions. The quarks from the squark decay are represented by smaller AK4 jets.

Once the fat jets are selected based on the double-b-tag score, the signal events are chosen based on the following criteria already present in the published analysis:

- $H_T > 1500 \text{ GeV}$.
- $p_T > 300 \text{ GeV}$ and $|\eta| < 2.4$, for the two AK8 jets corresponding to the two boosted $b\bar{b}$ pairs.
- At least one AK4 jet with $p_T > 300 \text{ GeV}$ and $|\eta| < 2.4$, corresponding to a quark from one of the squark decays. To guarantee that there is no overlap with the same PF candidates as the AK8 jets, the AK4 jet must be separated from both of them by a distance $\Delta R > 1.4$.

The $p_T > 300 \text{ GeV}$ threshold for jets was determined based on the p_T distribution of signal AK8 and AK4 jets [62]. After the trigger selection, the remaining QCD and $t\bar{t}$ events in the background Monte Carlo samples are reduced by 95% and 90%, respectively, based on the kinematic event selection. Additionally, approximately 93% of the remaining $W + \text{jets}$ and $Z + \text{jets}$ events are filtered out. About 17% of the events in the signal sample for $m_{H_1} = 80 \text{ GeV}$ and $m_{SUSY} = 2800 \text{ GeV}$ did not pass the selection criteria.

To compare between btagHbb/Softdrop and ParticleNet, $H_T > 1500 \text{ GeV}$ is used.

6.5 Performance Comparison and Cut Criteria

The purpose of the following sections is to show how the analysis benefits from switching from the old mass reconstruction and b-tagging algorithm to the new ones from ParticleNet. The selection thresholds applied to the new discriminating variables are chosen so that the event count for MC signal events is similar to that obtained with the old selection. This allows us to compare the background efficiency of the two selections.

6.5.1 Double-b-tag Selection

The double-b-tag discriminators for each of the two AK8 jets together form a 2D space. The two jets resulting from $b\bar{b}$ decays cluster in the upper right corner of the parameter space. Signal events are selected by a Tag Region (TR), defined as the area where the sum of the double-b-tag discriminator values exceeds a certain threshold. In the previous Run 2 analysis, a cut of 1.3 was applied for the sum of the btagHbb discriminators, as illustrated in Fig. 6.2.

Two additional regions were defined in this double-b-tag distribution to estimate the background from data: the Control Region (CR) and the Validation Region (VR). The CR, characterized by a low double-b-tag discriminator value for both selected AK8 jets, is expected to predominantly contain QCD background events and is used for background estimations. The VR is employed to validate this background estimation.

From this point, all the quoted selection efficiencies assume the application of the kinematic event selection. When the signal sample with $m_{H_1} = 80$ GeV and $m_{SUSY} = 2800$ GeV is considered, 63% of the events are retained after cutting on the btagHbb discriminator. The same signal efficiency is found when a cut on the sum of ParticleNet double-b-tag discriminator of 1.95, as illustrated in Fig. 6.2. This cut defines the new Tag Region for the ParticleNet tagger. The CR and VR are chosen to cover the same fraction of area as for btagHbb, with further optimization left for future work. The definitions of CR and VR are solely used for data-driven background estimation.

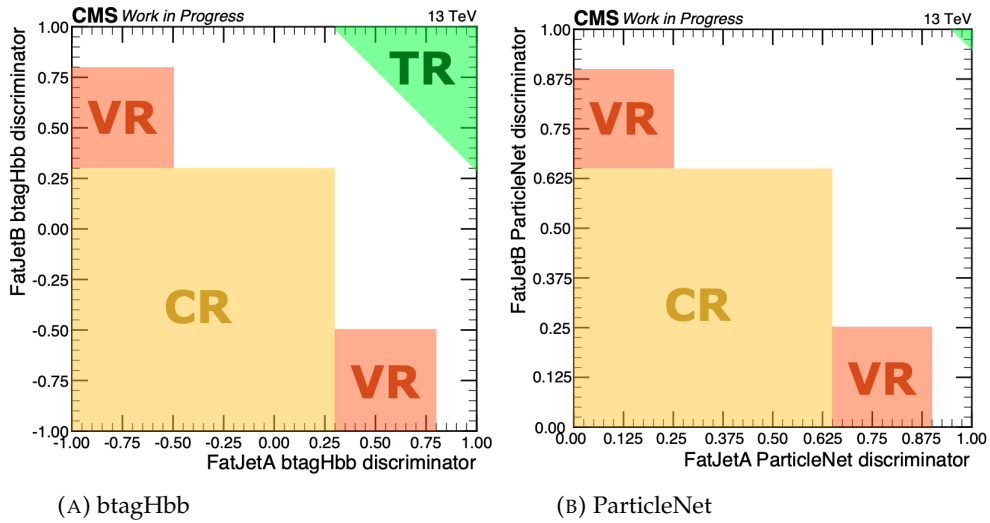


FIGURE 6.2: The two-dimensional space of btagHbb (left) and ParticleNet tagger (right) is divided into the signal region, also called Tag Region (TR) in green, the Control Region (CR) in yellow, and the Validation Regions (VR) in red. These latter are designated to estimate background events.

Figure 6.3 shows the 2D distributions of double-b-tag discriminator values for a signal sample with $m_{H_1} = 80$ GeV and $m_{SUSY} = 2800$ GeV, comparing btagHbb (left) and ParticleNet (right).

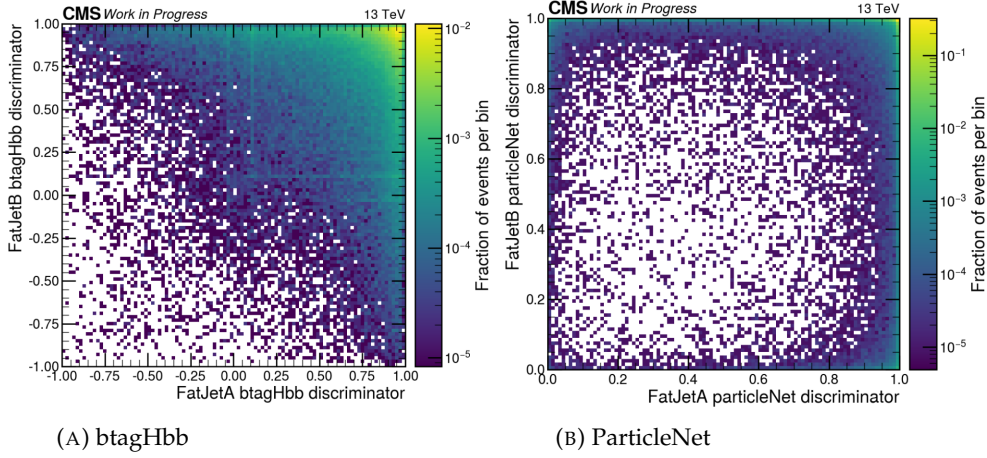


FIGURE 6.3: The double-b-tag discriminator values distribution for the two chosen AK8 jets for the signal sample with $m_{H_1} = 80 \text{ GeV}$ and $m_{SUSY} = 2800 \text{ GeV}$, comparing btagHbb algorithm (A) and ParticleNet (B).

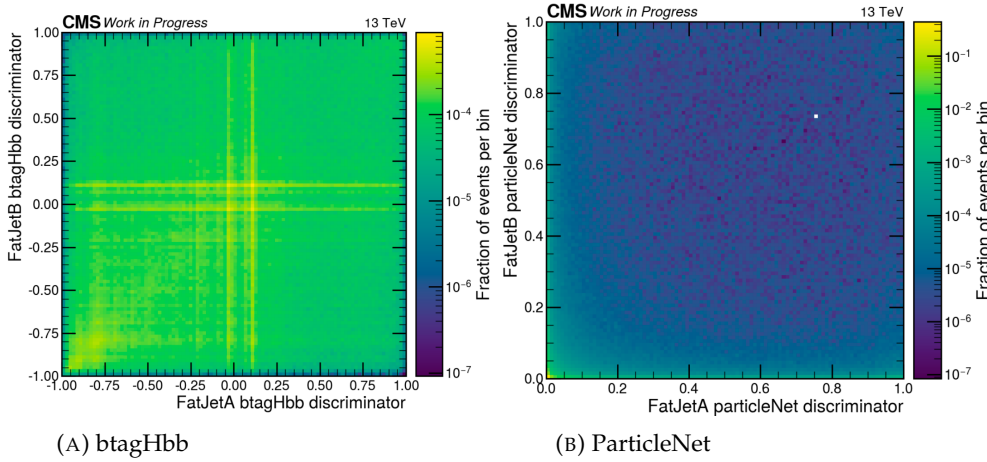


FIGURE 6.4: The double-b-tag discriminator values distribution for the two chosen AK8 jets for QCD background events, comparing btagHbb algorithm (A) and ParticleNet (B). The empty bin in the ParticleNet distribution is due to the limited number of events in the Monte Carlo sample.

Figures 6.4-6.5 report the 2D double-b-tag distributions for QCD and $t\bar{t}$ background samples². ParticleNet tagger exhibits much better signal-to-background discrimination, as signal events are concentrated in the top right corner, while background events in the bottom left.

The TR of the btagHbb (ParticleNet) tagger contains 2.4% (0.01%), 9.6% (0.5%), 2.4% (0.01%) and 4.4% (0.1%) of the remaining events for QCD, $t\bar{t}$, W+jets and Z+jets, respectively. The $t\bar{t}$ background shows a higher fraction of events with elevated double-b-tag values due to b quarks originating from top quark decays.

²In Figures 6.4-6.5, the horizontal and vertical lines near zero for btagHbb are where the double-b-tag tagger struggles to determine the presence or absence of a $b\bar{b}$ pair within the fat jet. This phenomenon is studied in [62], and is thought to be caused by the presence of jets with very small Softdrop mass. Importantly, as these lines are significantly distant from the TR, disagreements between data and simulation in this region cannot impact the analysis.

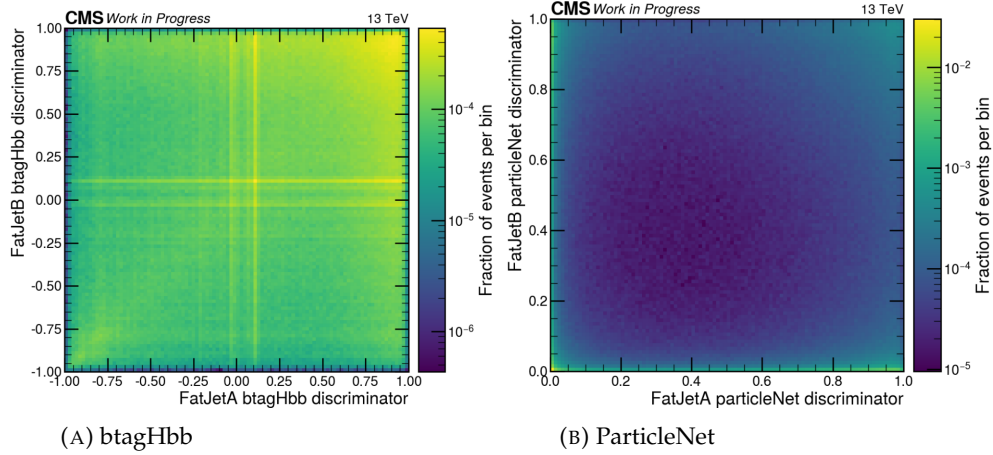


FIGURE 6.5: The double-b-tag discriminator values distribution for the two chosen AK8 jets for $t\bar{t}$ background events, comparing btagHbb algorithm (A) and ParticleNet (B).

6.5.2 Mass Event Selection

Two distinct methods for evaluating the jet masses are compared: the Softdrop technique [60], used in the Run 2 published analysis [14], and ParticleNet mass [71], explained in Section 6.3.

The comparison of the mass distribution for signal events is shown in Fig. 6.6. Notably, the ParticleNet mass exhibits superior performance compared to the Softdrop mass. A greater number of events closely around the true value of m_{H_1} indicate a reduced susceptibility to the issue of zero-mass jets. Furthermore, ParticleNet shows improved mass resolution.

The fat jets shown in Fig. 6.6 are the two AK8 jets with the highest double-b-tag values. Consequently, the observation that ParticleNet more accurately selects the correct fat jets than btagHbb could potentially influence both the spread and the peak around zero in the mass distribution. This highlights the complementarity of correctly identifying the relevant fat jets and their mass reconstruction.

Both selected AK8 jets are expected to originate from a light Higgs boson. Their masses should closely align with m_{H_1} . In the 2D space defined by the masses of the two AK8 jets, the signal should cluster at a point on the diagonal, depending on the value of m_{H_1} . The published analysis uses multiple signal regions on the diagonal to account for a spectrum of different masses. Together with these signal regions, sideband regions are also defined adjacent to the signal mass regions. These sideband regions will be used in data-driven QCD background estimation. Figure 6.7 illustrates an example of the employed signal and sideband regions. The same set of signal and sideband mass regions is applied to ParticleNet, in order to compare the algorithms.

Figure 6.8 presents the mass distributions of the selected AK8 jets for $m_{H_1} = 80$ GeV

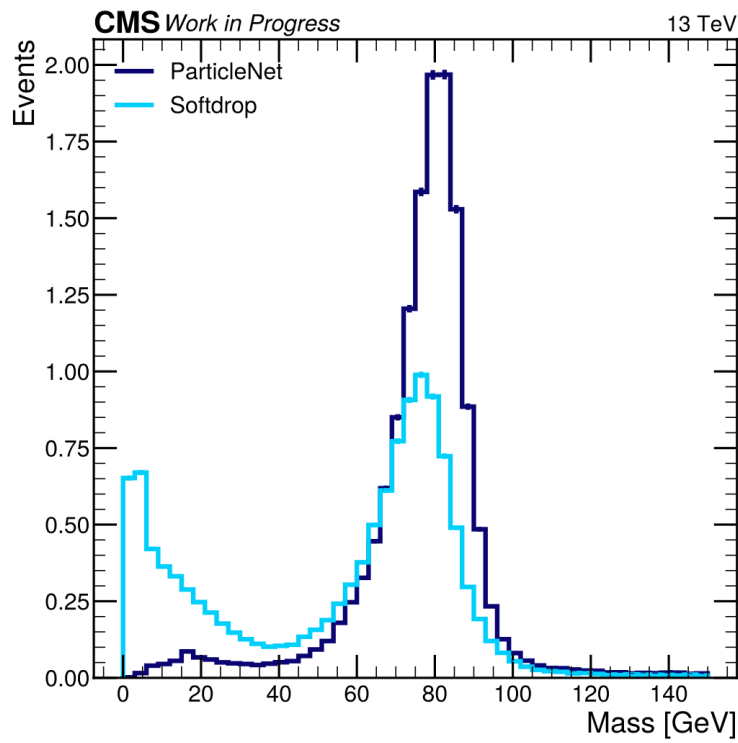


FIGURE 6.6: Comparison of the mass reconstructed by the Softdrop and ParticleNet algorithm in a signal sample with $m_{H_1} = 80$ GeV, $m_{SUSY} = 2800$ GeV.

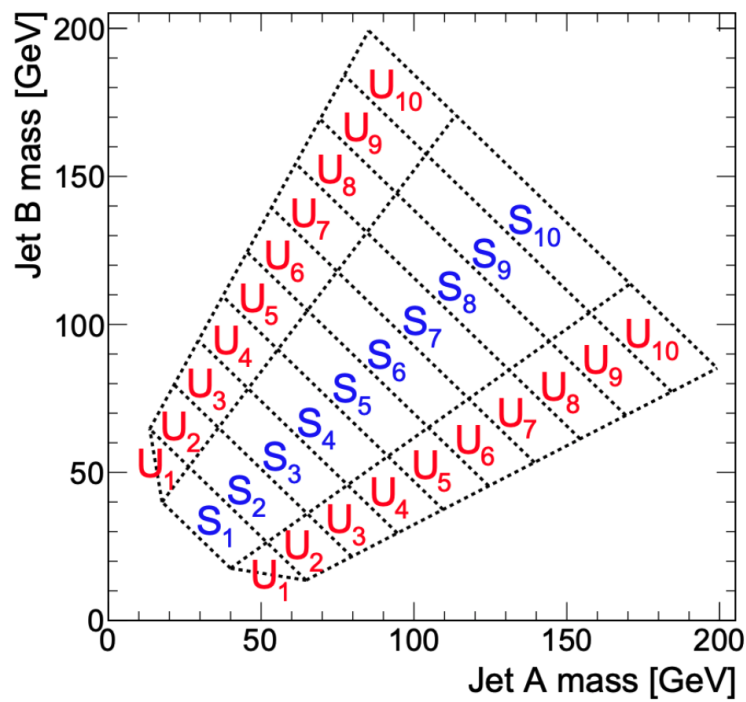


FIGURE 6.7: Scheme of the mass regions in the 2D jet mass plane. Signal mass regions are labelled as S_i , while sideband mass regions are identified as U_i .

before and after the double-b-tag selection and for $m_{H_1} = 40$ GeV only after the double-b-tag selection. The application of the double-b-tag selection eliminates events where only one of the two jets genuinely originates from $H \rightarrow b\bar{b}$, an effect particularly evident in the ParticleNet mass distributions. The figures also show that AK8 mass increases in accordance with m_{H_1} . Instead, m_{SUSY} has a much lower influence on the mass distributions [59].

As already discussed, the subjet grooming in the Softdrop algorithm is too aggressive and in many cases leaves jets with a single subjet (and thus close to zero mass). Instead, the ParticleNet distribution shows a larger fraction of events around the expected signal mass regions. In the signal sample with $m_{H_1} = 80$ GeV and $m_{SUSY} = 2800$ GeV, 49% of the events are selected within any of the signal mass regions for the Softdrop mass (after selection using the btagHbb tagger), while 98% are selected when ParticleNet mass reconstruction algorithm are used (after selection using the ParticleNet tagger).

Figure 6.9 shows the comparison between Softdrop and ParticleNet for the QCD background of the AK8 mass distributions after the double-b-tagger selection. The ParticleNet tagger eliminates a significant portion of the QCD background, with a final number of events in the signal region much smaller than using the previous tagger and mass algorithm. The final number of QCD events that passed all selections is a factor of 100 smaller for ParticleNet with respect to old algorithms used in the published analysis.

Figure 6.10 reports the comparison between Softdrop and ParticleNet for the $t\bar{t}$ background after the double-b-tag selection. A factor of 10 fewer $t\bar{t}$ events passed all selections for ParticleNet compared to the combination of btagHbb and Softdrop algorithms (the improvement is not as pronounced as for the QCD background).

Figure 6.11 shows the comparison between btagHbb and Softdrop algorithms and ParticleNet tagger and mass algorithms for W+jets and Z+jets backgrounds.

6.5.3 Yields in the Signal Mass Region

The yields within each signal mass region, after applying the event selections and estimated using Monte Carlo background and signal samples, are shown in Fig. 6.12. ParticleNet algorithm results in a considerably lower count of background events compared to the btagHbb and Softdrop algorithms, achieved while increasing the signal yields. In this figure, the lower H_T cut is moved from 1500 GeV (the threshold used in the published analysis) to 2000 GeV. This is because the H_T -binned $t\bar{t}$ background samples do not cover the region below 2000 GeV. However, this region plays a role in the analysis sensitivity only to signal with $m_{SUSY} < 1200$ GeV, which is already excluded in the published analysis. From this point on, future plots and studies will use this selection.

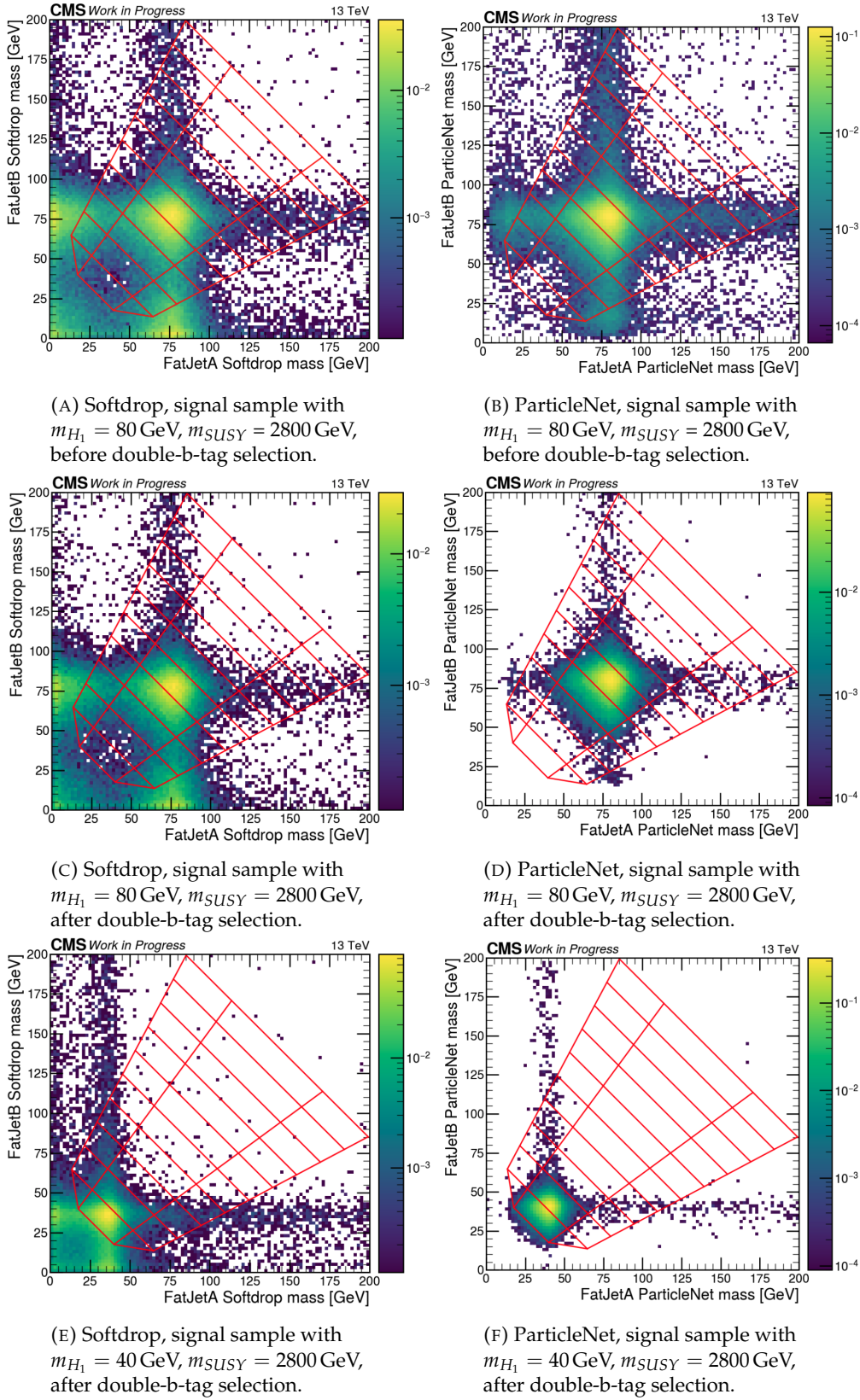
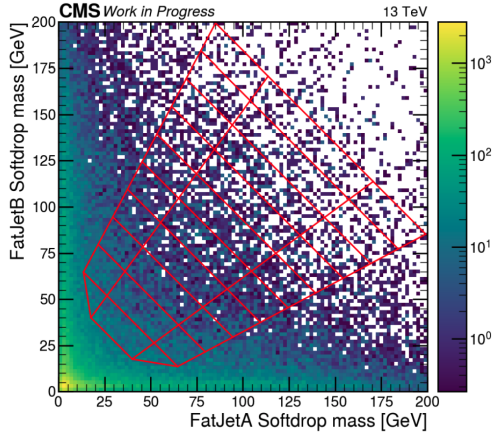
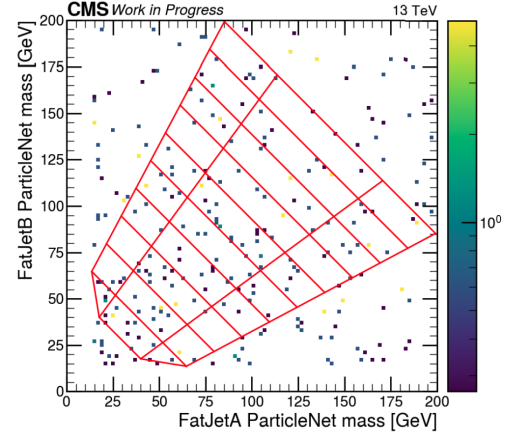


FIGURE 6.8: Distribution of the reconstructed jet mass for the two selected AK8 jets across various m_{H_1} masses and before and after the double-b-tagger selection, for the Softdrop (left) and the ParticleNet algorithm (right).

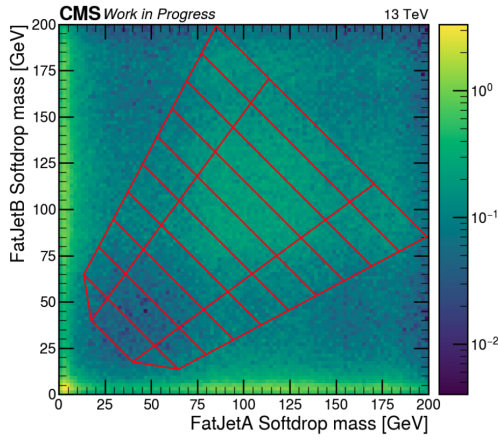


(A) Softdrop

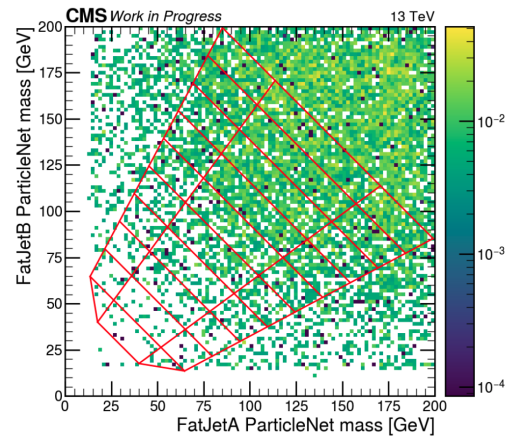


(B) ParticleNet

FIGURE 6.9: The distribution of the reconstructed jet mass for the two selected AK8 jets in the QCD background for the combination of btagHbb and Softdrop (left) and ParticleNet algorithms (right).



(A) Softdrop



(B) ParticleNet

FIGURE 6.10: The distribution of the reconstructed jet mass for the two selected AK8 jets in the $t\bar{t}$ background for the combination of btagHbb and Softdrop (left) and ParticleNet algorithms (right).

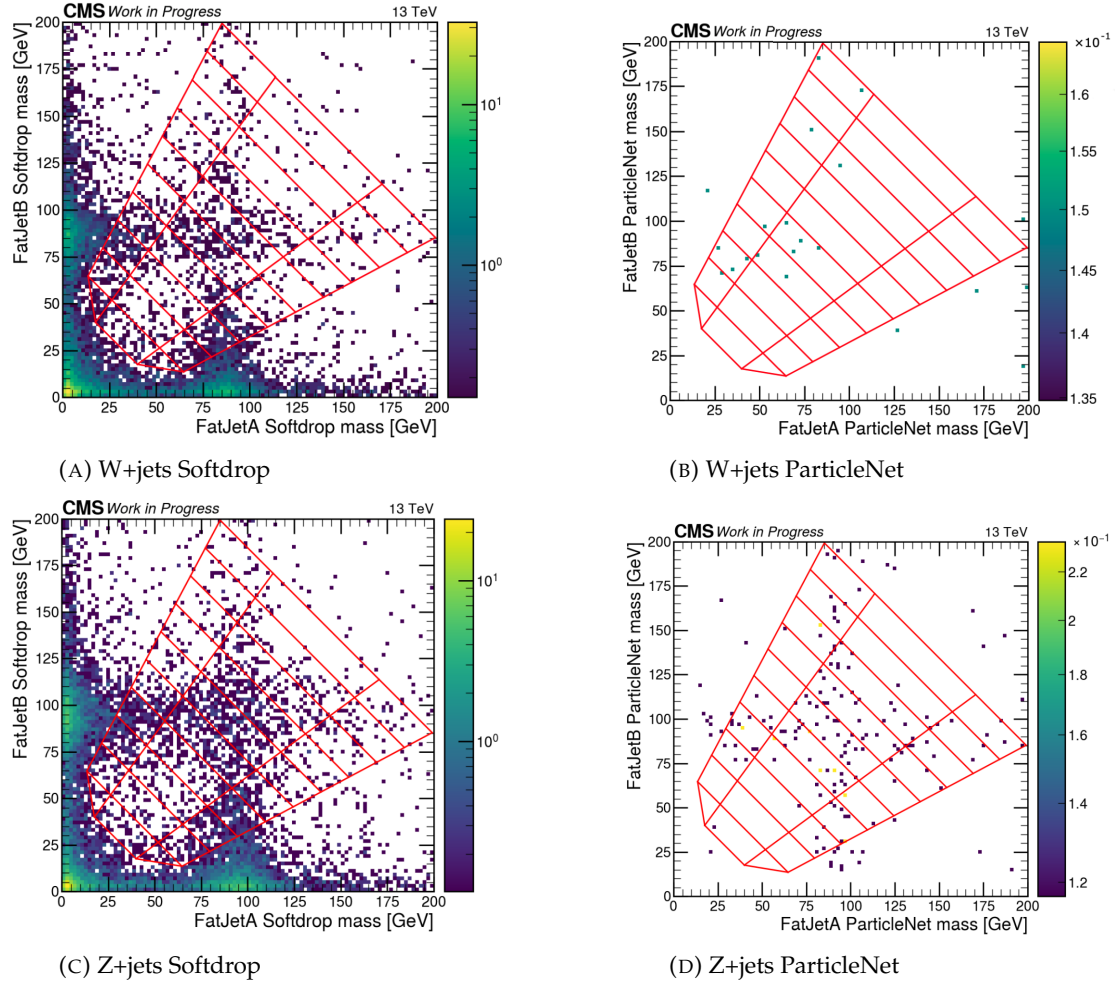


FIGURE 6.11: The distribution of the reconstructed jet mass for the two selected AK8 jets in the W+jets (A and B) and Z+jets (C and D) for the combination of btagHbb and Softdrop (left) and ParticleNet algorithms (right).

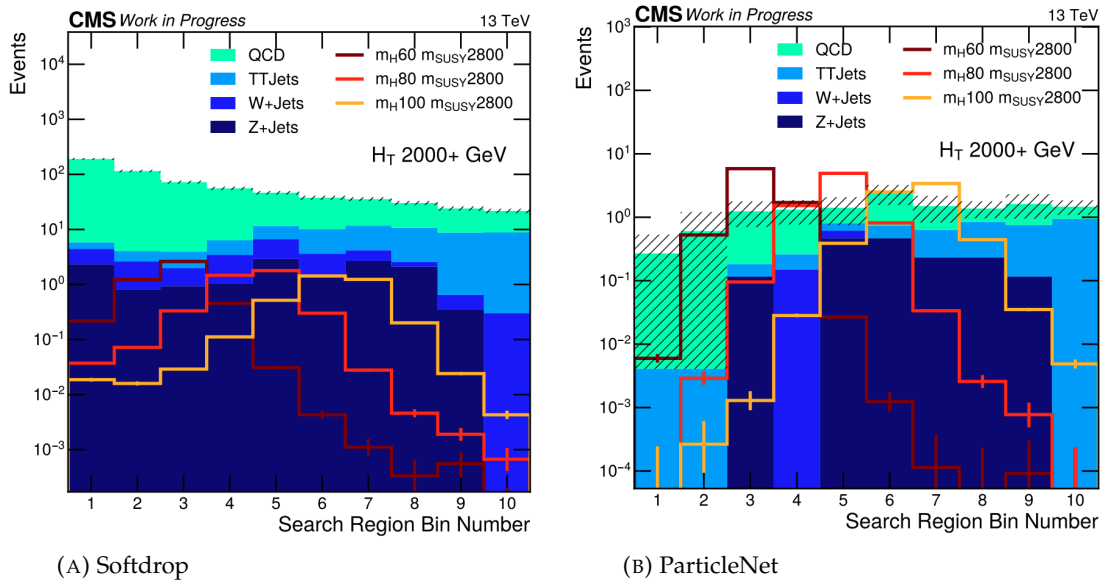


FIGURE 6.12: Yields in the signal mass regions after applying event selections with btagHbb and Softdrop algorithms (left) and the ParticleNet algorithms (right). The plotted signal and background yields are based on estimates derived from simulated events. The hatched bands in the figures denote the statistical uncertainty associated with the total background, while systematic uncertainties are not shown.

Chapter 7

Background estimations

For backgrounds with limited contributions, such as the W+jets and Z+jets backgrounds in this investigation, the background count in each signal bin is determined solely from Monte Carlo samples. While predictions from MC simulations carry substantial relative systematic uncertainties, their impact on the overall background uncertainty remains minimal. In the case of more significant backgrounds, obtaining a precise estimate becomes crucial, and a data-driven approach is often preferred.

The QCD background study will be described in Section 7.1. In the published analysis, the contribution of the $t\bar{t}$ background was small compared to the QCD one, so it was estimated from simulation. However, given the improved capability of ParticleNet in rejecting QCD background, this assumption is no longer valid. Section 7.2 explores the background estimation for the $t\bar{t}$ background.

After comparing the expected significance of the analysis with different double-b-tag values, we found out that the double-b-tag sum cut of 1.90 in the TR is preferable over 1.95. The value at 1.90 is also preferred because it permits higher statistics than the cut at 1.95. This is needed for both the QCD and the $t\bar{t}$ background studies. This value will be applied to the following studies and plots.

7.1 QCD Background Estimation

The estimation of the QCD background is accomplished using data and the sideband regions U_i as defined in Fig. 6.7, employing an "ABCD" method [72]. Each signal region S_i is intentionally designed to be sufficiently wide to ensure that only a negligible number of signal events is in the U_i sideband region. The dimensions of the sideband regions are chosen to match the area of their corresponding signal regions. Consequently, if the event density is approximately uniform across both the sideband U_i and the signal region S_i , the number of QCD background events in the two regions should be roughly equal.

The event yields in the signal and sideband regions are denoted as $\hat{S}_i$ and $\hat{U}_i$, respectively. To ensure that only background events are counted and to have more statistics, F_i is measured in the control region (CR). The ratio F_i between the event yields of S_i and U_i in the CR is defined as: $F_i := \hat{S}_i^{CR} / \hat{U}_i^{CR}$.

The QCD background in the Tag Region can then be estimated as: $\hat{S}_i^{TR} = F_i \hat{U}_i^{TR}$. Here, $\hat{U}_i^{TR}$ represents the observed yield in the TR after subtracting the MC simulated background estimations (currently from $t\bar{t}$, Z+jets, and W+jets).

A requirement of the ABCD method is that the two variables used to define signal and control region (which in this case are the mass reconstruction and the double-b-tag selection) should be independent. In the published analysis, this condition was verified for btagHbb and the Softdrop mass and it is still valid for ParticleNet [59].

While the ParticleNet double-b-tagger is explicitly designed to be uncorrelated with the jet mass, achieved through training with a flat distribution of Higgs masses, there may be residual correlations. To address this, the background estimation was checked using a methodology similar to the one employed for the previous analysis. This involved comparing $\hat{S}_i^{TR}$ and $\hat{S}_i^{VR}$ predicted by the ABCD method (exploiting the CR), against the simulated yield in the QCD MC sample (as shown in Fig. 7.1).

Figure 7.1 (left) provides a ParticleNet comparison between the calculated yield $\hat{S}_i^{TR}$ and the observed Tag Region (TR) yield in the simulated QCD sample.

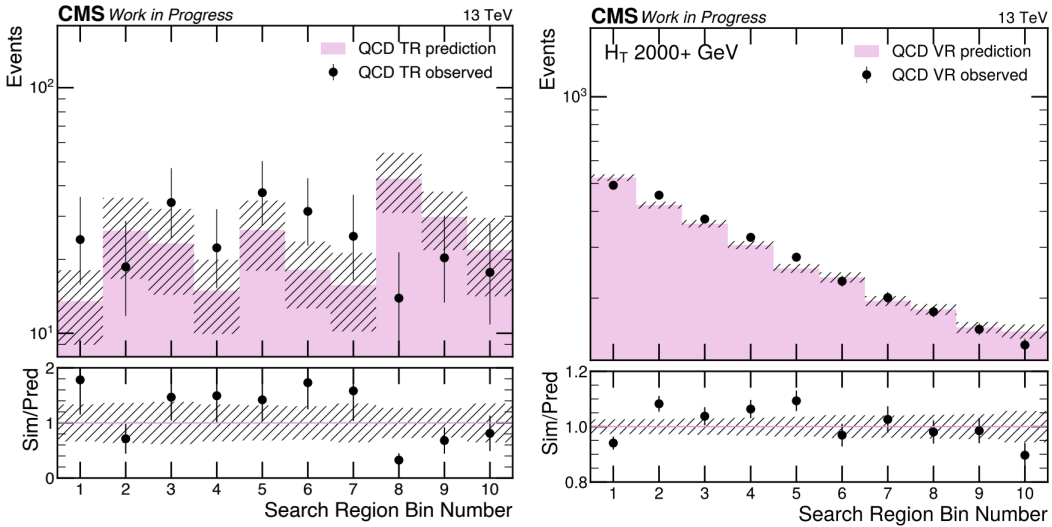


FIGURE 7.1: Predicted and Observed yields in the TR (left) and VR (right) regions for the QCD MC background. The shaded bands and error bars show the statistical uncertainties of predicted and observed values, respectively.

Given the strong performance of the ParticleNet double-b-tag discriminator, the yields in each bin are notably small, leading to substantial statistical errors which limit the sensitivity of this test. This limit is overcome if the comparison is performed in the VR region, shown in Fig. 7.1 (right), where more events are observed.

7.1.1 Data-driven QCD Estimation

A comparison between the simulated and predicted yields in the Validation Region (VR) using real data is presented in Fig. 7.2. The value of $\hat{U}_i^{VR}$ here represents the yield in data after subtracting the contributions from the MC simulated backgrounds (excluding QCD). The $t\bar{t}$ background is here estimated using MC. This is justified by the fact that the proportion of $t\bar{t}$ events in the VR is significantly smaller than in the TR. The estimated yields for the default double-b-tag discriminator cut are presented in Fig. 7.2, indicating effective performance in QCD background estimation. The outcomes are also in line with observations from a previously published analysis. In that analysis, a systematic uncertainty of 15% was assigned in H_T categories < 3500 GeV, and a 30% for $H_T \geq 3500$ GeV [14].

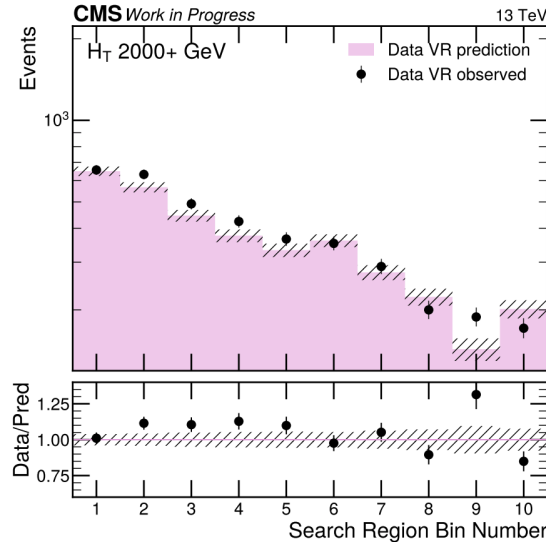


FIGURE 7.2: Predicted and observed yields in the VR region using the actual yields in Run 2 data from 2018. $H_T \geq 2000$ GeV is applied. The shaded bands and error bars show the statistical uncertainties of predicted and observed values, respectively.

7.2 The $t\bar{t}$ Background Estimation

As already discussed, the transition from Softdrop and btagHbb to ParticleNet has significantly altered the background composition in this analysis. In particular, as shown in Fig. 6.12, at a working point corresponding to a factor of 2 increase in signal efficiency, **the QCD background has been diminished by approximately a factor of ~ 100 and the $t\bar{t}$ background has experienced a reduction by a factor of ~ 10** . Given the substantial reduction in the QCD background in comparison to $t\bar{t}$, these two backgrounds have become comparable in magnitude. Consequently, the $t\bar{t}$ background, which was exclusively studied through Monte Carlo simulations in the previous analysis, now requires further study. In this section, we present a detailed study of the $t\bar{t}$ background, with the aim of introducing new cuts to mitigate the background and also to improve the background estimation technique beyond a simple MC prediction.

Following the comparison between Softdrop and ParticleNet algorithms, a revised mass grid was proposed [59]. It is depicted in Fig. 7.3 (left), while the signal sample with $m_{H_1} = 80 \text{ GeV}$, $m_{SUSY} = 2800 \text{ GeV}$ with the new grid is shown in Fig. 7.3 (right). The dimensions of the signal regions, denoted as S_i , have been marginally reduced, due to the enhanced mass resolution achieved with the ParticleNet algorithm. Conversely, the side-band regions, U_i , have been expanded in size relative to their previous dimensions. This adjustment aims to include a greater number of events within these regions, thereby reducing the statistical uncertainties associated with background estimation.

For all the studies in this Section we apply the following selections: $H_T > 2000 \text{ GeV}$, the events need to be inside the mass grid illustrated in Fig. 7.3, and they need to be inside the Tag Region (sum of the double-b-tag discriminants for FatJet A and B > 1.9). We call this the “Loose Signal Region”. This corresponds to the full signal selection, apart from the binning in H_T and mass regions.

In $t\bar{t}$ samples, two top quarks decay approximately 100% of the time to a W boson and a b quark [23]. When the top quarks are highly boosted, they are a source of massive fat jets containing a b quark, which can almost mimic a signal $H \rightarrow b\bar{b}$ jet. While such jets are still unlikely to pass the double-b-tag selection (corresponding to fat jets containing 2 b quarks), the large mass makes them more signal-like than typical b-jets in the QCD background sample. This is why the ParticleNet double-b-tagging reduces the $t\bar{t}$ background less than the QCD background, thus leaving $t\bar{t}$ as an important source of background in this analysis. The potential decays of the two W's originating from the top quarks are as follows:

- Fully hadronic: $W \rightarrow q\bar{q} + W \rightarrow q\bar{q}$
- Semileptonic: $W \rightarrow q\bar{q} + W \rightarrow l\nu$
- Fully leptonic: $W \rightarrow l\nu + W \rightarrow l\nu$

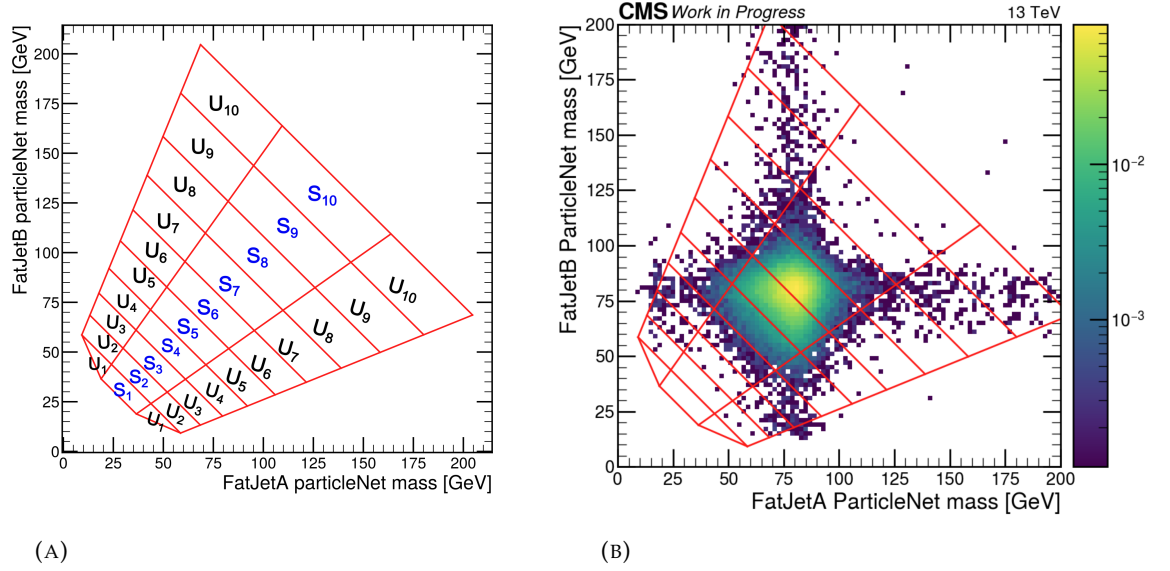


FIGURE 7.3: Scheme of the new mass regions in the 2D jet mass plane (left). ParticleNet, signal sample with $m_{H_1} = 80$ GeV, $m_{SUSY} = 2800$ GeV, after double-b-tag selection with new grid (right).

Table 7.1 reports the branching fractions and uncertainties for different decay modes of the W boson.

Mode	Fraction (Γ_i/Γ)
$e^+\nu$	$(10.71 \pm 0.16)\%$
$\mu^+\nu$	$(10.63 \pm 0.15)\%$
$\tau^+\nu$	$(11.38 \pm 0.21)\%$
<i>Hadrons</i>	$(67.41 \pm 0.27)\%$

TABLE 7.1: Decay modes of W^+ with relative branching fractions and uncertainties. W^- has the charge conjugate of these modes [23].

The fat jets in $t\bar{t}$ events can be broken down into different subclasses based on the W boson decay mode and whether they have sufficient p_T to contain a fully merged top quark decay. This is shown in Fig. 7.4. When the transverse momentum p_T exceeds 430 GeV, it indicates that the top quark products are likely to be sufficiently collimated to be encapsulated within the AK8 jets, as given by the approximate relationship $p_T \sim \frac{2m}{\Delta_R}$. Conversely, $p_T < 430$ GeV implies that the decay products of the b quark and the W boson will not be confined within a single fat jet. As explained in section 6.4, the decision to set a selection threshold of $p_T > 300$ GeV for jets was guided by the observed p_T distribution of signal AK8 and AK4 jets [62].

Another feature shown in Fig. 7.4 refers to the presence of B hadrons. It is important to note that only the contributions in the plot labelled “with B hadrons” are significant in the signal region of this analysis since fat jets without any B hadrons are very unlikely

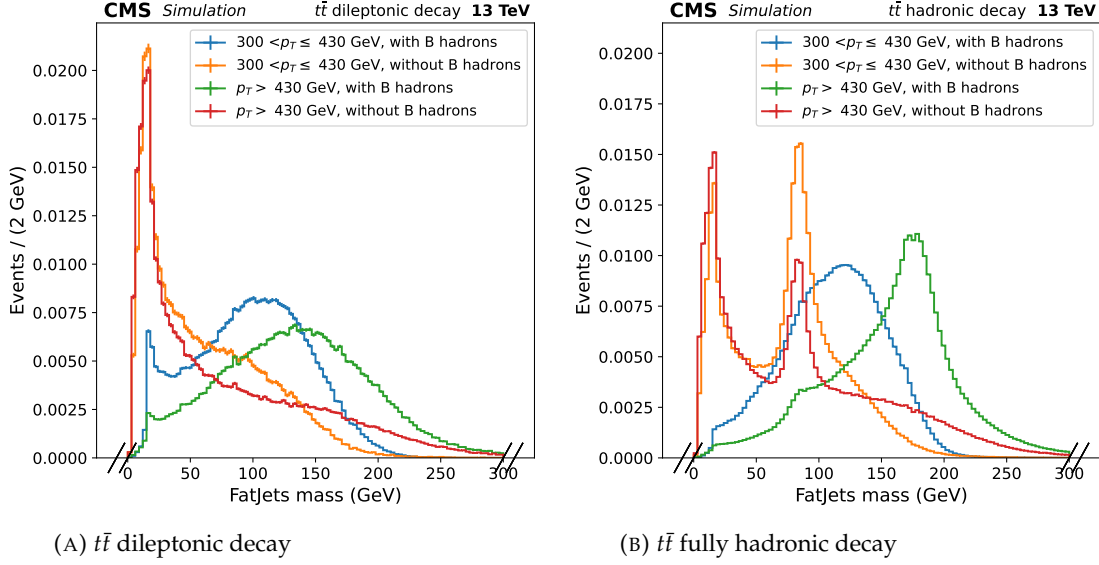


FIGURE 7.4: $t\bar{t}$ contributions to the fat jet mass distributions. The plot is normalized so that the area under each histogram integrates to 1. The $t\bar{t}$ dileptonic decay (A) and $t\bar{t}$ hadronic decay (B) are divided into four different contributions which are the combination of two different conditions: with or without B hadrons, with $(300 < p_T \leq 430)$ GeV, or $p_T > 430$ GeV. The latter is the approximate threshold for a top quark decay to be fully merged inside an AK8 jet.

to pass the double-b-tag selection. The major interest for our study is when we have both B hadrons and $p_T \geq 430$ GeV because these classes of events most closely mimic the analysis's signal.

After applying the **Loose Signal Region** selection of this analysis, the possible decay channels are distributed with the following percentages: fully hadronic $\sim 47\%$, semileptonic: $\sim 44\%$ and fully leptonic (or dileptonic): $\sim 9\%$.

We need to reduce the $t\bar{t}$ background by finding some selections that would identify these events while not removing an appreciable fraction of signal events. In the future section, we will proceed to analyze the following categories of $t\bar{t}$ decay channels in the following order:

- the electron study which comprises *dileptonic* and *semileptonic* decay,
- the muon study which comprises *dileptonic* and *semileptonic* decay,
- the tau study which comprises *dileptonic* and *semileptonic* decay,
- the hadronic top decay study comprises *fully hadronic* and the *semileptonic* decay.

All the figures that will be presented in the next sections are using data collected in 2018 and corresponding MC.

Selection criteria Before exploring individual cases, it is important to establish a key concept shared across the electron, muon and tau study. To achieve the best possible discrimination between signal and background, it is necessary to identify discriminators that selectively favour either background or signal events. To do so the lepton “Physics Object Groups” of the CMS experiment have prepared some common criteria. These criteria are broadly categorized into two main groups: cut-based ID and Multivariate Analysis ID.

Cut-based ID: are based on successive cuts on various particle properties such as track quality, particle quality and global particle status. In this case, the particles are classified into different categories (such as Loose, Medium, and Tight) based on meeting specific requirements. ID-based cuts are straightforward and rely on predefined rules.

MVA ID: these cuts employ machine learning algorithms such as neural networks and decision trees. They are trained using a dataset containing known particle types to recognize and learn the patterns that differentiate between them. The input variables used for the training are similar to the ones used in the cut-based approach.

In the following sections, we will analyze different selections for the different cases. For simplicity, I will use “ID” to refer to “Cut-based ID” and “MVA” to refer to MVA ID selections.

7.2.1 The Electron study

For the electron case, we studied some ID cuts and some MVA cuts. The electron ID selections are the *Tight*, *Medium* and *Loose* cut [73]. The MVA selections are MVA: *Very Loose* (VL), *Loose* (L), *Medium* (M), *Tight* (T), *Very Tight* (VT) and *Extra Tight* (ET). All these labels are characterized by different efficiencies of correctly identifying true electrons.

In Fig. 7.5, all the different selections considered in this study are presented. Figure 7.5 depicts on the y-axis the fraction of NMSSM signal that is removed by each cut and on the x-axis the fraction of $t\bar{t}$ events with at least one electron (one = semileptonic or two = dileptonic case) at the generator level removed by each cut. To determine the optimal selection cut, it is crucial to minimize the impact on signal reduction while maximizing the reduction in $t\bar{t}$ events. The MVAs overall seem to work better compared to the ID cuts. In these cases, we have a lower curve shifted to the right-hand side of the plot showing a better $t\bar{t}$ reduction for the electron case while retaining more signal.

The distinguishing feature of the MVA lies in its ability to effectively differentiate between prompt and non-prompt electrons, rendering it the optimal choice for our analysis. Prompt electrons are produced promptly in the collision. Non-prompt electron originates from the decay of hadrons that have longer lifetimes and can travel a measurable distance before decaying into electrons.

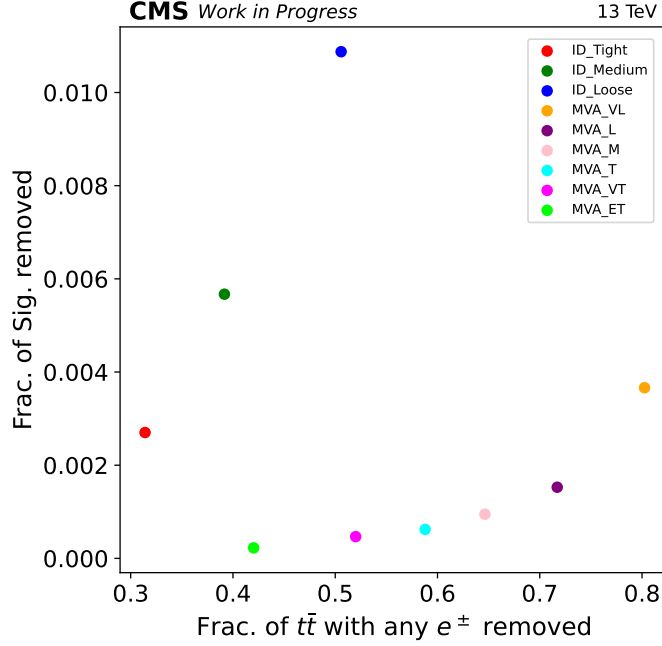


FIGURE 7.5: The fraction of signal vs the fraction of $t\bar{t}$ with at least one electron at the generator level removed by each selection. In this particular instance, the NMSSM sample being plotted corresponds to $m_{\text{SUSY}} = 2800 \text{ GeV}$ and $m_{H_1} = 80 \text{ GeV}$.

We noted that we could use another variable to distinguish between signal and background and this is the $\min_e(\min_{A,B} \Delta R_{j,e})$. While $\min_{A,B} \Delta R_{j,e}$ is the minimum ΔR between the leading electron passing the cut and the two fat jets, the $\min_e(\min_{A,B} \Delta R_{j,e})$ is the minimum ΔR between any electron passing the cut and the two fat jets.

As shown in Fig. 7.6, this variable helps to discriminate between signal and background because it shows two different distributions. While in $t\bar{t}$ background events the electron is expected to be prompt and inside the fat jet, $\min_e(\min_{A,B} \Delta R_{j,e}) < 0.8$, in the signal case there will be significant contributions from non-prompt electrons arising from other jets. For example, non-prompt electrons can arise from photon conversions following neutral pion decay. In Fig. 7.6 we present the five MVAs and the behaviour of $t\bar{t}$ and signal distributions varying on $\min_e(\min_{A,B} \Delta R_{j,e})$. All the events in Fig. 7.6 have at least one electron that has passed the specific cut at the reconstruction level. We noticed that by cutting on this variable, we could retain some signal, especially using selections such as MVA_VL and MVA.L that show more NMSSM events (other cuts have poor statistics). For example, by selecting only events above ~ 0.8 for this study, we would exclude almost all the $t\bar{t}$ background and we could retain all the NMSSM events above that value. The threshold at 0.8 is not necessarily ideal for every MVA cut, and a more optimal value can be found by studying the expected discovery significance of the analysis.

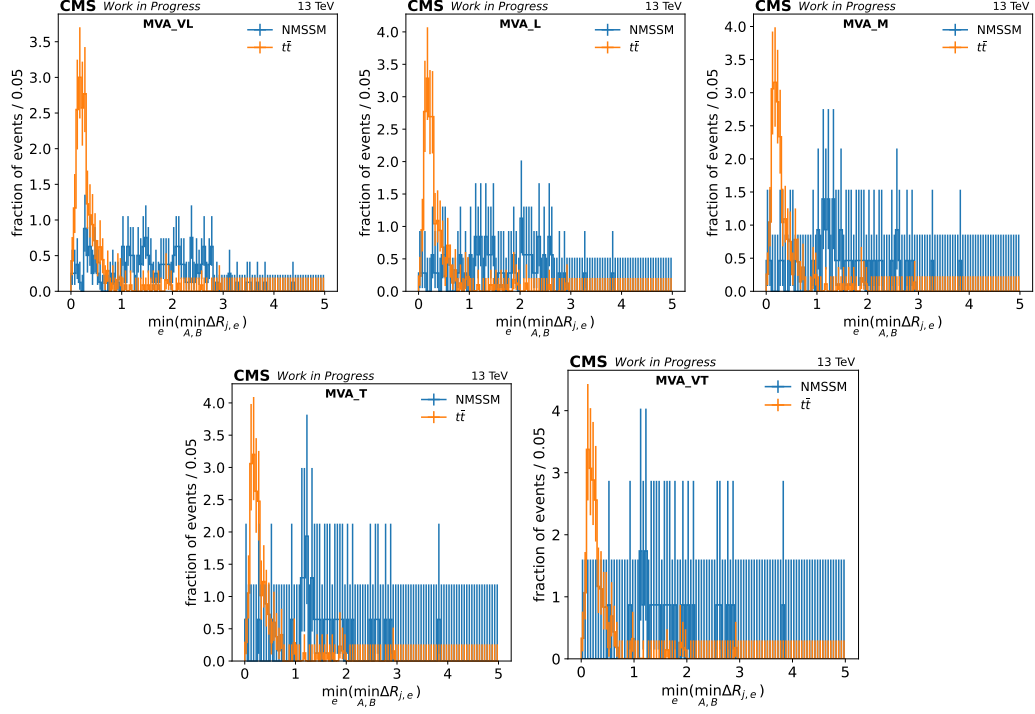


FIGURE 7.6: The distribution of $\min(\min \Delta R_{j,e})$ for NMSSM and $t\bar{t}$ events after applying the different selections.

An approximation for the expected discovery significance of an analysis is suggested by Ref. [74]:

$$\text{Significance} = \frac{S}{\frac{A}{2} + \sqrt{B}}, \quad (7.1)$$

where S represents the expected signal yield, B is the sum of the two main backgrounds of this analysis ($t\bar{t}$ background and QCD) and A represents the number of sigmas corresponding to both the desired significance threshold and the confidence level for setting exclusion limits¹. For a counting experiment in a single signal region, a value of A between 2 and 5 sigmas would be a reasonable choice. However, this does not account for the fact that in this study we are using a combination of the 10 signal regions, so if we want to estimate the significance for one of them we need to divide the background by approximately a factor of 10. However the signal yield is almost unchanged, because signal predominantly populates a single region. If we want to take into account also the division of the analysis in H_T bins the factor B will be further divided. We decided on a choice of $A=20$ to approximately account for these effects for a representative single signal region and verified that the results are not significantly affected by reasonable

¹The particularly simple expression for the significance in Eq. 7.1 is a special case that arises when the number of sigmas for both the desired significance threshold and confidence level for setting exclusion limits are chosen to be the same. The one described above is a figure of merit which allows optimization of the significance, but we will call it “Significance” for simplicity.

variations in the choice of A . This expression for the significance is used from this point onwards.

The significance curves vs $\min_{A,B} \Delta R_{j,e}$ are shown in Fig. 7.7. They have different peaks in significance for different selections. After deciding which ΔR cut is the best for each MVA cut, by taking the maximum of the corresponding curve, we applied these selections to the MVAs and plotted again the fraction of NMSSM and $t\bar{t}$ (with at least one electron) removed in Fig. 7.8. As expected, the new values represented with triangular

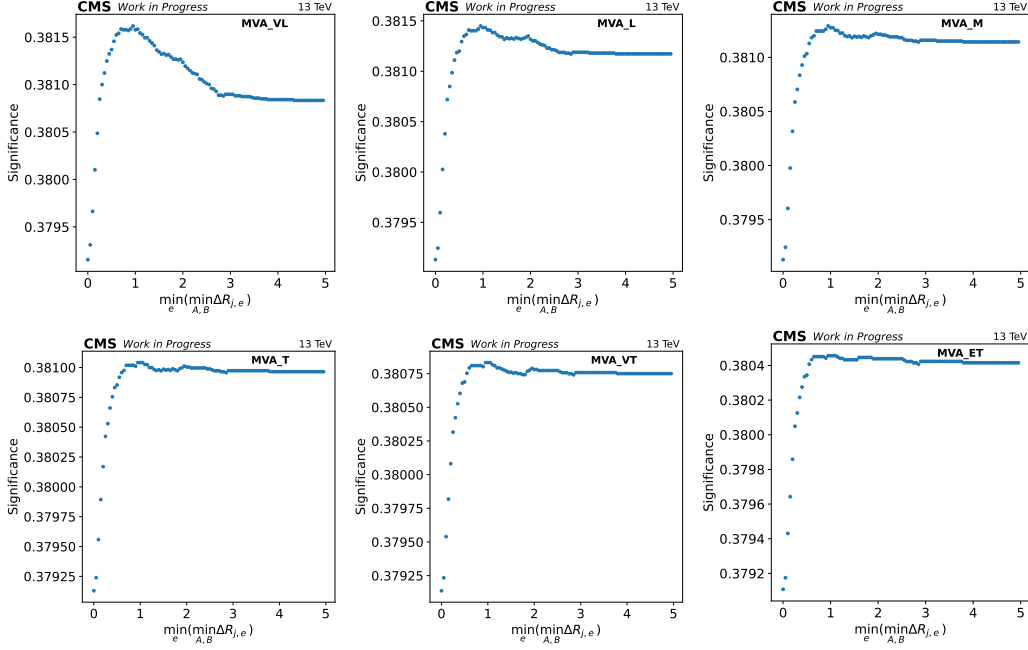


FIGURE 7.7: Significance values vs $\min_{A,B}(\min \Delta R_{j,e})$ for different selections studied for the electron case.

shapes in Fig. 7.8, retain more signal respect to the previous ones (circular shape), sacrificing only a negligible fraction of the $t\bar{t}$ background that was removed by the earlier cuts. The final selection for this study is the MVA.VL with $\min_{A,B}(\min \Delta R_{j,e}) > 0.95$ applied. With this selection, 77% of the $t\bar{t}$ background with electrons is removed and just 0.15% of the NMSSM signal is sacrificed. In the plots just showed, the signal values are specifically $m_{SUSY} = 2800$ GeV and $m_{H_1} = 125$ GeV, but this cut has been proven to be a good selection for all the other mass points of this analysis.

7.2.2 The Muon study

For the muon case, we tested muon MVAs [75] [76] and muon IDs. These latter were inappropriate (as shown in Fig. 7.9 A) because they identified non-prompt muons from our signal $H \rightarrow b\bar{b}$ jets, so they were not effective in discriminating signal from the background.

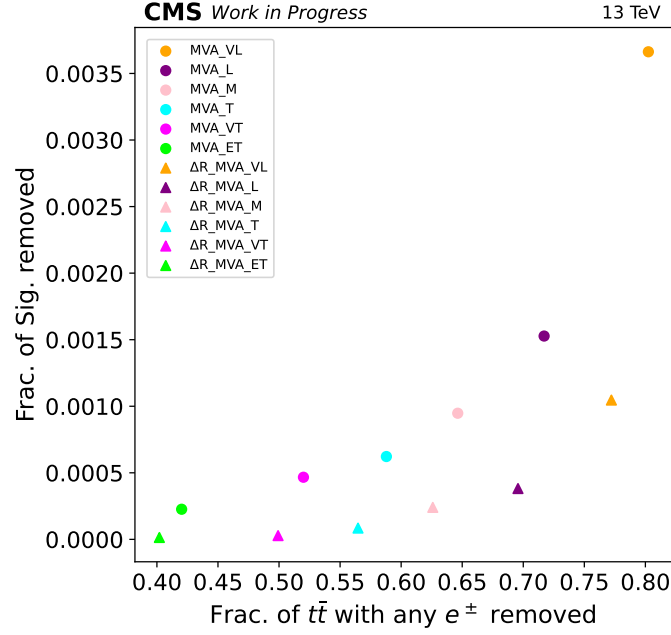


FIGURE 7.8: The fraction of signal vs the fraction of $t\bar{t}$ with at least one electron at the generator level removed by each selection after applying the $\min(\min_{e} \Delta R_{j,e})$ cuts.

Figure 7.9 B presents all the MVAs selection studied for this case: MVA *Loose*, *Medium*, *Tight* and *Very Tight*. The figure presents, on the y-axis the fraction of NMSSM signal that is removed by each cut and on the x-axis the fraction of $t\bar{t}$ events with at least one muon (one = semileptonic or two = dileptonic case) at the generator level removed by each cut. As for the electron case, in order to identify the most effective selection criterion, it is essential to minimize any decrease in signal yield while simultaneously maximizing the reduction in $t\bar{t}$ yield.

In the muon case, the ΔR analysis revealed less effectiveness compared to the electron scenario, as depicted in Fig. 7.10. The plot on the left illustrates the outcome after implementing the MVA *Loose* selection. This selection is presented as an illustrative example, but we verified that the behaviour observed is the same for the other selections as well. There is no obvious tail in this plot for $\Delta R > 0.8$ (or any other similar cut), showing no power to discriminate between signal and background using this variable, $\min(\min_{\mu} \Delta R_{j,\mu})$, in this case. This cut is not helpful for muons because light-quark jets are not a significant source of non-prompt muons (in contrast to electrons).

To confirm this, also the significance plot on the right does not appear to have a defined peak in the significance vs $\min(\min_{\mu} \Delta R_{j,\mu})$, showing that this variable is not the best to recall any lost signal in this case. The Significance is estimated following the Eq. 7.1 discussed for the electron scenario.

To estimate the best cut to apply to the muon case, we plotted a bar graph in Fig. 7.11 showing all the different significance values across the various selections. Although the

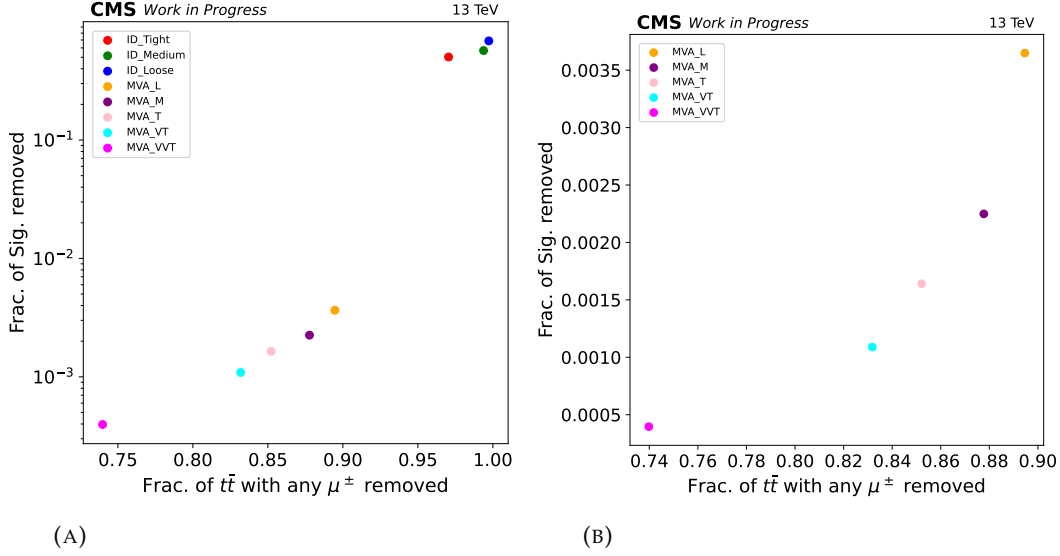


FIGURE 7.9: The fraction of signal vs the fraction of $t\bar{t}$ with muons at the generator level removed by each selection. In plot A, both the ID cuts and the MVA selections are shown, whereas plot B exclusively shows the MVA selections.

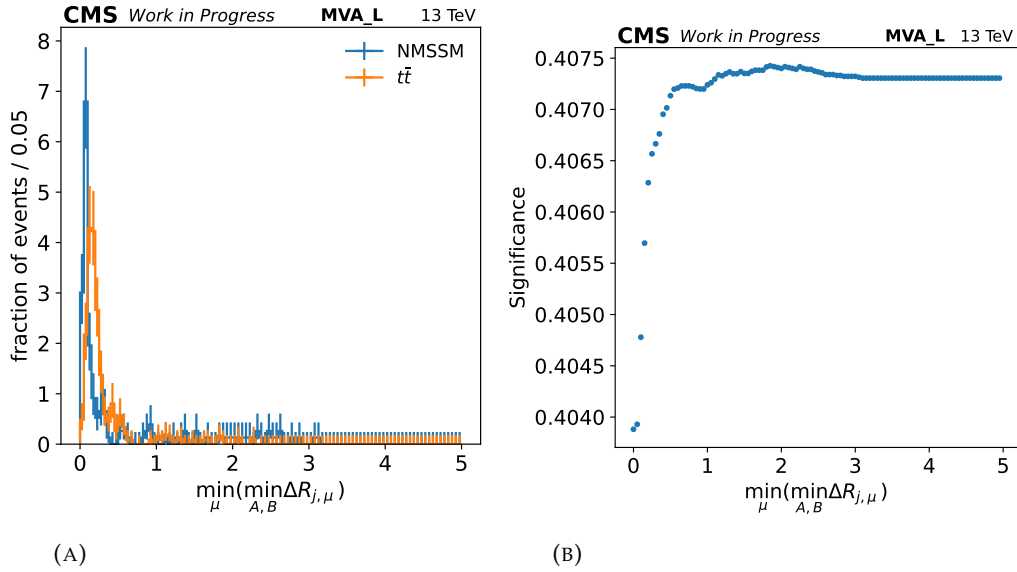


FIGURE 7.10: Plot A is obtained after applying the MVA.L selection. There is no obvious tail in this plot. There is no clear difference in $\min(\min \Delta R_{j,\mu})$ between signal and background. Plot B shows the significance curve vs $\min(\min \Delta R_{j,\mu})$, illustrating that there is no obvious peak in this case.

different selections are very similar in significance, varying among each other by approximately 0.2%, we chose the highest point, which is MVA *Very Tight*. This selection is removing 83% of of the muon $t\bar{t}$ background and just 0.1% of the NMSSM signal. In these plots, the signal values are specifically $m_{SUSY} = 2800 \text{ GeV}$ and $m_{H_1} = 125 \text{ GeV}$, but this cut has been proven to be a good selection for all the other mass points of this analysis. We can notice that even without any additional ΔR cut applied, the fraction

of signal events lost with a muon selection is smaller than the one lost in the electron case.

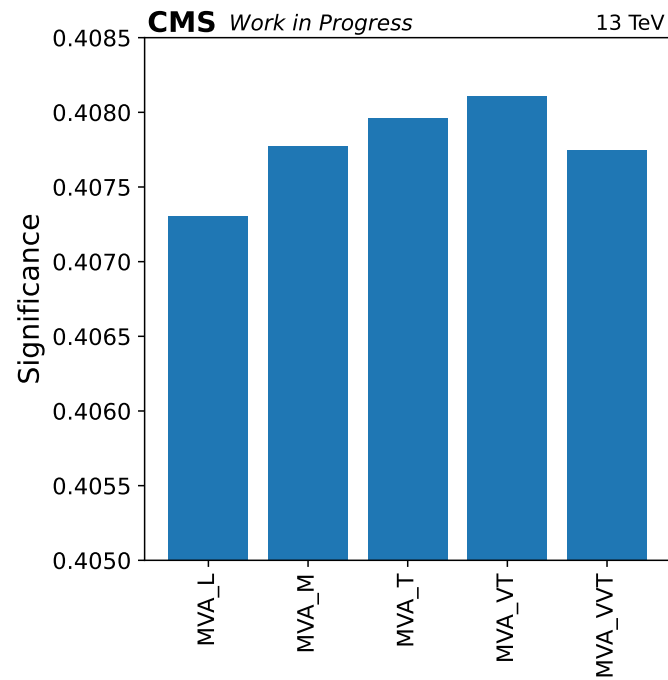


FIGURE 7.11: Bar graph showing the significance values for all the studied selections for the muon case.

7.2.3 The Tau study

The tau component is partially covered by the other studies. Some of the fully leptonic decays are already included by the electron and muon studies. As for the semileptonic case, it will be addressed by the hadronic top veto. The only scenario that remains uncovered is when both W bosons decay into tau and tau neutrino and both taus decay hadronically. The following study focuses on events with no electrons or muons and at least one hadronically decaying tau lepton.

For the tau study, the selections suggested by the Tau POG [77] are only MVA IDs. In Fig. 7.12 we can see the different selections and the relative fraction of signal removed and the fraction of $t\bar{t}$ with at least one τ removed.

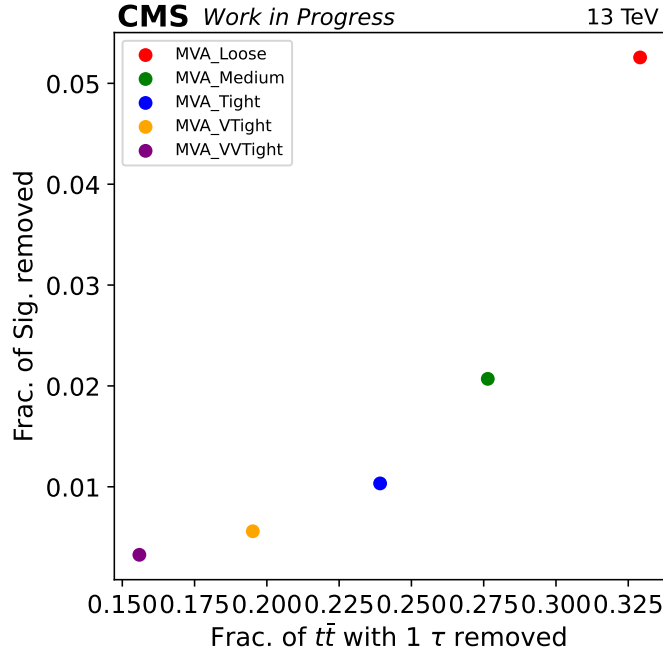


FIGURE 7.12: The fraction of signal vs the fraction of $t\bar{t}$ with at least one tau removed by each selection. In this particular instance, the NMSSM sample being plotted corresponds to $m_{\text{SUSY}} = 2800 \text{ GeV}$ and $m_{H_1} = 80 \text{ GeV}$.

Some of these selection show a characteristic $\min_{\tau}(\min_{A,B} \Delta R_{j,\tau})$ distribution that can be studied to retain some good signal events rejecting the same amount of $t\bar{t}$ background. All the significance curves vs $\min_{\tau}(\min_{A,B} \Delta R_{j,\tau})$ are shown in Fig. 7.13. From some of these curves is visible that the significance decreases with the tau selection applied. The first point in that curve indeed represents the total yield before applying any selection. Only for MVA_VTight and MVA_VVTight there is a small improvement in significance and it is worth it applying a ΔR cut around the value of 1.

Figure 7.14 show the selection after the ΔR is applied.

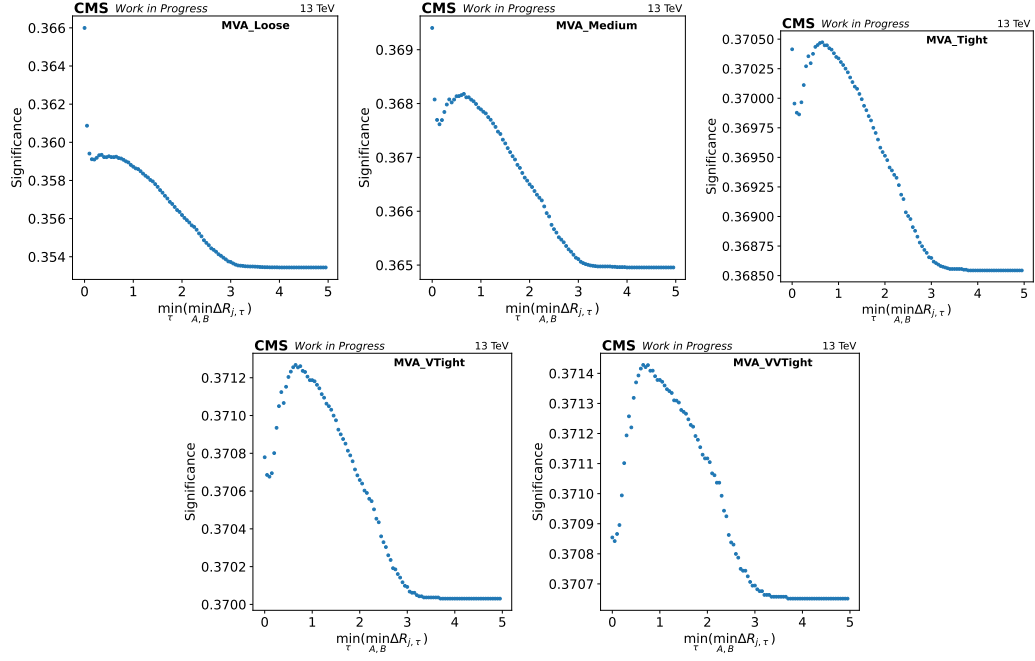


FIGURE 7.13: Significance values vs $\min(\min \Delta R_{j,\tau})$ for different selections studied for the τ case.

MVA_VVTight appears to be the highest selection in significance. The chosen selection, after the ΔR cut, removes 15% of the $t\bar{t}$ containing at least one τ , while removing 0.1% of the signal.

By applying both electron and muon vetoes simultaneously, we achieve a reduction of 39.7% of the whole $t\bar{t}$ background, while when combining electron, muon, and tau vetoes, the reduction increases to 42.8%. However, this improvement is expected to be partially offset by the increased systematic uncertainty associated with the efficiency of the tau ID. Considering this and the impact on computational complexity, we don't think there's a strong motivation for introducing a tau veto to the analysis. In contrast to the substantial improvements from the electron and muon vetoes, the potential enhancement achieved by a tau veto is comparatively modest.

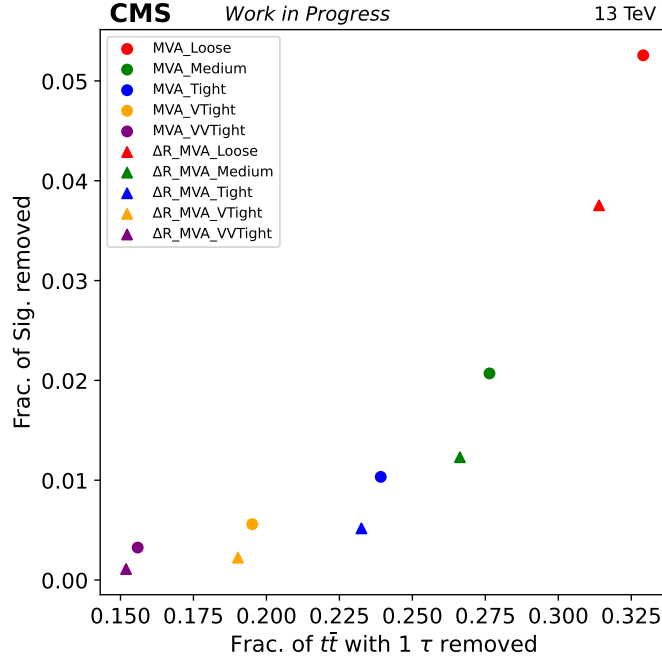


FIGURE 7.14: The fraction of signal vs the fraction of $t\bar{t}$ with at least one tau removed by each selection, after applying the $\min(\min_{\tau} \Delta R_{j,\tau})$ cuts.

7.2.4 The Hadronic study

In addition to the ParticleNet fat jet mass regression and double-b-tagging discriminant used in this analysis, ParticleNet can also be trained for the purpose of discriminating between fat jets from hadronic top quark decays and those from QCD background. For simplicity, we refer to the `FatJet_particleNet_TvsQCD` discriminant as the “ParticleNet-top-tagger.” This tagger represents the relative probability of a fat jet to originate from a merged hadronic top decay.

We introduce two selections based on this variable:

- The 1D discriminant: the maximum value of the ParticleNet-top-tagger among all fat jets.
- The 2D discriminant: the maximum value of the sum of the ParticleNet-top-tagger among any combination of 2 fat jets.

The first discriminant ranges from 0 to 1, while the second ranges from 0 to 2 due to the sum of 2 fat jets. In Fig. 7.15, the distribution of these variables is depicted for the $t\bar{t}$ background in Fig. (A) and for a signal sample with $m_{SUSY} = 2800 \text{ GeV}$, $m_{H_1} = 125 \text{ GeV}$ in Fig. (B).

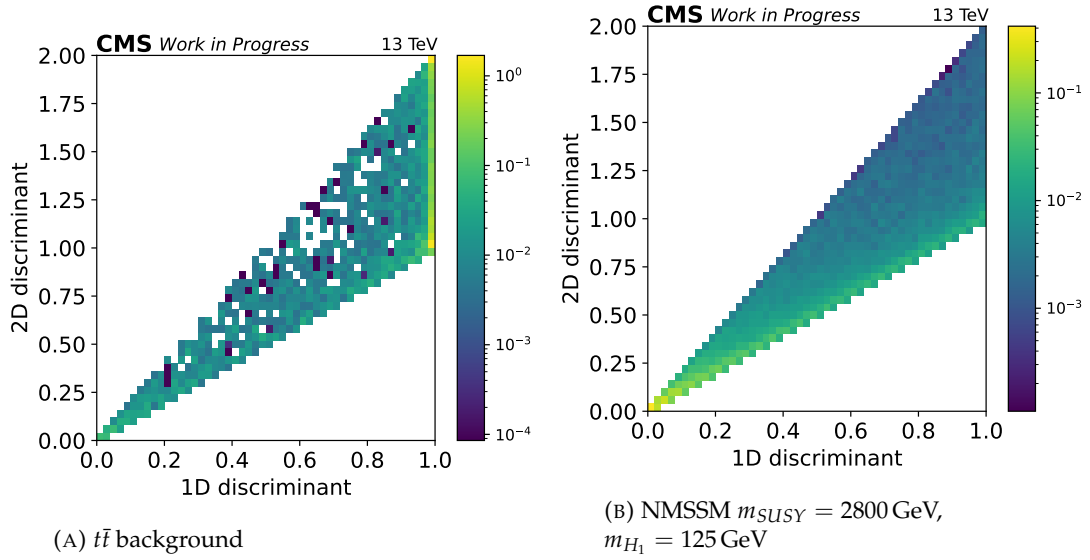


FIGURE 7.15: The 1D discriminant spans a range from 0 to 1, while the 2D discriminant, derived from the sum of two fat jets, ranges from 0 to 2. Figure (A) shows the distribution of $t\bar{t}$ background for both variables, while Fig. (B) illustrates the distribution of a signal sample with a Higgs boson mass of $m_{H_1} = 125$ GeV and $m_{SUSY} = 2800$ GeV for both variables.

The 1D discriminant selection discriminates against residual semileptonic $t\bar{t}$ background as well as the all-hadronic $t\bar{t}$ background. In contrast, the 2D discriminant specifically targets the all-hadronic contribution of the $t\bar{t}$ background, as both tops in this case decay hadronically.

For the signal, the distribution of 1D discriminant and 2D discriminant is notably influenced by the Higgs boson mass, with higher values of m_{H_1} leading to a closer resemblance of $H \rightarrow b\bar{b}$ to hadronic top decays. Exploring all feasible combinations, a consistent zone of interest across different m_{H_1} datasets is identified when the 1D discriminant ranges from 0.91 to 1 and the 2D discriminant ranges from 1.4 to 1.99.

Figures 7.16- 7.18- illustrate a scan on 2D discriminant values (x-axis) and 1D discriminant cuts (coloured curves) for different signal masses. Each curve represents the significance values calculated among the 2D discriminant range and for specific 1D discriminant cut values (coloured curves). The steps in these plots are due to a lack of statistics in the QCD background sample.

To determine optimal single values for the 1D and 2D cuts, we conducted an averaging process over various m_{H_1} values, encompassing 30, 40, 60, 80, 100, and 125 GeV. In Tab. 7.2 are illustrated the ten best values coming from the average in m_{H_1} , sorted from highest significance to the lowest. The derived optimal cuts are identified as 1D-cut < 0.99 and 2D-cut < 1.48. Applying these cuts results in a substantial reduction of the $t\bar{t}$ background to 37.6% of its original magnitude and maintains the Signal yield at 98.9%, on average. These percentages refer to the hadronic top vetoes only.

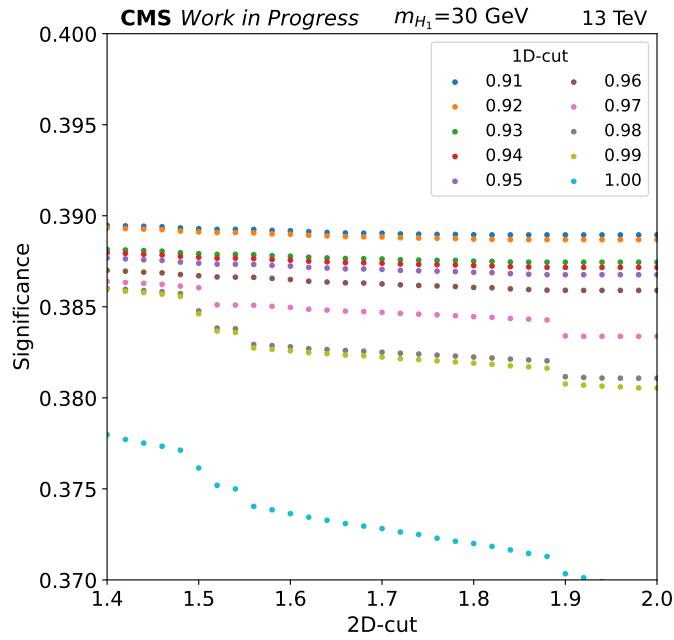


FIGURE 7.16: Significance vs the 2D discriminant scan for a signal mass of 30 GeV for various 1D-cut discriminant values (coloured curved).

7.2.5 Final choice of $t\bar{t}$ veto cuts

Summarizing the impact of the different applied selections, the decision to utilize MVA_VL with $\Delta R < 0.95$ for the electron study and MVA_VT for the muon study was completed by a dual cut in the hadronic study: 1D-cut < 0.99 and 2D-cut < 1.48 .

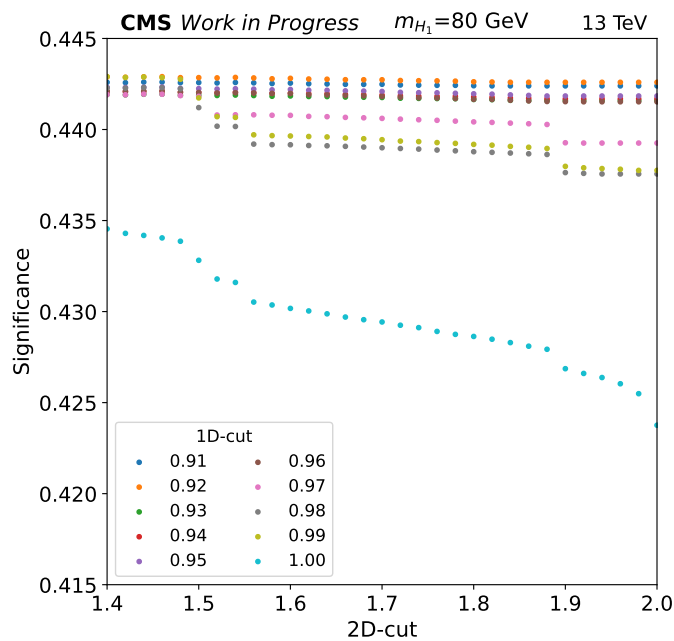


FIGURE 7.17: Significance vs the 2D discriminant scan for a signal mass of 80 GeV for various 1D-cut discriminant values (coloured curved).

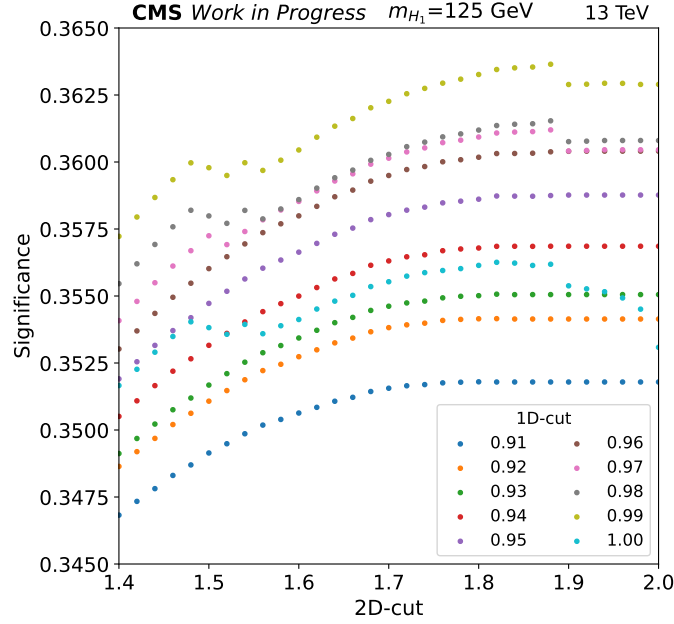


FIGURE 7.18: Significance vs the 2D discriminant scan for a signal mass of 125 GeV for various 1D-cut discriminant values (coloured curved).

Sign. Av. (%)	1D	2D	$t\bar{t}$ (%)	QCD (%)	NMSSM (%)
98.36	0.99	1.48	37.6	93.1	98.9
98.33	0.99	1.46	37.0	93.1	98.8
98.32	0.96	1.82	35.3	91.7	98.4
98.31	0.96	1.78	35.0	91.7	98.4
98.31	0.96	1.80	35.2	91.7	98.4
98.31	0.96	1.76	34.9	91.7	98.4
98.31	0.96	1.84	35.6	91.7	98.4
98.30	0.96	1.86	35.7	91.7	98.4
98.30	0.96	1.90	35.9	91.7	98.4
98.30	0.96	1.88	35.9	91.7	98.4

TABLE 7.2: This table showcases the top ten values derived from the average, arranged in descending order of percentage of maximum significance. The 1D and 2D cuts are presented along with corresponding fractions for signal, QCD, and $t\bar{t}$.

Upon implementing these cuts, only 22.7% of the original $t\bar{t}$ background remains and approximately 98.7% of the original signal (this is an averaged number that varies depending on the specific data sample), and around 92.7% of the original QCD. As shown in Tab. 7.2 the primary cause of the modest reduction of QCD background is the top tag cuts. It is a consequence of QCD jets faking a top (which is more likely for the double-b-tagged QCD jets that appear in our signal region).

Preceding this study, the adoption of ParticleNet resulted in the QCD background being reduced by approximately a factor of 100, and the $t\bar{t}$ background being reduced

by about a factor of ~ 10 compared to the published analysis. The reduction in the $t\bar{t}$ background from the new $t\bar{t}$ veto cuts is substantial, amounting to a factor of ~ 5 , or a total factor of ~ 50 compared to the published analysis. These cuts bring back the QCD background as the dominant background in every signal region.

7.2.6 Uncertainties associated with the $t\bar{t}$ veto cuts

Up to this point, the $t\bar{t}$ study has been based purely on MC simulations, which are used to model the efficiencies of the vetoes introduced in the previous section. However, the efficiencies of these vetoes when applied to $t\bar{t}$ events in the real data are not necessarily well-modelled by the MC, and each one is subject to a data-MC scale factor (SF), defined as the ratio between the efficiency in data and the efficiency in simulation. SFs are measured in control samples of data, and are subject to both statistical and systematic uncertainties. Usually, scale factors are computed in bins of transverse momentum p_T and η .

To evaluate the effectiveness of our method, it is important to quantify the uncertainties related to the numbers presented in the previous section. To investigate these, we analyzed the efficiencies of the various selections applied and calculated their uncertainties.

Scale factors can be estimated using the tag and probe method [78]. In this technique, events selected for tag and probe analysis involve the decay of resonant particles like the Z boson into two same-flavor leptons. One lepton, designated as the “tag”, is required to pass stringent quality criteria, including trigger filters, identification, and isolation requirements. The other lepton, known as the “probe,” is then used to assess the performance of the identification algorithm under study. The properties of the probe lepton can then be used to calculate the efficiency of various identification algorithms. For example, the efficiency for a tight identification working point can be calculated as the number of tight probes within the resonance peak divided by the total number of probes within the peak.

Currently, we have not calculated the scale factors for our selection efficiencies. This is due to the fact that the Run 3 data-taking phase is still ongoing, as well as the Monte Carlo production for Run 3. After the completion of this phase, scale factors will be computed using the tag and probe technique.

At the moment, it is important to assess the magnitude of uncertainties associated with both the scale factors and our selection efficiencies. This evaluation ensures the validity of our methodology and confirms that the reported numbers are accurately calculated and the uncertainties are under control.

To accomplish this, we referred to analogous analyses that employed similar selection criteria, as described in Refs. [79, 80, 81]. By reading these analyses, we estimated

not only the potential range of our scale factors but more importantly the range of uncertainties associated with them.

The relative efficiencies are the efficiency of the electron ID, the muon ID and the efficiency of the top tagging. We can write the number of events in the veto region in terms of these efficiencies. The Veto region is defined as the one characterized by a high concentration of $t\bar{t}$ background, delineated by the cuts detailed in the preceding sections. Conversely, the Signal region represents the opposite scenario, an area where the signal event is expected to prevail, resulting in a notable reduction of $t\bar{t}$ background contamination.

The formula given estimates the number of events in the Veto Region (VR) for different decay channels of $t\bar{t}$:

$$\begin{aligned}
 N_{t\bar{t}}(\text{VR}) = & N_{1e} \times \eta_{1e} \times \text{SF}_e + N_{2e} \times (1 - (1 - \eta_{1e} \times \text{SF}_e)^2) \\
 & + N_{1\mu} \times \eta_{1\mu} \times \text{SF}_\mu + N_{2\mu} \times (1 - (1 - \eta_{1\mu} \times \text{SF}_\mu)^2) \\
 & + N_{2D} \times (\sqrt{\eta_{2D}} \times \text{SF}'_{1D})^2 \\
 & + N_{1D}^{\text{residual}} \times \eta_{1D} \times \text{SF}_{1D} \\
 & + N_{2D}^{\text{residual}} \times (1 - (1 - B \times \eta_{1D} \times \text{SF}_{1D})^2).
 \end{aligned} \tag{7.2}$$

Breaking down the formula in the different components:

- $N_{1e}, N_{2e}, N_{1\mu}, N_{2\mu}$ are the number of events with 1 electron, 2 electrons, 1 muon, and 2 muons respectively. N_{2D} is the number of events with the fully hadronic $t\bar{t}$ decay. N_{1D}^{residual} is the number of events tagged by the 1D discriminant that were not included in the lepton cuts. N_{2D}^{residual} is the number of fully hadronic decay events tagged by the 1D discriminant.
- $\eta_{1e}, \eta_{1\mu}, \eta_{1D}, \eta_{2D}$ are the efficiencies for selecting events with 1 electron, 1 muon, 1 $q\bar{q}$, and 2 $q\bar{q}$ from the $t\bar{t}$ background.
- $\text{SF}_e, \text{SF}_\mu, \text{SF}_{1D}, \text{SF}'_{1D}$ are the scale factors associated with the efficiencies. The SF-related uncertainty in the signal region is the same as in the veto region.
- B is a constant factor (approximately 0.69) correcting for the correlation between the 1D and 2D cuts in the hadronic scenario.

The equation accounts for various channels of $t\bar{t}$ decays and the efficiencies and scale factors associated with each channel. It uses the efficiencies and scale factors to estimate the number of events passing the veto criteria in the Veto Region.

For the first two lines in Eq. 7.2, it has been proven that the efficiencies of tagging one or two electrons (muons) are independent. Therefore, deriving the efficiency of tagging two electrons (muons) from that of tagging one is straightforward.

The third line of Eq. 7.2 makes the approximation that the effect of the 2D cut can be factorized by the product of two 1D cuts, which then allows us to use the scale factor SF'_{1D} derived from a single top tag. This comes from a limitation in the tag and probe method, that using $t\bar{t}$ events can only measure the SF of a single top to be tagged.

The last two terms in Eq. 7.2 are approximations. The first one represents the fact that for the events which do not get vetoed by the lepton IDs but belong to that category, there is another chance they could get vetoed by the 1D top tag. In the last one $N_{2D}^{residual}$ is the number of events that belong to the fully hadronic case, that have not been vetoed by the 2D cut but are vetoed by the 1D cut. B is a constant factor that corrects the correlation between the hadronic 1D cut and the 2D cut.

In order to estimate the uncertainties in $N_{t\bar{t}}(VR)$, we define two possible scenarios for the scale factor uncertainties: an optimistic one based on the systematic uncertainties of the cited existing analysis, and a pessimistic scenario assuming considerably inflated systematic uncertainties.

The systematic uncertainties arise from using the SFs in different phase spaces than those in which they are measured. For example, in the case of the lepton SFs, we are applying them for leptons that are overlapping with fat jets, while the SFs are measured in $Z \rightarrow ll$ events with isolated leptons. Translating these SFs for use in our analysis implies a larger systematic uncertainty.

Similarly, we would ideally want to measure the top tag SF in jets with large double-b-tag scores instead of the more inclusive $t\bar{t}$ sample used in Ref [82]. The evaluation of scale factors needs a dedicated study and as already mentioned it will be better evaluated in future.

All four different scale factors: $SF_e, SF_\mu, SF_{1D}, SF_{2D}$ were set to 1 and the scale factor uncertainties were estimated from other analysis using similar selections [79, 80, 81]. The optimistic scenario is the one where the scale factor uncertainties for the electron and the muon selection are $\sim 2\%$, and on the 1D and 2D cuts are $\sim 5\%$. The pessimistic case scenario is instead the one where the electron and the muon scale factor uncertainties are of the order of $\sim 5\%$, and 1D and 2D uncertainties of the order of $\sim 10\%$. The other sources of error are the four efficiencies themselves. Based on the statistical power of the MC sample for 2018 these numbers will be between 3 and 4%. With more MC events, these numbers will go down.

In the 2018 MC sample, the total number of $t\bar{t}$ events weighted is 27.3. $N_{t\bar{t}}(VR)$ is equal to 20.45 ± 0.73 in the optimistic scenario and 20.45 ± 1.17 in the pessimistic one. $N_{t\bar{t}}(SR) = N_{t\bar{t}}(all) - N_{t\bar{t}}(VR)$ is equal to 6.88 ± 0.73 in the optimistic scenario and 6.88 ± 1.17 in the pessimistic scenario. Obviously, the percentage uncertainty is higher

in the signal region compared to the veto region. The propagation of the errors is always calculated using Mathematica [83], based on Eq. 7.2.

7.2.7 Data-driven $t\bar{t}$ Estimation

We want now to evaluate the number of $t\bar{t}$ expected in the signal region (SR) using a data-driven estimation method based on the data in the VR. Figure 7.19 shows the distributions of data and MC background events in the VR. We calculate how many $t\bar{t}$ events there are in data in the veto region, subtracting from the data the number of QCD events, as we expect to have only QCD and $t\bar{t}$ events in the veto region. We multiply this value by the MC ratio $N_{t\bar{t}}^{MC}(SR)/N_{t\bar{t}}^{MC}(VR)$ and we obtain in this way the number of $t\bar{t}$ in the signal region. We use the MC ratio to predict how many $t\bar{t}$ events will be in the signal region using data:

$$N_{t\bar{t}}^{data}(SR) = [N_{t\bar{t}}^{data}(VR) - N_{QCD}^{MC}(VR)] \times \frac{N_{t\bar{t}}^{MC}(SR)}{N_{t\bar{t}}^{MC}(VR)} \quad (7.3)$$

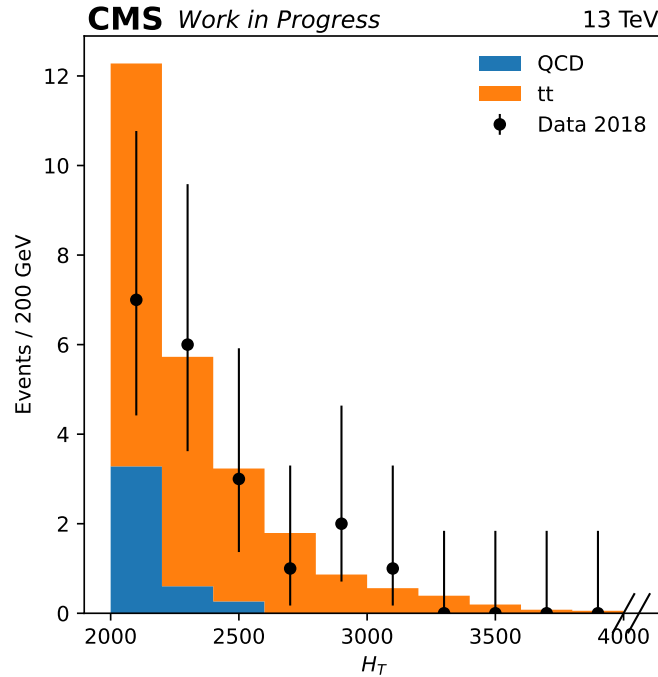


FIGURE 7.19: H_T distribution of 2018 data, QCD and $t\bar{t}$ MC background events in the VR.

Using data from 2018, $N_{t\bar{t}}^{data}(VR) = 20$ and $N_{QCD}^{MC}(VR) = 4.15$. The most important numbers are the MC ratio and its uncertainty, which can be used as projections for the Run 3 analysis. The MC ratio $N_{t\bar{t}}^{MC}(SR)/N_{t\bar{t}}^{MC}(VR)$ is equal to 0.34. The uncertainty on this ratio comes largely from the SFs and is calculated using Mathematica as in Section 7.2.6. This uncertainty is 0.047 (that makes $\sim 14\%$ uncertainty) in the opti-

mistic scenario and 0.076 ($\sim 22\%$ uncertainty) in the pessimistic one. The uncertainty of $N^{data}(VR) - N_{QCD}^{MC}(VR)$ scales like $1/\sqrt{N}$ with the number of data and MC events.

It is important to notice that this method estimates the number of $t\bar{t}$ events in the combination of all signal mass regions and H_T bins. The distribution of events among these bins will be taken from the $t\bar{t}$ MC simulation, meaning there is still some residual effect of the systematic uncertainties in the MC prediction. While a data-driven estimate for individual mass regions might be viable (particularly for the higher mass regions where $t\bar{t}$ is an important background), this is likely not the case for the H_T bins, because the number of $t\bar{t}$ events (and thus events in the VR in data) falls off rapidly with H_T , as shown in Figure 7.19. Thus, instead of relying on $t\bar{t}$ MC for both the normalization and shape of the $t\bar{t}$ background, we now take the normalization in the Loose Signal Region from the data with a systematic uncertainty determined from the SFs (Eq. 7.3).

We conclude by saying that the $t\bar{t}$ background has been reduced such that it is small compared to the QCD background and its uncertainties are well under control.

7.2.8 Discussion of Expected Limits

The ultimate objective of this analysis is to publish a comprehensive study integrating data from both Run 2 and the ongoing Run 3 data-taking period. The publication of this analysis is expected following the conclusion of Run 3. The work presented in this thesis is crucial for achieving this objective.

Although drawing definitive comparisons based solely on the 2018 data analyzed in this thesis is premature, we can still make preliminary observations based on our current findings. Comprehensive comparisons will only be possible after analyzing the full Run 2 dataset (including 2016 and 2017) and incorporating data from the ongoing Run 3.

Compared with the published analysis, we have achieved an increase in signal efficiency of a factor 2 as discussed in Section 7.2, together with a substantial reduction in background, with the QCD background decreased by a factor of 100 and the $t\bar{t}$ background reduced by a factor of 50 across all signal regions.

Reviewing a plot from the published paper allows us to make an observation. The plot in Fig. 7.20 shows different m_{SUSY} values at fixed $m_{H_1} = 70$ GeV signal, together with all different background sources. We expect a background of ~ 0 in the last HT bin, where the background sources were already very low. Looking at $m_{SUSY} = 2800$ GeV signal in the last HT bin about 4 events were expected which now turn to approximately 8 events, thanks to the enhancement in signal efficiency. The expected signal yield for $m_{SUSY} = 2800$ GeV is large enough that it would be excluded at the 95% CL if 0 or 1 events were observed in the SR (as shown in Table 40.3 of Ref. [84]).

We should be able to comfortably exclude this signal using the Run 2 data alone. When including Run 3 data, the expected cross-section limits should improve further, scaling approximately with square root of the integrated luminosity, $\sqrt{\int L(t)dt}$, because the analysis is statistically limited in the high HT bin.

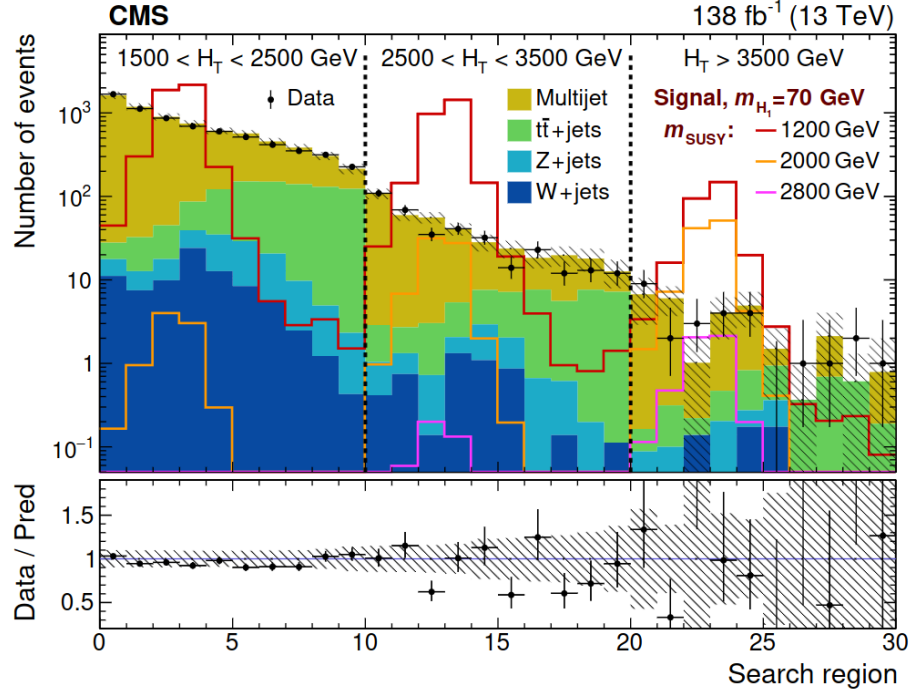


FIGURE 7.20: Observed and predicted yields in the TR for each of the 30 search regions, aggregated across the three years of data collection (2016-2018). The multijet background (QCD) is estimated from data while other backgrounds are simulated. Example signal distributions are provided for $m_{H_1} = 70$ GeV and $m_{SUSY} = 1200, 2000,$ and 2800 GeV. Error bars indicate statistical uncertainties, while the hatched bands represent systematic uncertainties. Figure from Ref. [14]

Part IV

A new algorithm for the L1 trigger at CMS at HL-LHC

Chapter 8

The CMS Phase-2 upgrade

8.1 The High-Luminosity Large Hadron Collider

The High-Luminosity Large Hadron Collider (HL-LHC) project has a mission to elevate the LHC's performance, extending the potential for groundbreaking discoveries. The primary goal of HL-LHC is to amplify the integrated luminosity by a factor of 10 beyond the originally designed value of the LHC. A timetable for the Large Hadron Collider (LHC) that goes from 2011 to 2029 is shown in Fig. 2.2. HL-LHC is anticipated to start operating in early 2029

In its ultimate configuration, the HL-LHC is expected to achieve a peak instantaneous luminosity of $7.5 \times 10^{34} \text{ cm}^2\text{s}^{-1}$, leading to an increase in the average number of proton-proton collisions per bunch crossing (pileup) to approximately 200. This enhanced performance could facilitate the collection of 400 to 450 fb^{-1} of integrated luminosity annually, potentially providing a total of 4000 fb^{-1} for each of the CMS and ATLAS experiments.

Given these challenging running conditions, the CMS detector needs a trigger and data acquisition system with outstanding performance. The upgrade of the trigger system [16] aims to improve physics selectivity and ensure sustained performance over the entire 10-year HL-LHC program, which includes heavy ion operations.

8.1.1 Overview of the CMS detector upgrade

The CMS upgrade planned for HL-LHC includes the replacement of both the Strip and Pixel tracking detectors. The Inner Tracker will incorporate small-sized pixel sensors, while the Outer Tracker will feature strip and macro-pixel sensors, extending their coverage to $|\eta|=3.8$ [8]. Stacked strip modules in the Outer Tracker will reduce hit multiplicity, enabling the reconstruction of track candidates for the L1 trigger up to $|\eta|=2.4$,

thereby introducing significant possibilities for the L1 trigger. More about the tracker upgrade will be discussed in Section 8.1.3.

For the barrel calorimeters, readout electronics will be replaced to achieve finer granularity and provide timing information. The endcap calorimeters will be substituted with the high-granularity calorimeter [85] (HGCAL), featuring over 6 million readout channels. This sampling calorimeter is designed to facilitate shower separation and identification, specifically adapted to the harsh conditions in the forward region of the detector.

The redundancy of the muon detection system, achieved through a combination of drift tubes (DTs), resistive plate chambers (RPCs), and cathode strip chambers (CSCs), will be maintained with upgraded electronics [86]. Additional improved RPC (iRPC) chambers and gas electron multiplier (GEM) chambers will be installed to extend the coverage up to $|\eta|=2.4$ and $|\eta|=2.8$, respectively. Introducing a minimum ionizing particle (MIP) timing detector (MTD) in front of the barrel and endcap calorimeters will enable precise timing measurements of charged tracks.

In tandem with sub-detector upgrades, a comprehensive replacement of the trigger (L1 and HLT) and data acquisition (DAQ) system is planned [16].

Similar to the Phase-1 (Run 3 setup), the Phase-2 (HL-LHC setup) HLT will have access to the full granularity of the detector, with an input rate of 750 kHz in the ultimate HL-LHC scenario. The selection algorithms are designed to achieve rate reduction, resulting in an output bandwidth of 7.5 kHz.

8.1.2 Physics motivations and objectives

The CMS Phase-2 physics programme aims to fully use the capabilities of the HL-LHC to conduct explorations into new physics beyond the Standard Model (BSM) and achieve exceptionally precise measurements of the Standard Model (SM). Increased luminosity will allow for more precise measurements of Higgs properties, such as couplings and branching ratios, and enable the observation of rare decays and production modes. Additionally, the larger dataset will facilitate studies of Higgs self-coupling providing deeper insights into the Higgs potential. The availability of unprecedentedly large data samples from the HL-LHC and the enhanced capabilities resulting from the detector upgrade will enable a diverse range of physics analyses. To effectively navigate the challenges posed by the enlarged HL-LHC data samples, particularly in the presence of 200 pileup (PU) events, more sophisticated selection algorithms are needed at L1. Additionally, the upgraded detector needs advanced triggers to specifically target certain event topologies, such as rare B-meson decays. Furthermore, a dedicated scouting system will be employed to identify low-mass resonances.

8.1.3 Upgrade of the Outer Tracker

While the previous section offered an overview of the upgrade of all individual sub-detectors of the CMS experiment, the focus of this section is specifically on the Outer Tracker (OT) upgrade.

This upgrade is directly related to the work I have conducted on the vertex finding algorithm, which will be discussed in Section 9. This algorithm relies on tracks firstly selected by the outer tracker.

The CMS tracker will be completely replaced, both the inner (IT) and outer (OT) trackers, with better radiation-tolerant materials, to cope with the very high radiation level that will be present at the HL-LHC. Both the inner and outer trackers' angular coverage will be expanded, improving their sensitivity to forward physics events. A fourth of the improved tracker's layout is shown in Fig. 8.1, with notable differences from the Phase-1 tracker structure shown in Fig. 2.6. One of the largest upgrades is to the pixel detector which will have more endcap layers increasing the pseudorapidity range to 4, which, when combined with the high granularity endcap calorimeters, results in far better particle flow reconstruction at high η .

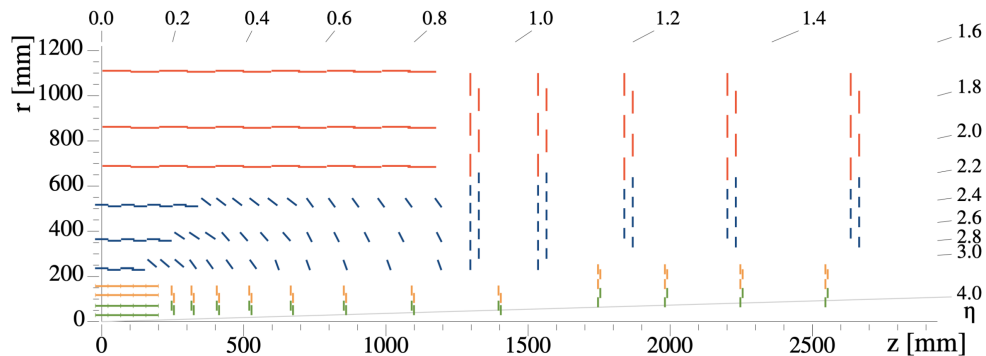


FIGURE 8.1: One-quarter of the Phase-2 CMS tracker in the $r - z$ plane. The inner tracker is represented in green and yellow, with varying pixel sizes across different modules, with smaller pixels positioned closer to the interaction point. The outer tracker is depicted in blue and red. The blue sections correspond to pixel strip modules, utilizing both pixel and strip detectors. The red sections represent two strip module detectors that solely utilize strip detectors [8].

The main feature of the OT in Phase-2 is the ability to send information to the L1 trigger for track reconstruction. Utilizing tracking data in the L1 trigger requires the tracker to transmit self-selected information during each bunch crossing. This functionality depends on local data reduction within the front-end electronics to manage the data volume transmitted at the 40 MHz rate. This is accomplished through the utilization of modules capable of rejecting signals from particles with a p_T below a specific threshold, known as p_T modules. These p_T modules consist of two closely spaced layers of silicon with a tunable window allowing for a cut for charged particles with a $p_T > 2$ GeV. A threshold of 2 GeV results in a reduction in data volume of approximately one order

of magnitude. This reduction is adequate to facilitate the transmission of stubs at the required 40 MHz rate.

This concept is shown in Fig. 8.2 with the acceptance window shown in green. As a charged particle traverses the CMS detector in the $r - \phi$ plane it curves in the magnetic field, with the radius of curvature inversely proportional to the particle's p_T . A high p_T particle passing through two silicon layers follows a straight path falling within a specific geometric window and is registered as a "stub" for L1 reconstruction. A stub is defined as the hit information from a particle that passes through both layers of the silicon detector within this window. Low p_T particles, which have more curved trajectories and fall outside this window, are not utilized for L1 reconstruction. The p_T modules employ various sensor spacings, tailored to optimize the desired p_T filter across different regions of the detector.

The hits in the detector layers are stored regardless of their p_T so that the full p_T spectrum of tracks can be read out and reconstructed in the HLT and offline.

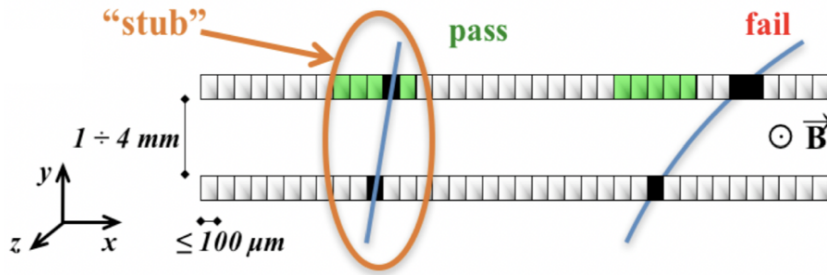


FIGURE 8.2: An example of the p_T module concept. Low- p_T particles can be rejected by correlating signals from closely spaced sensors; the green-coloured channels indicate the selection window that defines an acceptable stub. [8]

8.2 The upgrade of the CMS Trigger System

The Phase-2 upgrade of the L1 trigger system aims not only to sustain signal selection efficiency at the Phase-1 performance level (discussed in Section 2.2.8) but also to significantly enhance the identification of potential new physics manifestations. This involves expanding the phase space for high-precision measurements of physics processes, including improved trigger coverage in the forward region of the detector. Crucially, advanced techniques employed in offline reconstruction and analyses, such as global event reconstruction based on particle-flow methods, become viable at the L1 trigger with the introduction of L1 tracks from the upgraded Outer Tracker. The increased granularity of calorimeter information further supports these algorithms. The utilization of Field Programmable Gate Arrays (FPGA) and high-speed optical links, building on Phase-1 experience, enables hardware implementation of trigger processing capabilities.

The Phase-2 upgrade maintains a two-level strategy while increasing the L1 maximum rate to 750 kHz, with a total latency extended to 12.5 μ s. **This allows for the inclusion of tracker information for the first time.**

The extended latency facilitates higher-level object reconstruction, identification, and evaluation of complex global event quantities and correlation variables, optimizing physics selectivity. This upgrade opens the door to implementing sophisticated algorithms, including particle-flow reconstruction techniques and machine-learning approaches. Additionally, a proposed 40 MHz scouting system captures trigger primitives from sub-detectors and trigger objects produced at various levels of the system. This scouting system uses reconstructed Level-1 trigger objects and quantities to perform real-time data reduction and analysis. The Level-1 scouting system offers the advantage of systematically exploring correlations among multiple contiguous bunch crossings, enabling the identification of potential signatures beyond the reach of standard trigger selection processes.

Various architecture designs were thoroughly examined, ultimately guiding the final decision towards a central "Correlator Trigger" system where advanced triggering algorithms receive backend information.

In tandem with the L1 upgrade, substantial enhancements are envisaged for the Data Acquisition (DAQ) and High-Level Trigger (HLT) systems.

Technological challenges

Complexity will increase for the detectors and related readout electronics in the demanding HL-LHC operating environment. The primary technological features of the upgraded CMS L1 trigger system include:

- Extensive use of Field Programmable Gate Arrays (FPGAs) and processors is planned to achieve optimized reconstruction, identification, isolation, and energy calibration of trigger candidates.
- High-speed optical links will be employed to facilitate the aggregation of data from the entire detector, enabling complex global processing.
- The trigger system will adopt a flexible and modular architecture that can be re-configured to adapt to diverse HL-LHC running conditions and varied physics requirements.

Figure 8.3 illustrates the functional diagram of the Phase-2 trigger system's architecture and data flow.

To balance flexibility and robustness, four independent data processing paths have been implemented: tracking, calorimetry, muon systems, and particle-flow techniques.

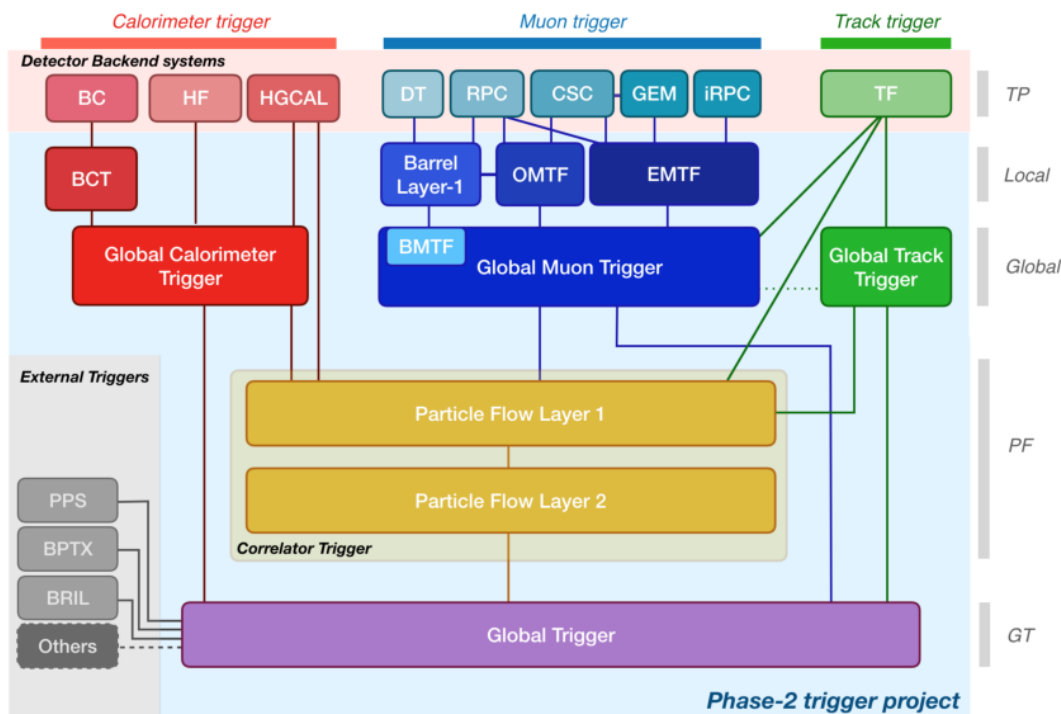


FIGURE 8.3: Functional diagram of the CMS L1 Phase-2 upgraded trigger design. The Phase-2 L1 trigger integrates inputs from the calorimeters, muon spectrometers, and the track finder. Calorimeter trigger inputs encompass data from the barrel calorimeter (BC), high-granularity calorimeter (HGAL), and hadron forward calorimeter (HF). This comprises a barrel calorimeter trigger (BCT) and a global calorimeter trigger (GCT). The muon trigger receives inputs from detectors including drift tubes (DT), resistive plate chambers (RPC), cathode strip chambers (CSC), and gas electron multipliers (GEM). It consists of a barrel layer-1 processor and muon track finders, processing data from three distinct pseudorapidity regions-BMTF, OMTF, and EMTF for barrel, overlap, and endcap, respectively. Muon track finders transmit muon candidates to the global muon trigger (GMT), allowing combination with tracking information. The track finder (TF) provides tracks to various components, including the global track trigger (GTT). The correlator trigger (CT) at the centre (yellow area) has two layers dedicated to particle-flow reconstruction. All objects feed into the global trigger (GT) for the final L1 trigger decision. External triggers, such as potential inputs from the MTD, are also depicted. Dashed lines represent links with potential exploitation. Components under development within the Phase-2 L1 trigger project are grouped in the light blue area. Various processing levels are indicated on the right: trigger primitives (TP), local and global trigger reconstruction, particle-flow trigger reconstruction (PF), and global decision.

Calorimeter Trigger Path The system utilizes a barrel calorimeter trigger (BCT) and the HGCAL backend to process high-granularity calorimeter information, generating high-resolution clusters and identification variables for subsequent stages. Outputs from the BCT, HGCAL, and the hadron forward calorimeter (HF) are directed to a global calorimeter trigger (GCT), where calorimeter-only objects such as e/γ candidates, hadronically decaying tau leptons candidates, jets, and energy sums are constructed.

Track Trigger Path The track finder (TF) reconstructs tracks from the Outer Tracker, forming a crucial component of the Phase-2 trigger system. Processed track parameters will be provided to achieve precise vertex reconstruction and matching with calorimeter and muon objects. A global track trigger (GTT) is introduced to reconstruct primary vertices along with tracker-only-based objects like jets and missing transverse momentum. The GTT can also be utilized to propagate extra copies of tracks to any part of the trigger system, potentially reducing interconnections and simplifying logic.

Muon Trigger Path The muon track finder algorithms process trigger primitives (TP) in three separate regions: barrel (BMTF), overlap (OMTF) and endcap (EMTF). Standalone muon stubs and L1 tracks are sent to a global muon trigger (GMT), where muon stubs contain reconstructed local information extracted from detector hits, including position, bend angle and timing. The GMT generates track-matched muons and L1 tracks matched to muon stubs, known as tracks plus muon stubs. This is achieved through the propagation of tracker tracks through muon system layers. The GMT receives tracks directly from the TF based on the baseline design.

Particle-Flow Trigger Path The correlator trigger (CT) plays a central role, implementing sophisticated algorithms for higher-level trigger object production, particle identification and providing a sorted list of objects to the global trigger. The CT is structured in two layers: Layer-1 produces particle-flow candidates from the matching of calorimeter clusters, muons information and tracks and Layer-2 builds and sorts final trigger objects. The GTT transmits primary vertex candidates to CT Layer-1. Additionally, Layer-2 can receive tracker-only objects from GTT, track-matched muons (and/or muon track plus stubs) from GMT, and calorimeter-only objects from GCT applying further isolation criteria as needed. Isolation of muons may also be calculated in the GMT depending on available resources.

Global Trigger Outputs from GCT, GMT, GTT, and CT are combined in the global trigger (GT), calculating a trigger decision based on a menu of algorithms. The GT evaluates sophisticated correlation variables among various object types to enhance

selectivity. The Level-1 Accept signal is transmitted to the Trigger Control and Distribution System (TCDS), initiating readout to the Data Acquisition System (DAQ). The GT serves as the interface to external triggers including: precision proton spectrometer (PPS), beam position and timing monitors (BPTX) and luminosity and beam monitoring (BRIL) detectors.

8.2.1 The Global Track Trigger

The global track trigger serves the purpose of conducting global event-level reconstruction by utilizing tracks provided by the track finder. An essential function of the GTT is the **reconstruction of the hard interaction**, a critical task that allows the rejection of pileup interactions through the association of tracks and objects to vertices.

The GTT is one of the components that receives tracks from the track finder (TF) as shown in Fig. 8.3. The GTT boards will host three different algorithms: the Vertex Finding algorithm, the Jet reconstruction algorithm and another algorithm to compute missing E_T . The algorithm that will be presented in the following section aims to serve as the Vertex Finding algorithm for implementation on the GTT boards. To be implemented on GTT boards there are constraints in resources and latency that will be discussed in Section 9.3.2, where also additional details about the GTT will be provided.

Chapter 9

The Peak Finding Algorithm

This section will discuss the implementation of a vertex finding algorithm that will be referred to as the "Peak Finding Algorithm" (PFA). Its primary objective is to reconstruct the hard vertex, also called primary vertex (PV), for each event at HL-LHC.

The chapter aims to provide a comprehensive description of the Peak Finding Algorithm and to explore its potential applications within the specific environmental conditions present in the HL-LHC.

This algorithm has undergone modifications and adaptations over time. In the subsequent sections, a chronological approach will be adopted to outline its evolution. The series of tests conducted and the diverse considerations that eventually culminated in the final definition of the algorithm will be explained.

Until now, the approach which is planned to be utilized as the vertex finding algorithm in the GTT system is known as FastHisto (FH) [16]. The PFA yields competitive results, comparable to those of the FastHisto algorithm while developing certain features absent in the latter.

Simple algorithms such as the PFA and the FH are easily implementable on Field Programmable Gate Arrays (FPGAs) and for this reason, ideal to be used in the GTT system. More about this will be discussed in Section 9.3.

Most of the plots in this study are done using an inclusive $t\bar{t}$ production sample with 200 pileup (PU) events overlaid. When the used sample is different, it will be specified.

9.1 Designing the algorithm

In CMS the distribution of the pp interaction vertices extend along the z-axis, where the beam spot profile is Gaussian shaped, with a RMS width of about 5 cm. A minimal spread is observed in the x-y plane. Consequently, our focus will be directed along the z-axis, which separates the pileup vertices.

One of the main variables used in this study is the track's z_0 which is the longitudinal impact parameter (shown in Fig. 9.1), which corresponds to the z coordinate of the point of closest approach of the track to the z -axis. From this point on, we will refer to the z_0 parameter as “ Z ” for simplicity.

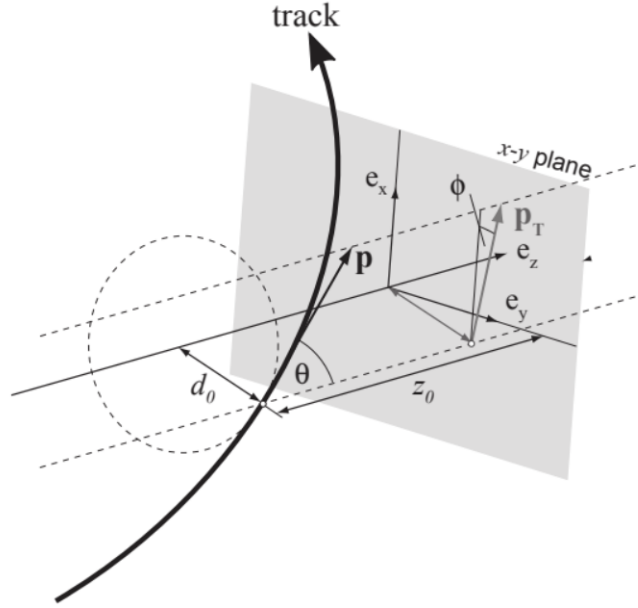


FIGURE 9.1: Sketch of the transverse impact parameters (d_0) and longitudinal impact parameters (z_0) of a track inside the detector. Figure from Ref. [15].

A single CMS event contains multiple pileup vertices along with the primary vertex of interest. The measurement uncertainties of the tracker induce a Gaussian-shaped smearing to the track's Z measurement. The width of the Gaussian distribution is linked to the tracker resolution. As a first approximation, we assume that this resolution is fixed and equal for all tracks. If we look at the distributions of the **reconstructed** Z of the tracks produced in an event, there will be multiple clusters of tracks, each with a Gaussian shape distribution centred around the interaction vertices. Some particles have a long lifetime such as B hadrons so their decay products will not originate from interaction vertices. This may slightly widen the Gaussian distribution. Recognizing these patterns can help in identifying most of the interaction vertices.

The PFA algorithm is illustrated through a simple sketch depicted in Fig. 9.2. Figure 9.2 (A) shows the ensemble of tracks along the z -axis. Most of these tracks, depicted in yellow, come from pileup vertices and the ones represented in red come from the primary vertex. The PFA algorithm uses the Z and p_T information of the tracks in two different steps. In the first step, it evaluates the sum over all tracks of a Gaussian weight, $f(Z)$, at points along the z -axis spaced δ from each other:

$$\sum_{i=1}^N f(Z_i). \quad (9.1)$$

The Gaussian weight of each track is defined as

$$f(Z_i) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{Z_i - \mu}{\sigma}\right)^2}, \quad (9.2)$$

where σ is the Gaussian width, Z is the longitudinal impact parameter of the track and μ is the position of the center of the Gaussian. In Fig. 9.2 (A) this sum is evaluated in three points that are a distance δ from each other. For illustrative purposes, in this sketch δ is chosen to be large, in reality, there is more overlap between the Gaussians. The distribution of the output of this first step, as a function of Z , has distinct peaks, and this is shown both in Fig. 9.2 (B) (with magenta dots) and in a more realistic representation of an event in Fig. 9.3 (blue line).

After the evaluation along the whole z -axis range, distinct peaks emerge in the outputs. The Z position of each of these local maxima is associated with a vertex candidate, concluding the first step of the PFA algorithm.

The next step only considers the local maxima found by the first step. Only for this subset of points, another weighted sum takes into account not only the Gaussian weight but also the transverse momentum of each track:

$$\sum_{i=1}^N f(Z_i) p_{Ti}. \quad (9.3)$$

The local maxima are ranked with respect to the output of Eq. 9.3. The point associated with the highest output is identified as the location corresponding to our primary vertex candidate in that event. This second step is shown in Fig. 9.2 (B) (with the blue dot) and in Fig. 9.3 (with red markers).

In the following section, we will discuss the evaluation and optimization of the algorithm's parameters.

9.1.1 Distribution of the residuals and algorithm's parameters

The algorithm's parameters include the width of the Gaussian distribution, used to evaluate the weighted sums, which will be called *Gaussian width* or σ , and the *step size* or δ , indicating the distance, along the z -axis, between the points in which the algorithm is evaluated. The PFA performances are evaluated by looking at $Z_{reco} - Z_{true}$ distribution, where Z_{reco} denotes the Z -coordinate reconstructed by the algorithm, and Z_{true} represents the true position of the simulated primary vertex from the Monte Carlo samples.

Establishing a truth-matching criterion aids in assessing the vertex reconstruction quality by evaluating how many of the tracks which contribute to the reconstructed vertex are actually generated from the primary vertex. We define ϕ which is the ratio between

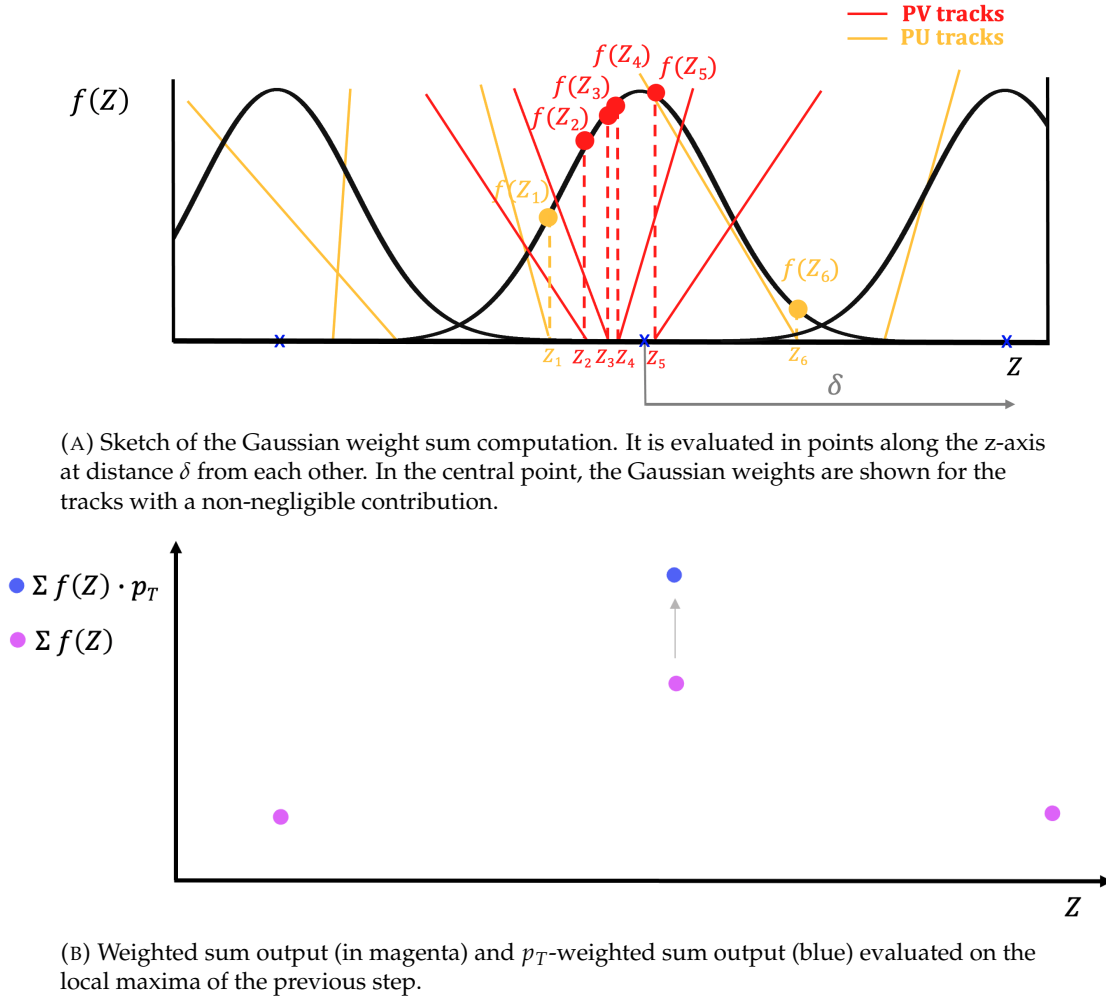


FIGURE 9.2: Visual description of the simplest implementation of the Peak Finding Algorithm.

the cumulative weights of MC true tracks utilized in reconstructing the primary vertex and weights of all tracks involved in the reconstruction process,

$$\phi = \frac{\sum_{i=1}^N f(Z^i) p_{T, \text{TRUE}}^i}{\sum_{i=1}^N f(Z^i) p_{T, \text{ALL}}^i}. \quad (9.4)$$

This metric provides valuable insights into the accuracy and reliability of our reconstruction methodology, offering a nuanced understanding of the track quality achieved.

In Fig. 9.4, the distribution in light blue indicates the events with $\phi > 0.1$, so events in which a minimum of 10% of the tracks originate from the true Monte Carlo (MC) vertex, while the surrounding data depicted in red represents noise, typically originating from events with numerous pileup occurrences as less than 10% of true tracks are used in the PV reconstruction. This approach shows that the pile-up tracks have a smooth distribution in $Z_{\text{reco}} - Z_{\text{true}}$ while the “signal” ones have an approximately Gaussian distribution.

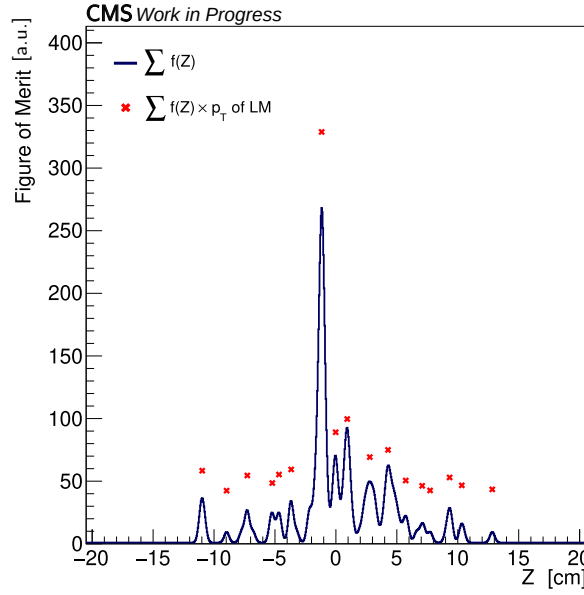


FIGURE 9.3: Image representing two steps of the PFA algorithm in a realistic event.

In Fig. 9.5 we fitted the distribution of all $t\bar{t}$ events. Fixing with $\delta = 0.05$ cm and $\sigma = 0.2$ cm and producing a log-scale plot for the distance $Z_{reco} - Z_{true}$ in all events, we fit the distribution with two different Gaussians, a narrower one that represents the events in which the PV is supposedly “well-reconstructed” and a wider Gaussian which should represent failed reconstruction where Z_{reco} differs significantly from Z_{true} . The examination of the narrow Gaussian fit reveals a standard deviation (σ) of 0.15 cm. We understand that within a range of $Z_{reco} - Z_{true} \sim 0.5$ cm ($\sim$ three standard deviations), we encapsulate all well-reconstructed peaks. Beyond this range, further integration adds lots of vertices reconstructed from pile-up tracks. Another crucial observation is that the signal distribution, represented by the narrow Gaussian bump fitted at the centre of Fig. 9.5, will fluctuate with changes of σ and δ parameters. Indeed, as the algorithm parameters are adjusted, this distribution will either narrow or widen. The primary objective of the algorithm optimization is to minimize this distribution, aiming for the narrowest possible shape. A narrower distribution signifies an improved resolution of the algorithm, hence enhancing its performance.

In conclusion, the two parameters that are interesting to check while modifying the algorithm will be the number of reconstructed vertices within the “ $Z_{reco} - Z_{true} \leq 0.5$ cm” range (or tighter cuts) and the width of the $Z_{reco} - Z_{true}$ distribution.

As previously mentioned, our algorithm comprises two key parameters: δ (the step size) and σ (the Gaussian width). The step size is a relatively straightforward variable to model. We can reduce the step size incrementally to enhance accuracy, but eventually, this improvement reaches a plateau as it becomes constrained by the experimental resolution. Upon varying the step size parameters, we found that below 0.05 cm, there

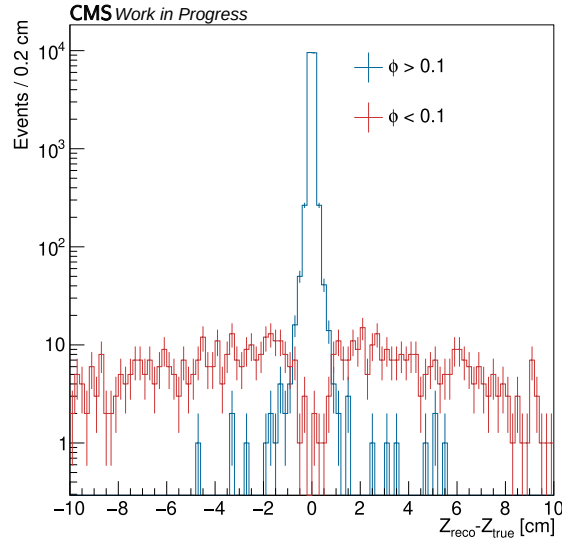


FIGURE 9.4: The light blue distribution with $\phi > 0.1$ represents the events where at least 10% of tracks originate from the true MC vertex. Conversely, the surrounding data depicted in red denotes noise, often stemming from events with significant pileup occurrences ($\phi < 0.1$), indicating that less than 10% of the tracks used for reconstruction are sourced from the true MC vertex.

is no significant gain in resolution (we have reached the limit of our experimental resolution) as visible from Fig. 9.6. At this stage we are not discussing a computational optimization of this algorithm, to be implemented on FPGA (that will be discussed in Section 9.3), so we choose the best limit case.

Let's now explore a significant parameter of the algorithm: the width of the Gaussian. Conducting a first simple study and plotting three σ values in the interval between 0.05 cm and 0.3 cm in Fig. 9.7, we notice that the resolution is slightly better for values around 0.15 cm.

Zooming more around the value of 0.15 cm, we study how the distribution of residuals, $Z_{reco} - Z_{true}$, evolves changing σ , while keeping δ fixed at 0.05 cm. This is shown in Fig. 9.8 where the colour axis shows the fraction of reconstructed events using a certain σ and within a certain $Z_{reco} - Z_{true}$ range.

The algorithm's efficiency is defined by the ratio between the number of events with the PV reconstructed within a certain $Z_{reco} - Z_{true}$ range and the total number of events. In Fig. 9.8, we extract two key insights. Firstly, by examining the "efficiency value" (represented by the colour scale) at $Z_{reco} - Z_{true} \approx 0.5$ cm, we can gauge the algorithm's efficiency. Secondly, as we move below the value of 0.5 cm, we observe how the resolution varies changing sigma. For instance, going down the range from 0.5 to 0.1 cm, maintaining the same sigma, a slower decline in efficiency signifies better resolution for that specific σ value.

In Fig. 9.9 (A), we observe the algorithm's efficiency with varying σ , revealing a peak

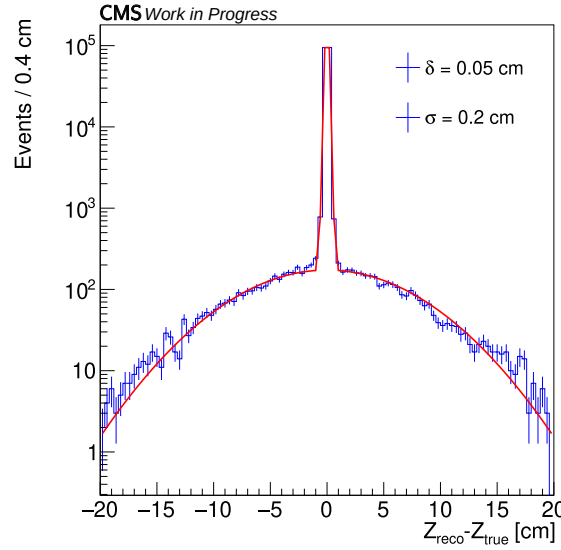


FIGURE 9.5: Log-scale plot illustrating the distribution of $Z_{\text{reco}} - Z_{\text{true}}$ for events with fixed $\delta = 0.05$ cm and $\sigma = 0.2$ cm. The distribution is characterized by two distinct Gaussian fits. A narrower distribution corresponds to events with a successful reconstruction, juxtaposed with a broader distribution encompassing events falling below, reflecting varying degrees of reconstruction quality.

in the curve at 0.22 cm. This plot is generated while fixing $Z_{\text{reco}} - Z_{\text{true}}$ at 0.5 cm. The efficiency represents the fraction between the number of events where the hard vertices are reconstructed at most 0.5 cm from the true MC primary vertex and the total number of events. The plot on the right of Fig.9.9 (B) is instead produced fixing σ at 0.22 cm and showing the different efficiency for different $Z_{\text{reco}} - Z_{\text{true}}$ ranges. For values bigger than 0.5 cm the curve almost reaches a plateau; there are still some efficiency improvements but as explained before they come from vertices with low-quality reconstruction, primarily comprising tracks originating from pileup.

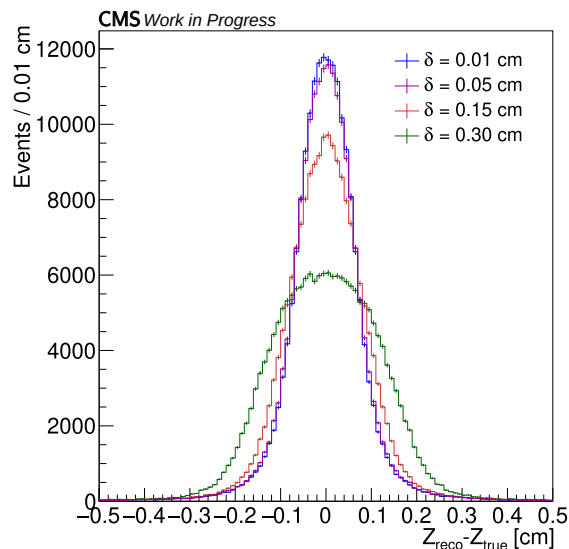


FIGURE 9.6: The impact of varying δ is explored while maintaining σ at 0.15 cm. The algorithm's performance is inherently tied to the value of δ , serving as a limitation on achievable resolution. Decreasing δ initially improves resolution until reaching the experimental resolution limit. This plot shows that beyond $\delta = 0.05$, the gains in resolution become insignificant.

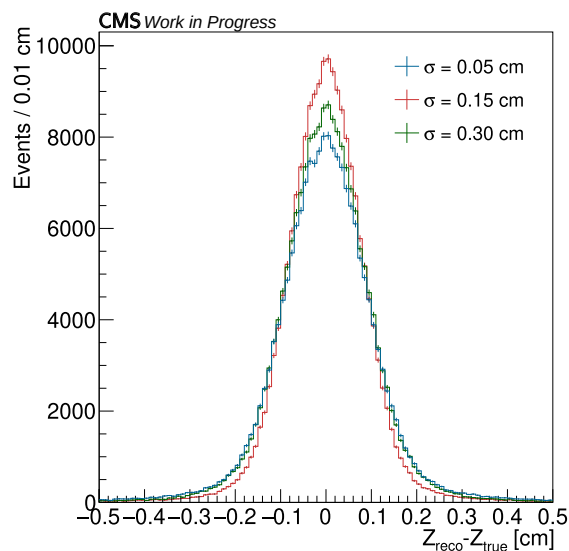


FIGURE 9.7: Exploring the impact of varying σ while setting the step size to 0.05 cm. The plot depicts the behaviour of the algorithm with σ assuming three different values. Interestingly, the analysis suggests an optimal range for σ between 0.05 cm and 0.30 cm, with 0.15 cm the best performance among the three values.

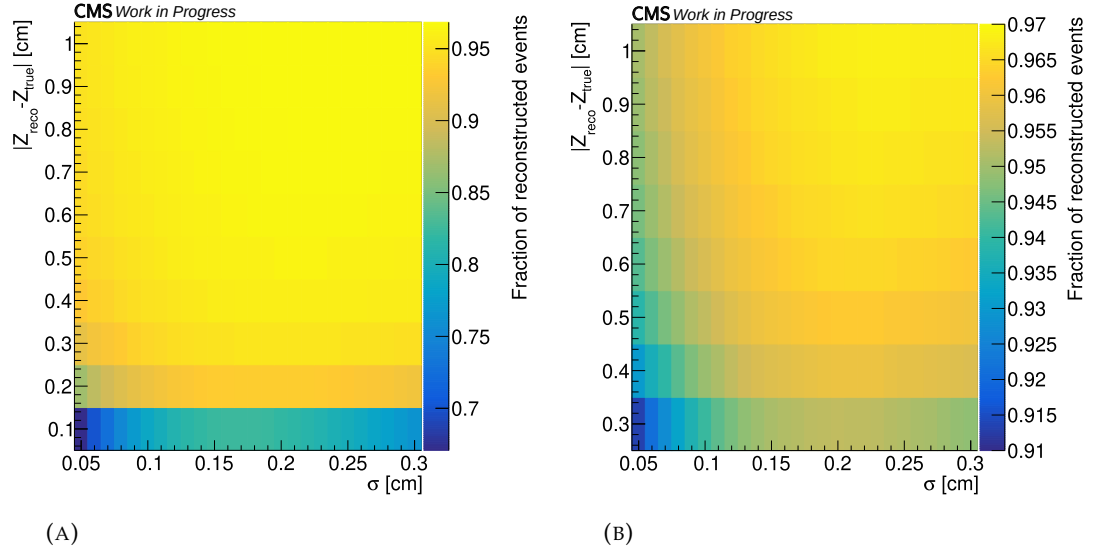


FIGURE 9.8: Bi-dimensional scan of the algorithm across various σ values and $Z_{\text{reco}} - Z_{\text{true}}$ ranges at fixed $\delta = 0.05$ cm. The colour scale illustrates the fraction of reconstructed events within different ranges. The plot on the right (B) is a zoom of the one on the left (A), to show more granularity between $Z_{\text{reco}} - Z_{\text{true}} \leq 0.3$ and $Z_{\text{reco}} - Z_{\text{true}} \leq 1$.

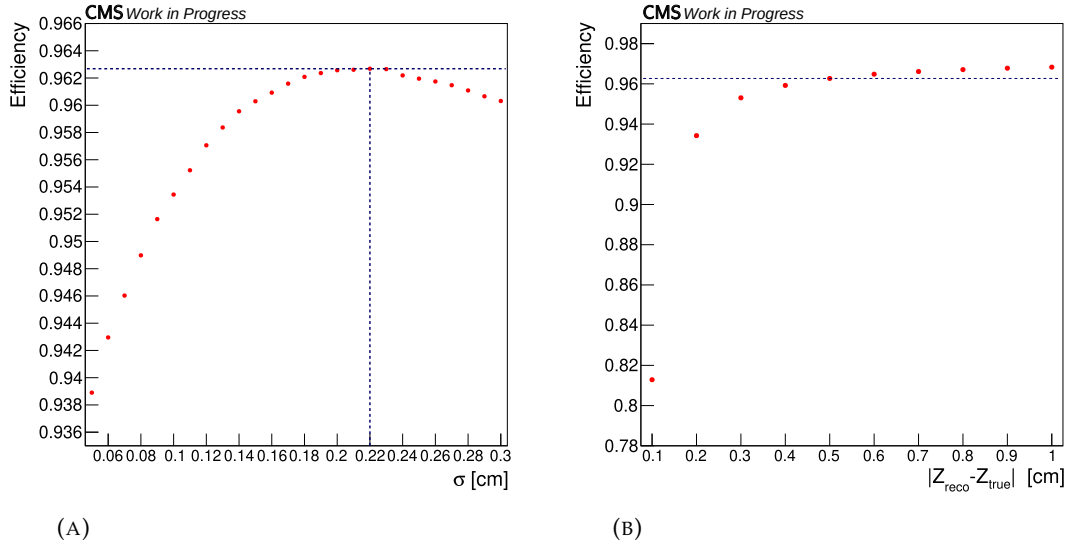


FIGURE 9.9: In Figure (A), the algorithm's efficiency varies with σ , peaking at 0.22 cm of 0.96. The plot is generated with $Z_{\text{reco}} - Z_{\text{true}}$ fixed at 0.5 cm. The plot on the right side (B) fixes σ at 0.22 cm, displaying varying efficiency for different $Z_{\text{reco}} - Z_{\text{true}}$ ranges. Beyond 0.5 cm, the curve nearly plateaus, indicating marginal efficiency improvements attributed to the addition of low-quality reconstructed vertices with numerous tracks from the pileup.

9.1.2 Adding the η information

So far the Peak Finding Algorithm works only by taking into account the Z values of the tracks and their p_T . Other variables could help the algorithm's reconstruction. Studying the resolution of the linear impact parameter z_0 , we noticed that, it maintains a uniform distribution across the p_T values, as shown in Fig. 9.10.

On the other hand, studying the z_0 resolution varying $|\eta|$, a clear dependence between the two variables is visible, as shown in Fig. 9.11.

The η variable is a spatial coordinate that describes the angle of a particle relative to the beam axis. It is defined by the Eq. 2.2. The η information provides the direction of the track inside the detector. If the track is pointing toward the barrel it will have an $|\eta|$ value < 1.5 . On the other hand, if it is pointing in the direction of the endcap it will have a $|\eta| > 1.5$. The plot in Fig. 9.11 is based on a $t\bar{t} + \text{PU} = 200$ MC sample and shows with two different curves the behaviour of the z_0 resolution as a function of η .

The curves show the value of track resolution at different particle $|\eta|$. The black (red) curve shows the range around the true z_0 where we can find 68% (90%) of the tracks with $p_T > 2$ GeV. The resolution degrades as the $|\eta|$ value changes, this means that tracks pointing toward the endcaps have a worse resolution compared to the ones pointing towards the barrel.

It is important to notice that the p_T and the η of a track are very often strictly correlated, meaning that tracks with low p_T are often the ones going toward the endcap direction, while the ones with high p_T are usually pointing towards the barrel.

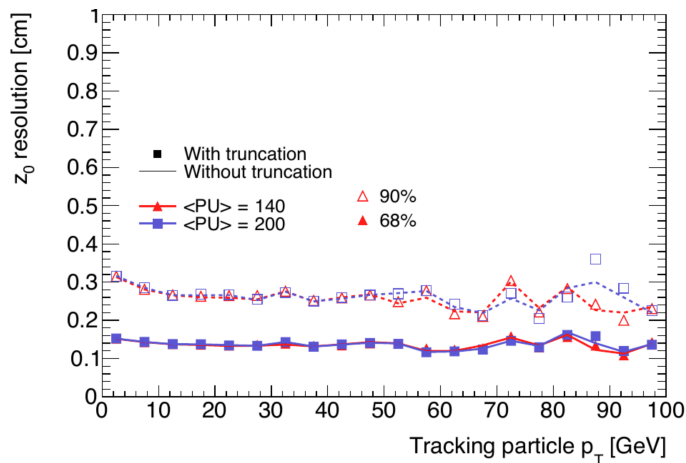


FIGURE 9.10: The Z_0 track parameter resolutions for muons in $t\bar{t} + \text{pileup}$ events as a function of the muon p_T [8].

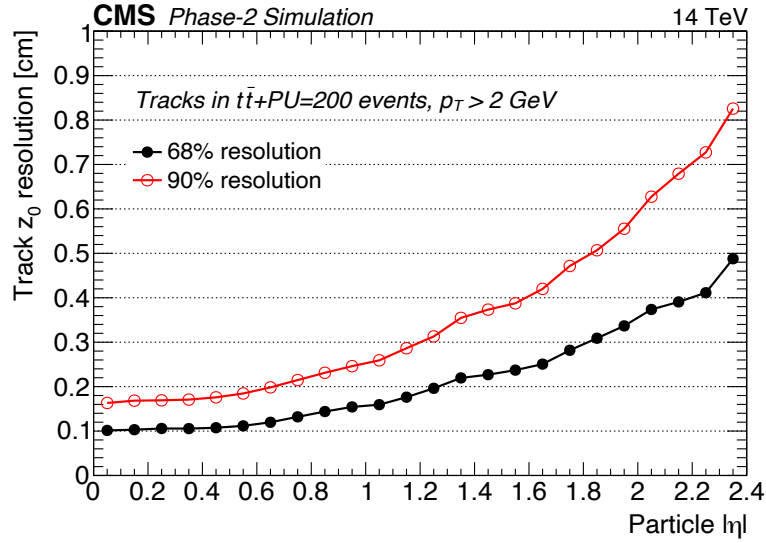


FIGURE 9.11: The z_0 resolution for tracks in $t\bar{t}$ events, with an average pileup of 200, is depicted as a function of $|\eta|$. The z_0 resolutions are represented by intervals encompassing 68% (indicated by filled black markers) or 90% (represented by open red markers) of all tracks with $p_T > 2 \text{ GeV}$ [16].

In order to include the dependency between the track z_0 resolution and $|\eta|$ in the PFA the following approach was used. A parabolic fit for the 68% or the 90% resolution curve in Fig. 9.11 was performed. By imposing the track z_0 resolution, to be equivalent to the width of a Gaussian distribution σ , we derive the following parabolic fit equation for the curve at 68%:

$$\sigma = c_0 + c_1 \cdot \eta + c_2 \cdot \eta^2 \quad (9.5)$$

This simplified parabolic fit provides a practical representation of the track z_0 resolution. The fitted curves for both 68% and 90% can be shown in Fig. 9.12.

The algorithm modification, that incorporates the η value, involves the adaptation of σ for each track (the step size δ remains fixed as before). This adjustment aims to tailor the σ to the characteristics of each specific track. The algorithm now takes into consideration the track's η value, dynamically adjusting the σ with each iteration to best accommodate the particularities of the given track.

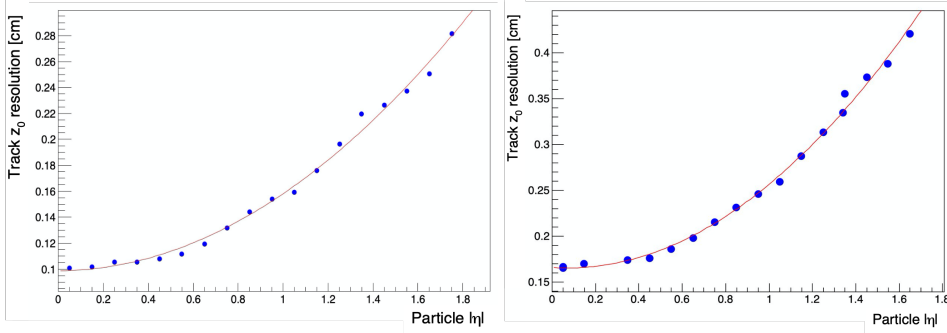


FIGURE 9.12: The figures illustrate a simplified parabolic fit representing the track Z_0 resolution of the 68% resolution curve on the left and of the 90% resolution curve on the right.

Figure 9.13 shows the PFA algorithm efficiency across three distinct scenarios. The original algorithm, depicted in red, is contrasted with iterations incorporating the η variable following the 68% resolution curve in blue, and 90% in green. The implementation of η improves the resolution of the algorithm especially with a low value of $Z_{reco} - Z_{true}$.

Figure 9.14 presents two plots: one on the left labelled (A) and another on the right labelled (B). These plots depict the $t\bar{t}$ residual in three distinct scenarios. In the first scenario (depicted in red), the plot represents the $t\bar{t}$ residual when using the PFA without considering the η contribution. In the second scenario (shown in blue), the plot illustrates the $t\bar{t}$ residual with the η contribution incorporated, following the curve at 68% resolution. Lastly, in the third scenario (displayed in green), the plot demonstrates the $t\bar{t}$ residual with the η contribution included, following the curve at 90% resolution. The plot on the right (B) is a zoomed-in view of the plot on the left (A) within the range of interest.

Plots in Fig. 9.15- 9.19 show the algorithm residual plot on the left (A) and efficiency plot on the right (B) for QCD samples with different energy ranges. In each plot, a comparison is drawn between the original PFA and its adaptation utilizing the η distribution at a 68% resolution.

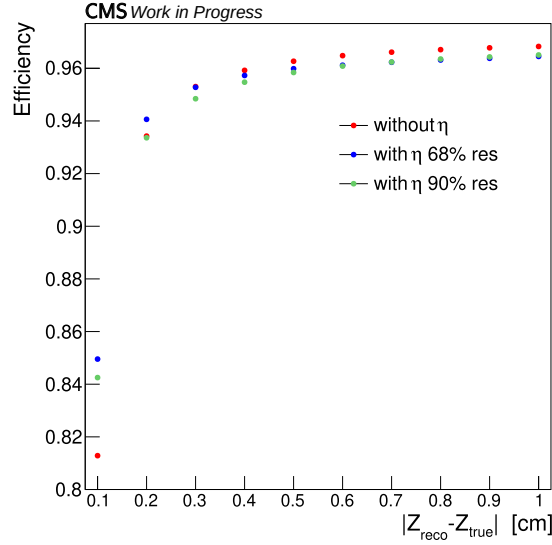


FIGURE 9.13: Algorithm's efficiency in three different cases for $t\bar{t}$ sample. The old algorithm in red is compared with the one with the η variable implementation following the 68% resolution in blue and the 90% resolution in green.

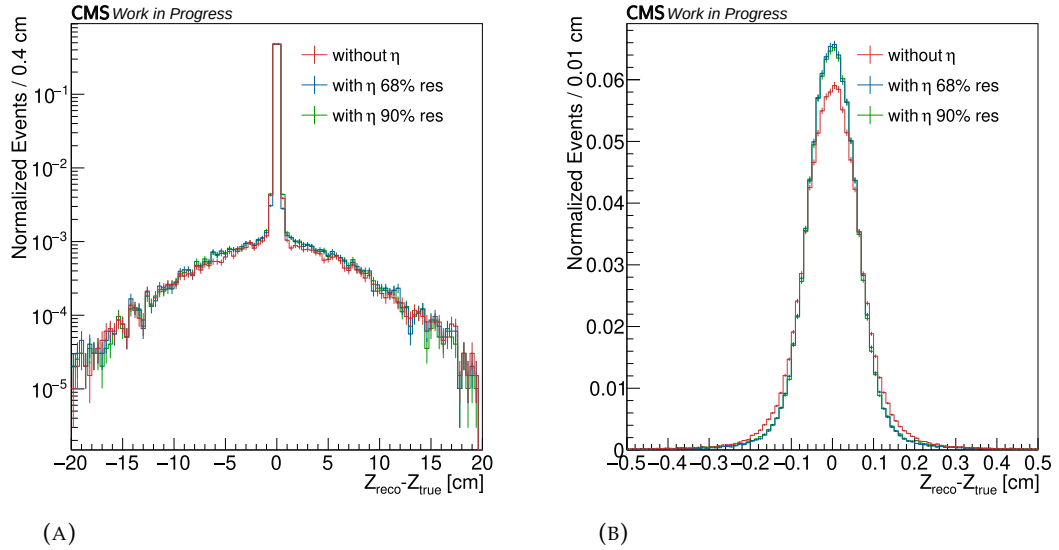


FIGURE 9.14: The plot on the left (A) and the plot on the right-hand side (B) represent $t\bar{t}$ residual in the three different cases: PFA without η contribution (in red), η contribution following the curve at 68% resolution (blue) and 90% resolution (green). The left and the right plot are based on the same data. The left plot uses a log scale for the y-axis with an x-axis range of -20 to 20 cm. The right plot uses a linear y-axis scale with an x-axis range of -0.5 to 0.5 cm.

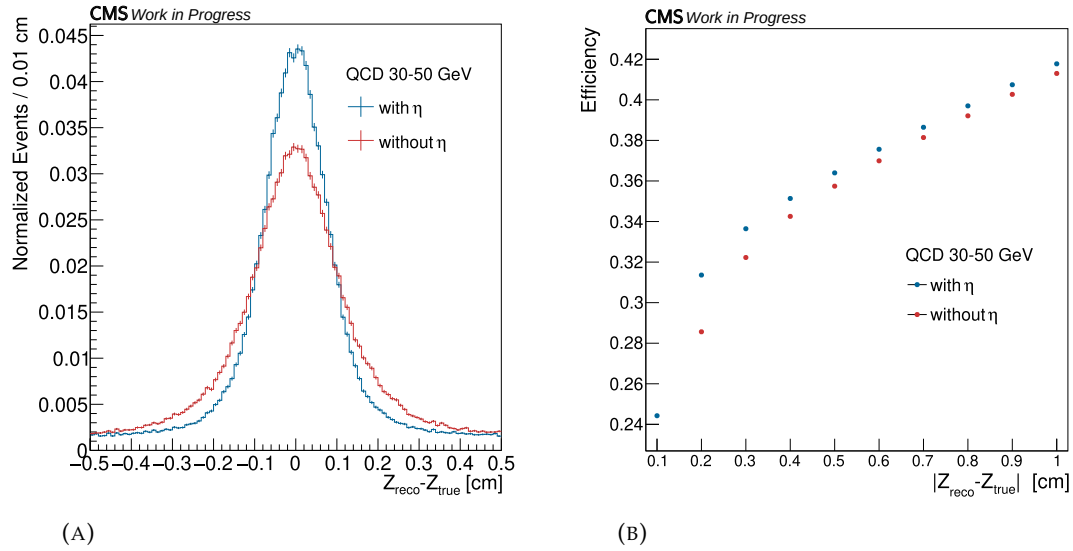


FIGURE 9.15: Residual plot (A) and efficiency plot (B) of the algorithm, using the QCD 30-50 GeV sample. A comparison between the original PFA and the one using the η distribution at 68% resolution is shown.

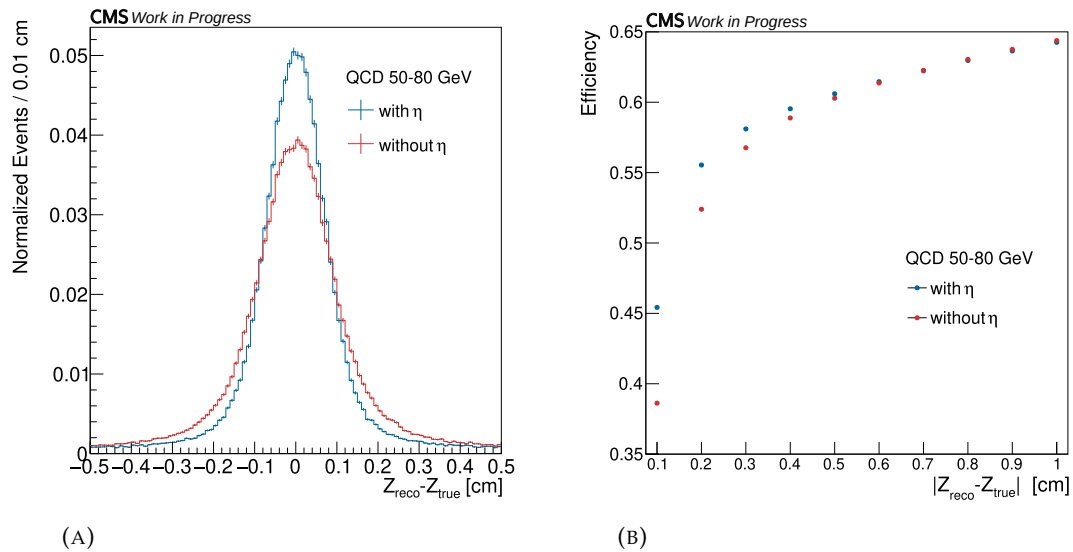


FIGURE 9.16: Residual plot (A) and efficiency plot (B) of the algorithm, using the QCD 50-80 GeV sample. A comparison between the original PFA and the one using the η distribution at 68% resolution is shown.

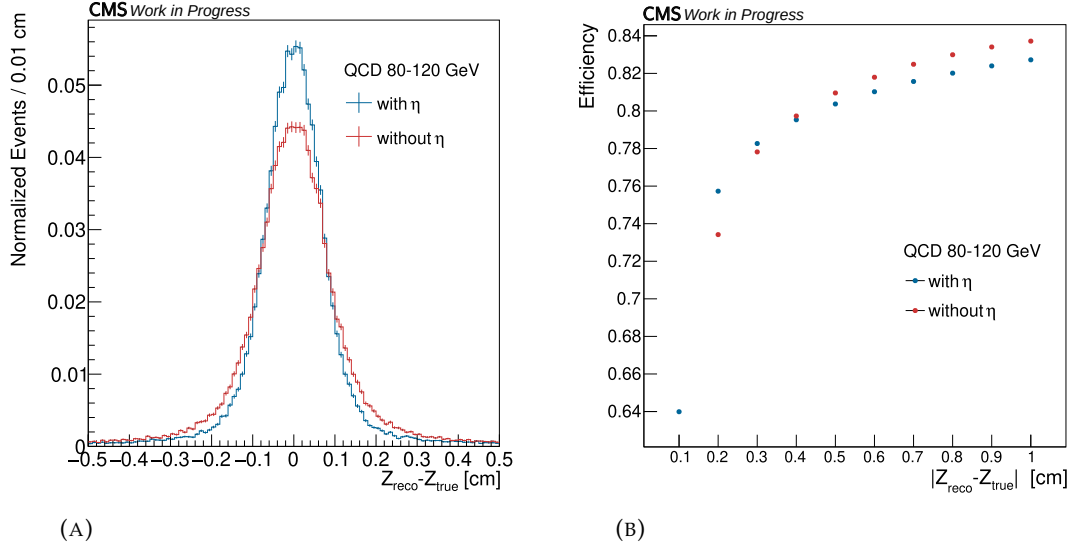


FIGURE 9.17: Residual plot (A) and efficiency plot (B) of the algorithm, using the QCD 80-120 GeV sample. A comparison between the original PFA and the one using the η distribution at 68% resolution is shown.

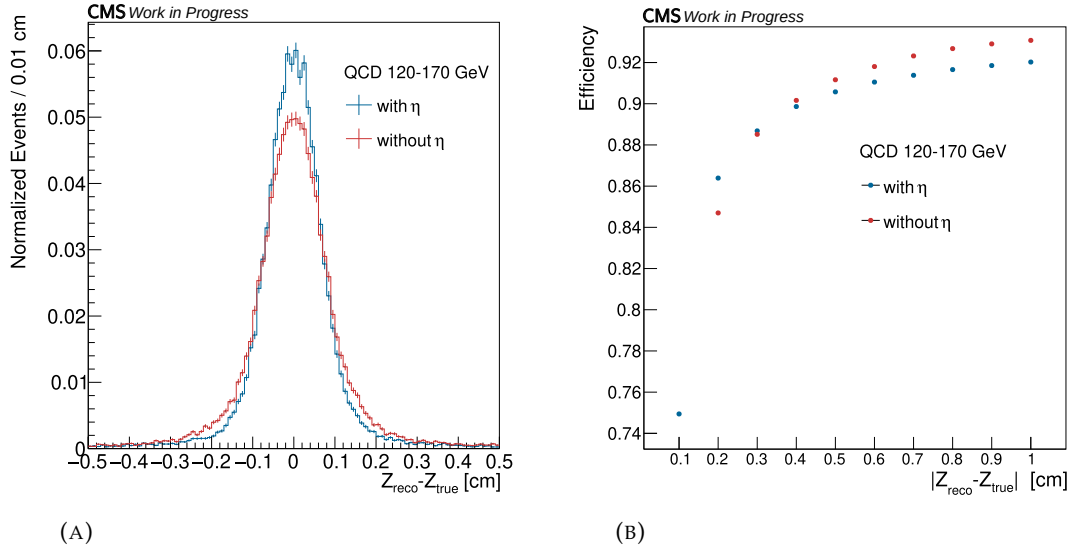


FIGURE 9.18: Residual plot (A) and efficiency plot (B) of the algorithm, using the QCD 120-170 GeV sample. A comparison between the original PFA and the one using the η distribution at 68% resolution is shown.

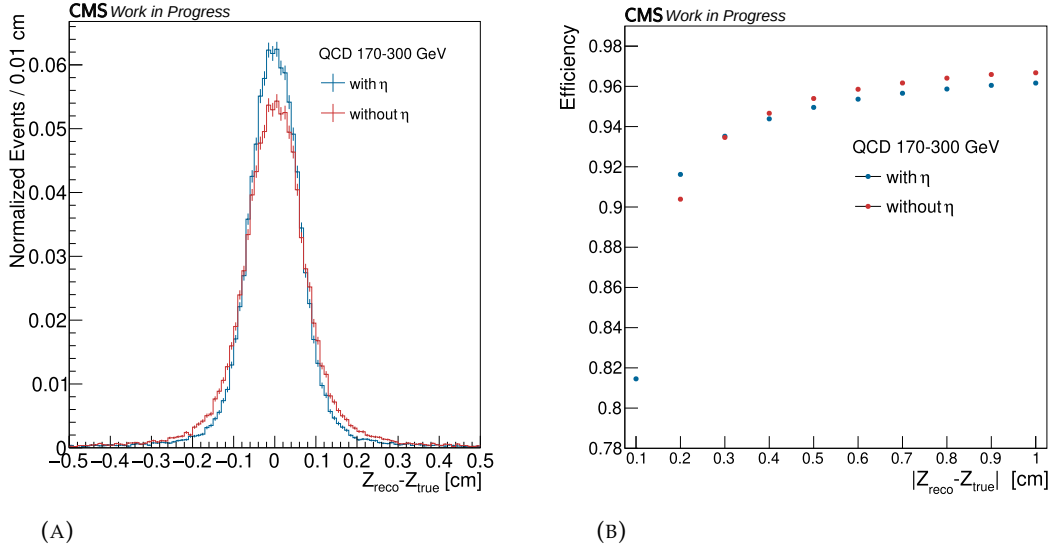


FIGURE 9.19: Residual plot (A) and efficiency plot (B) of the algorithm, using the QCD 170-300 GeV sample. A comparison between the original PFA and the one using the η distribution at 68% resolution is shown.

From these plots, we can state that the resolution of the reconstruction always improves with the addition of the η contribution. For small values of $Z_{\text{reco}} - Z_{\text{true}}$, there is consistently an enhancement. For instance, in $t\bar{t}$ events, we reconstruct approximately 85% of the events within 1 mm from the true MC vertex. Conversely, without considering η , we reconstruct only around 81% of the events within this range.

In low energy samples like QCD events with 30-50 GeV or 50-80 GeV, the enhancement remains relatively consistent across $Z_{\text{reco}} - Z_{\text{true}}$ ranges. While in more energetic samples at some point in the efficiency curve, there is a swap between the two configurations.

This can be due to the presence of more tracks pointing toward the endcaps in low energy samples compared with the high p_T sample. While the low-energy samples always benefit from the inclusion of low-resolution tracks pointing toward the endcap, the samples that have already a high reconstruction efficiency, such as $t\bar{t}$, may be disadvantaged by this process above a certain $Z_{\text{reco}} - Z_{\text{true}}$ threshold. It is important to notice anyway that the $Z_{\text{reco}} - Z_{\text{true}}$ range we are mostly interested in, is the one within 0.5 cm.

To better understand the different η distribution among the different samples, in Fig. 9.20 and 9.21 the jet axis orientation in η is presented. The jets plotted in figures are only the ones associated with the primary vertex. As previously mentioned η is strictly correlated to the p_T of the tracks, and the samples with lower p_T values have more tracks ending up in the endcaps. This is shown also in the jet η plots, where lower energy samples show more jets pointing toward the endcap, compared with higher energy

samples. Figure 9.22 shows the distribution of the resolution of the algorithm (with η implementation) for different samples.

Another factor contributing to the significant disparity in efficiency between high and low p_T samples stems from limitations within the algorithm. Specifically, the algorithm encounters challenges in distinguishing QCD events from pileup (PU) events, particularly when QCD events are generated within a similar p_T range as PU tracks and do not exhibit a significantly higher multiplicity compared to pileup events.

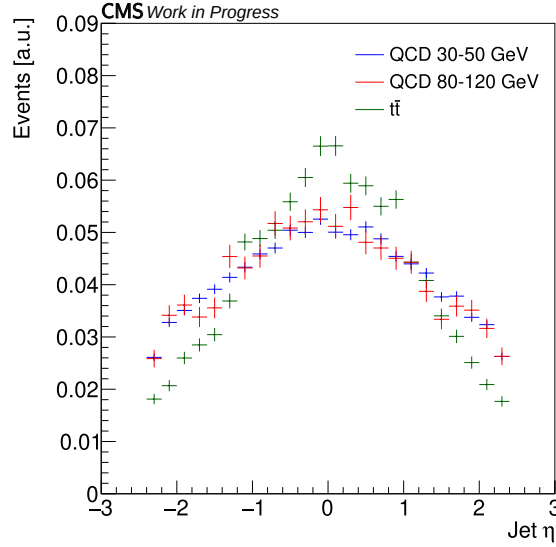


FIGURE 9.20: Normalized representation of the η of the Jets associated with the PV for three different samples: QCD 30-50 GeV, QCD 80-120 GeV and $t\bar{t}$ sample.

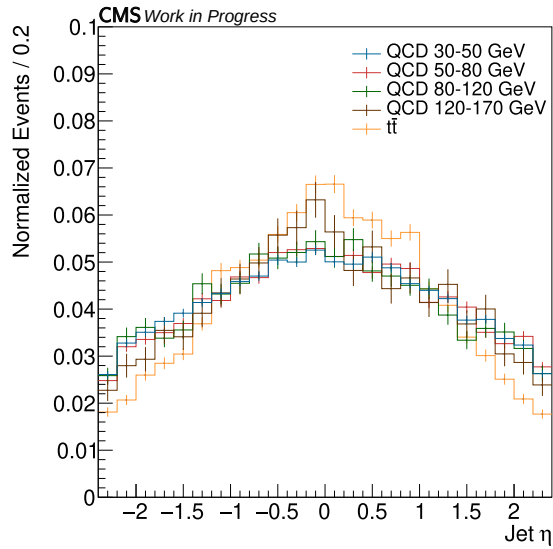


FIGURE 9.21: Normalized representation of the η of the Jets associated with the PV for different samples: QCD 30-50 GeV, QCD 50-80 GeV, QCD 80-120 GeV, QCD 120-170 GeV and $t\bar{t}$ sample.

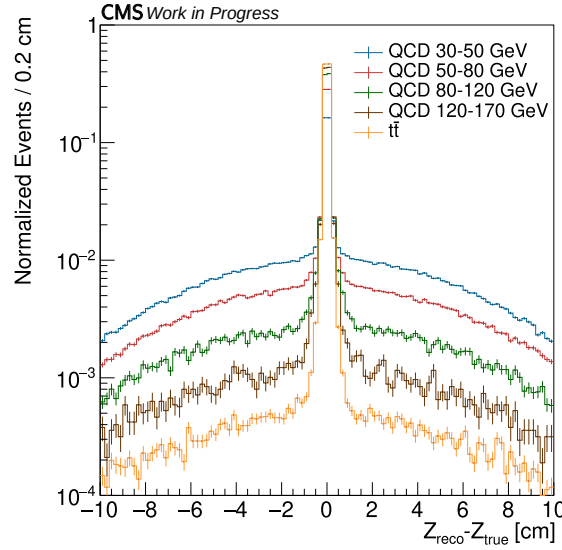


FIGURE 9.22: Residuals, showing the resolution of the algorithm with η for different samples: QCD 30-50 GeV, QCD 50-80 GeV, QCD 80-120 GeV, QCD 120-170 GeV and $t\bar{t}$ sample.

9.1.3 Exploring supplementary discrimination factors

In this section, we explore additional variables with the aim of enhancing the performance of the Peak Finding Algorithm. While some variables yield interesting outcomes, others do not. This section provides a comprehensive report of this study's results.

In the PFA, each peak can be considered as an interaction vertex candidate. Vertex candidates are then ranked based on the strategy outlined in section 9.1 in order to find the primary vertex of interest. Hence, together with the primary vertex candidate, i.e. the first vertex in the ranking, we also have information on the other reconstructed vertices, which can be exploited to enhance the algorithm's performance. So far the algorithm efficiency is computed using only the first peak and a primary vertex candidate is considered well reconstructed if the first peak is within a certain threshold distance from the true hard vertex MC position. For example, in the $t\bar{t}$ sample the fraction of badly reconstructed hard vertices is about 4% when a threshold distance of $|Z_{\text{reco}} - Z_{\text{true}}| \leq 0.5$ cm is chosen. The efficiency can be determined by employing both the first and second peaks in the ranking. In this case, the reconstruction is considered correct if at least one of the two selected candidates is within the threshold distance from the true hard vertex MC position. In the $t\bar{t}$ sample the number of badly reconstructed hard vertex can be halved to 2%. In order to actually obtain this performance enhancement it is necessary to swap the first and the second peak in the ranking when the latter is more accurate. Of course, this cannot always be known a priori, but it might be possible to

identify a region in the space of a few variables where this exchange is almost always advantageous. This is the purpose of this section.

In the following, a certain track is associated with a reconstructed vertex (and thus the track is considered to contribute to the reconstruction of that vertex) if the distance between track Z and the position of the reconstructed vertex is within 2σ , i.e. $|Z_i - Z_{reco}| \leq 2\sigma_i$, where σ is defined as in Eq. 9.5.

In Fig. 9.23 the sum of the transverse momentum ($\sum p_T$) of tracks contribution to the reconstruction of the first and the second peak and their average p_T is depicted. On the left-hand side, Fig. (A) and (C) represent the case in which the first peak is closer to the true MC PV than the second peak. Fig. (B) and (D) describe the opposite scenario: in this case, the second reconstructed peak is closer to the true MC PV and it would be advantageous to select this value instead.

In this figure, plots (A) and (B) depicting the sum of transverse momenta ($\sum p_T$) do not reveal clear exploitable features. The region of interest shown in Fig. (B), where the second peak outperforms the first, overlaps with that of Fig. (A). Therefore, swapping depending on this variable would instead decrease the efficiency. Regarding Fig. (C) and (D), however, the average (p_T) reveals an intriguing pattern: above 40 GeV, Fig. 9.23 (D) exhibits a higher number of events compared to Fig. (C). This observation suggests the potential for improvement by selecting the second reconstructed vertex.

The same plots are reported in Fig. 9.24 for two different variables such as the multiplicity (number of tracks involved in the reconstruction) and the standard deviation., defined as $\tilde{\sigma} = \sqrt{\frac{1}{N} \sum_{i=1}^N (Z^i - \bar{Z})^2}$, where N is the tracks multiplicity, Z^i represents the Z for each individual track and $\bar{Z}$ the mean of Z values for that specific peak. In terms of multiplicity, within the range where multiplicity is less than 5, Fig. (B) contains considerably more events than Fig. (A). This observation suggests that swapping the two peaks in this region could enhance the efficiency. Similarly, when examining the plots of standard deviation, a similar effect is observed when looking at events where the standard deviation of the first peak is around 0.

The insight that can be extracted from the features observed in Fig. 9.23 (D) and Fig. 9.24 (B) and (D) have a common explanation traceable to multiplicity. The average multiplicity of pileup vertices is sensibly lower than the typical multiplicity of $t\bar{t}$ vertices. Even if in a $t\bar{t}$ event, high p_T tracks are produced, the high multiplicity decreases the average p_T per track contributing to the vertex reconstruction. Hence very high average p_T is here associated with pileup vertex with low multiplicity. Moreover, events in Fig. 9.24 (D) with a standard deviation around 0 correspond to those with a multiplicity of few units.

Another investigation was conducted concurrently with the examination of the best peak. The best peak refers to the best peak that could possibly be reconstructed by the PFA, with its Z value distant from the Z_{true} MC less than half the step size. In Fig. 9.25

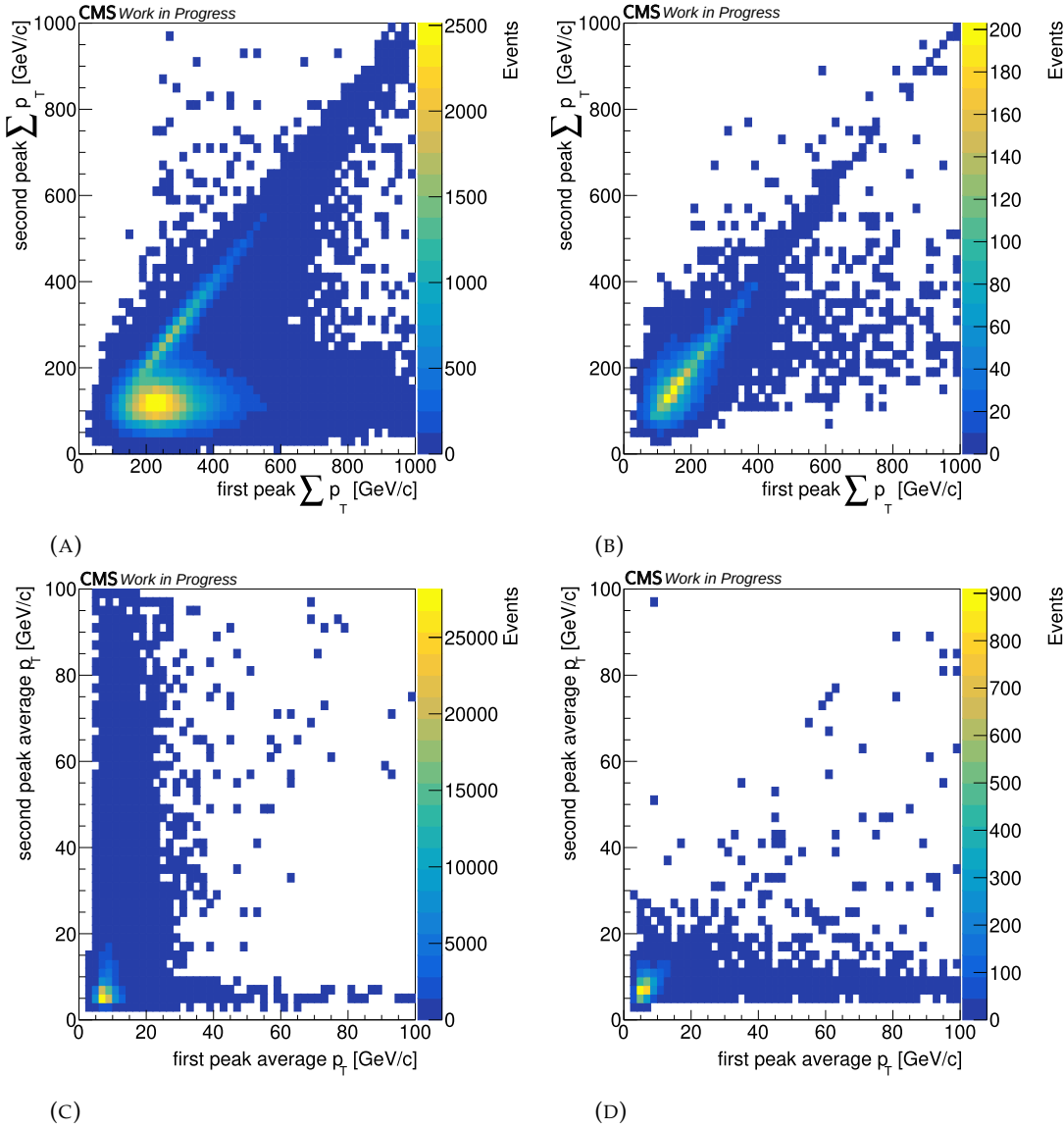


FIGURE 9.23: In Fig. (A) and (B) the sum of the transverse momentum ($\sum p_T$) of tracks reconstructing the second peak vs the first peak is shown. Figure (A) represents the case in which the first peak is closer to the true MC PV than the second peak. While Fig. (B) describes the opposite scenario. The same concept is depicted in Figures (C) and (D) for the average p_T quantity.

the $\sum(p_T)$, average (p_T), multiplicity and standard deviation are shown for the best peak vs the first one. These plots are a confirmation of findings from the initial investigation between the first and the second peaks. Specifically, concerning the average transverse momentum (p_T), a tail is observed for values exceeding 40 GeV, indicating that the best peak exhibits a lower average p_T in that zone. At low multiplicities, the first peak is more commonly badly reconstructed. The algorithm would benefit from selecting a peak with higher multiplicity instead. A pattern similarly is noted in the standard deviation plots where values around 0 indicate inadequate reconstruction.

Another investigation concerning the first and second peaks is depicted in Fig. 9.26.

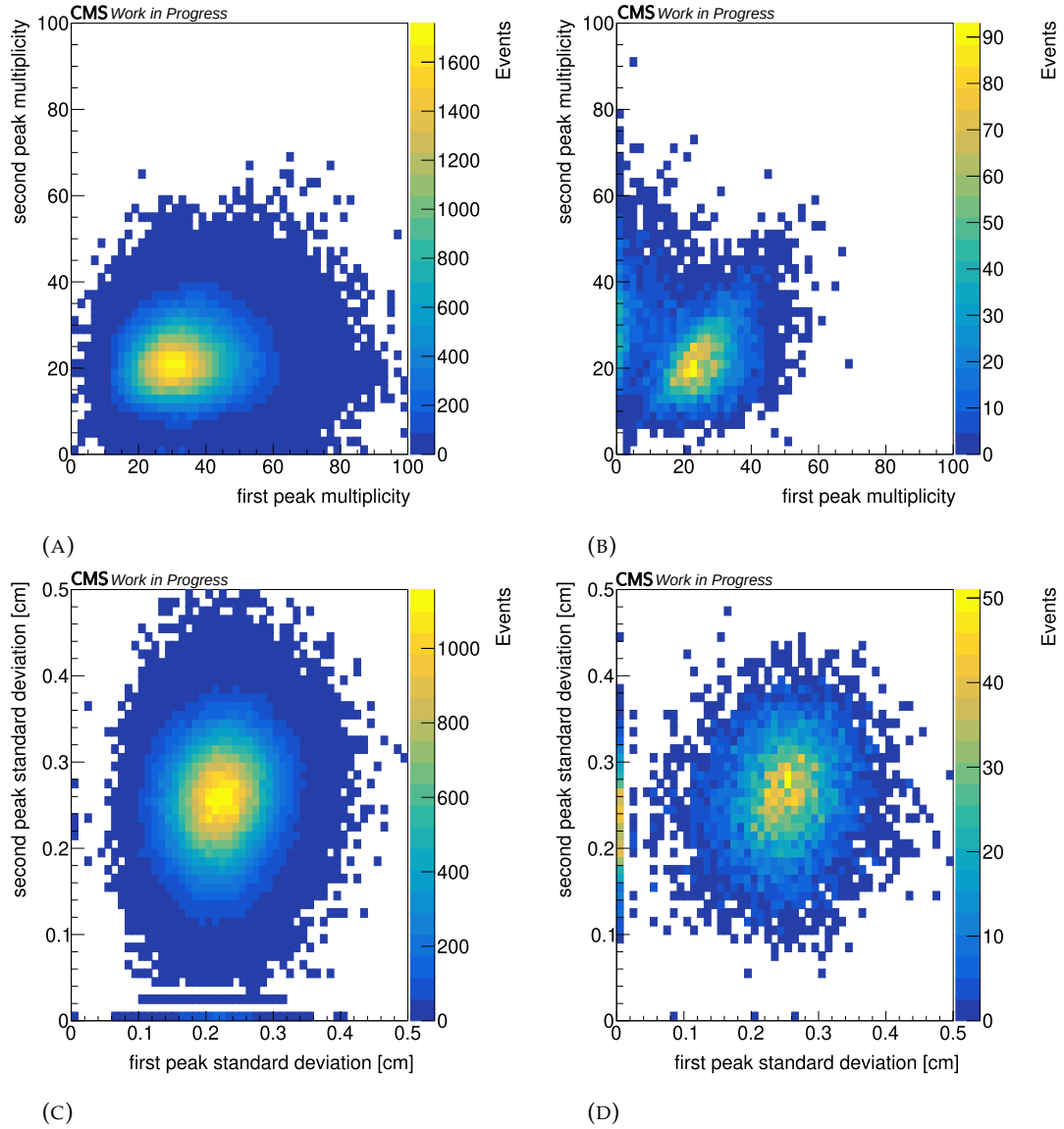


FIGURE 9.24: In Figs. (A) and (B) the multiplicity of tracks reconstructing the second peak vs the first peak is shown. Figure (A) represents the case in which the first peak is closer to the true MC PV than the second peak. While Fig. (B) describes the opposite scenario. The same concept is depicted in Figs. (C) and (D) for the average standard deviation quantity.

The blue curve illustrates scenarios where the first peak outperforms the second peak, while the red curve represents instances where the second peak surpasses the first. More precisely, the blue curve displays the ratio between the quantity taken into consideration (for instance $\sum p_T$) associated with the first peak and the same quantity for the second peak in cases where swapping peaks is undesired (i.e. when the first peak is closer to the true MC PV). Conversely, the red curve portrays the same ratio but in the opposite scenario, it is desired to swap in this case. The quantities under scrutiny include the sum of transverse momenta (p_T), average p_T , standard deviation, and multiplicity.

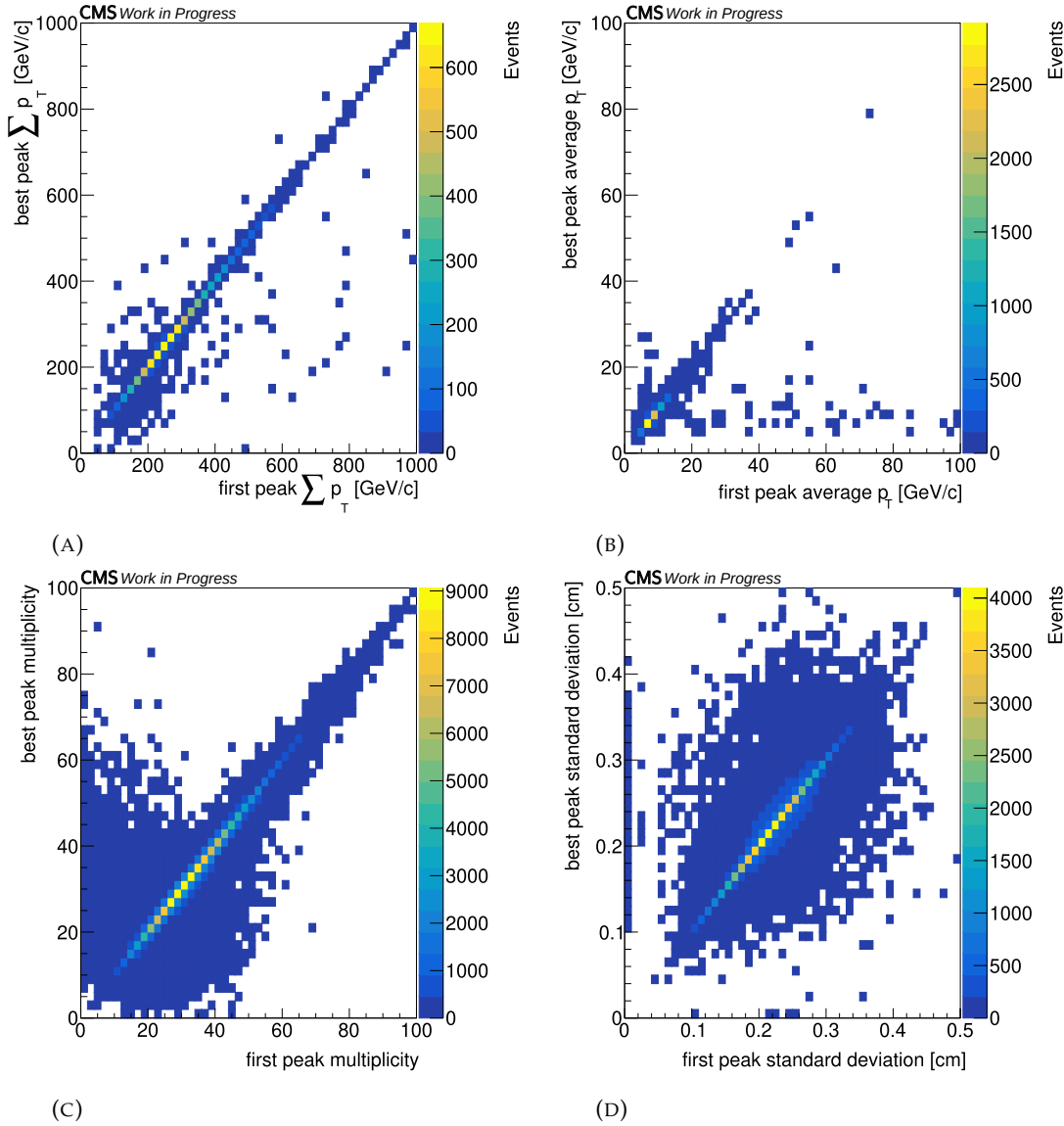


FIGURE 9.25: The sum of transverse momenta ($\sum p_T$), average transverse momentum (p_T), multiplicity, and standard deviation are depicted for the best reconstructable peak compared to the first one.

These plots offer insights into relationships that can be represented visually in a 2D plot as a line $y = mx$, where y represents the value of the observable for the second peak, x represents that one of the first peak, and m denotes the slope. Analyzing the multiplicity plot, we observe that when the ratio $x/y < 0.3$, it is advantageous to interchange them. This means that above the line $y = (1/0.3)x$ in the 2D plot, there is a benefit in swapping the two variables.

Following the implementation of the described cuts and swaps, the algorithm's efficiency shows an improvement of approximately 0.7% as visible in Fig. 9.27.

The same features and similar performance improvements shown here, in the example case of the $t\bar{t}$ sample, are also observed in QCD samples, both at low and high energies,

as shown in Appendix B. This demonstrates that there could be an improvement in applying these cuts. However, for low multiplicity processes, it may not be an advantage. Checking the effect on these samples is left for future study. The result of this study is therefore not implemented in the FPGA implementation (described in 9.3.3).

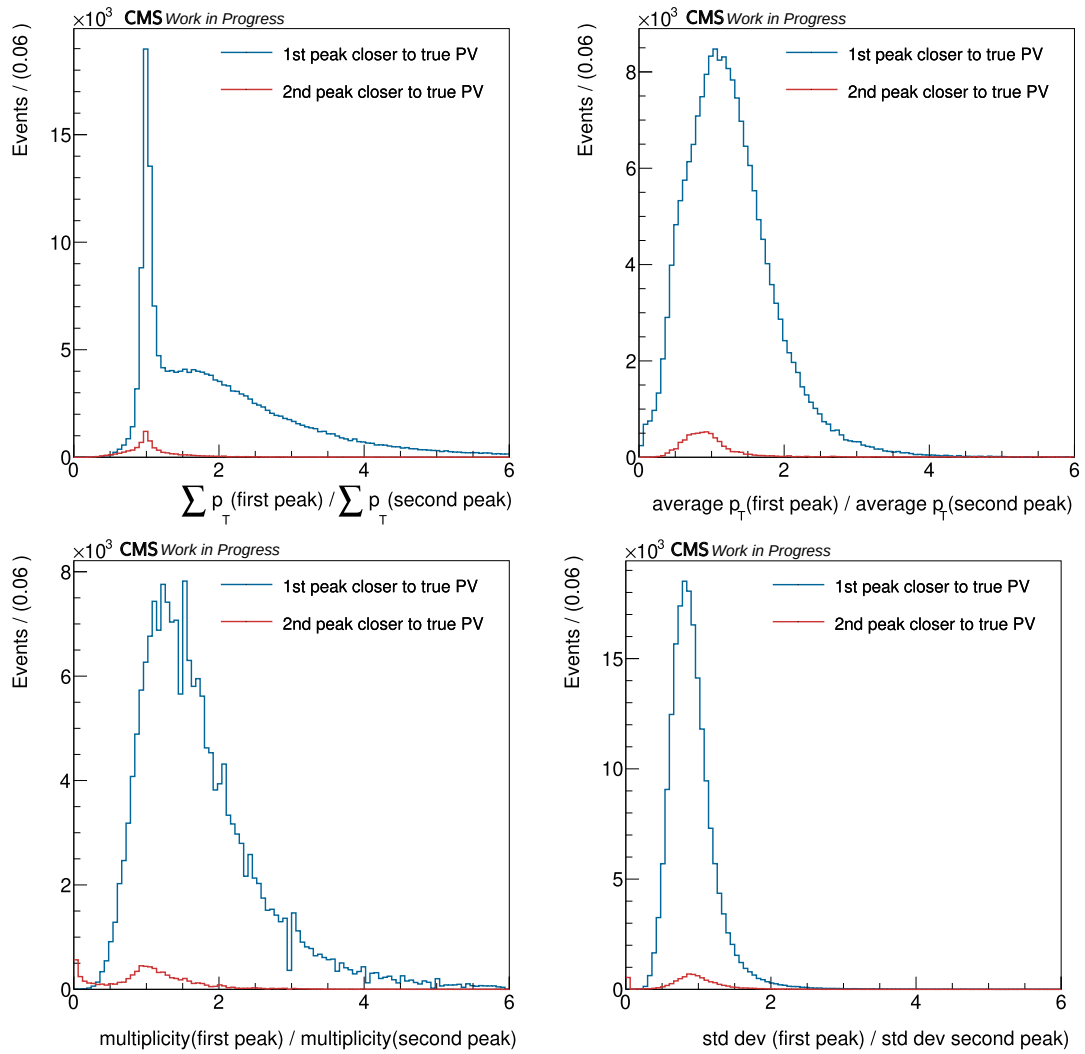


FIGURE 9.26: The four plots represent the ratio between four different quantities (sum of transverse momenta (p_T), average p_T , standard deviation, and multiplicity) calculated for the first and the second peak. The blue curve showcases situations where the first peak is closer to the MC PV than the second peak, while the red curve illustrates cases where the second peak is closer than the first to the true MC primary vertex.

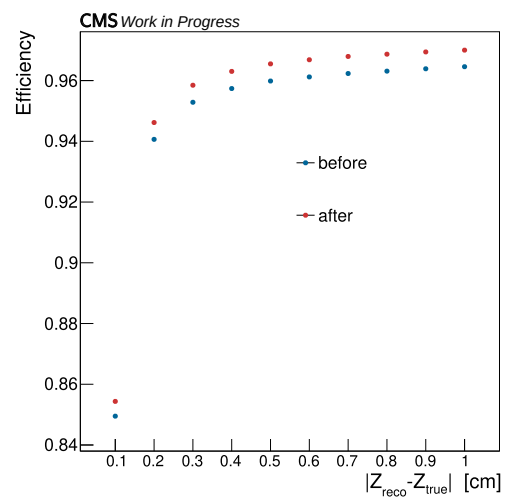


FIGURE 9.27: After implementing the described cuts and swaps, the algorithm's efficiency demonstrates an enhancement of approximately 0.7%.

9.2 Comparison between the PFA and FastHisto

This section provides a brief description of the FastHisto algorithm, followed by a comparison between the PFA and the FastHisto algorithm and an evaluation of the performance differences between the two.

The CMS L1 Phase-2 upgrade baseline approach for vertex finding is known as FastHisto [16]. Initially, the FastHisto algorithm receives tracks detected by the track finding system. Once the tracks have been received, they are sorted into bins based on their positions along the z -axis. Each track's contribution is weighted by its transverse momentum p_T . A three-bin window slides across the entire histogram summing all the track p_T in that window. The PV position is determined by computing the p_T -weighted average of the Z positions of the three bins with the maximum scalar sum of p_T .

The baseline method employs a 256-bin histogram with a bin width of 0.16 cm, covering an interval in Z of -20.47 to 20.47 cm. An example of one event is shown in Fig. 9.28.

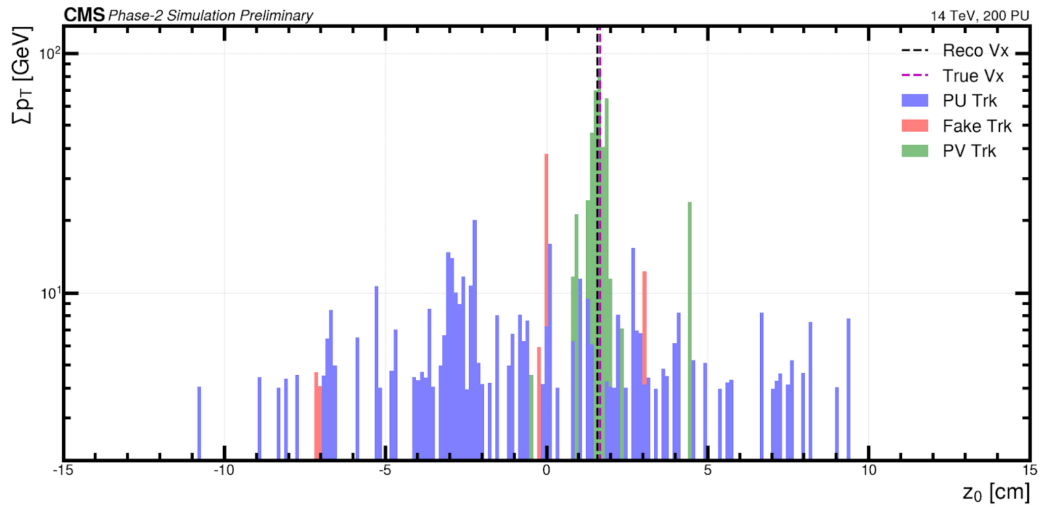


FIGURE 9.28: The FastHisto histogram is shown for a single $t\bar{t}$ event with 200 PU, illustrating the Z distribution in 256 bins, weighted by p_T for each track. Different track types are distinguished by colours: blue represents evenly spread PU tracks, red denotes randomly distributed fake tracks, and green signifies primary vertex tracks clustered around $z_0 = 2$ cm, corresponding to the true vertex location (indicated by the purple dashed line). The reconstructed vertex by FastHisto is depicted in black, closely aligned with the true vertex, signifying a well-reconstructed vertex. The success of this reconstruction can be attributed to the high p_T peak at the vertex, with clearly visible high p_T PV tracks, simplifying the vertex reconstruction process for FastHisto [17].

After computing the position of the primary vertex, FastHisto performs a track-to-vertex association step. Tracks are assigned to the vertex based on the PV position and the Z positions of the tracks. For this assignment, the simplest method is to associate each track with a predetermined distance from the PV point.

A more sophisticated, multivariate technique is developed to associate individual tracks to the PV. This method utilizes additional track parameters η , p_T , and the distance between the track's Z and the reconstructed PV. These input variables are used to train a boosted decision tree (BDT) to classify the association of each track with the PV. By incorporating parameters that are known to correlate with the track's z_0 resolution, the BDT enables a more selective association of tracks with the PV.

The PFA, differently from FastHisto, incorporates the η information during the process of identifying the primary vertex. The FastHisto uses this variable only for the track-to-vertex association that happens after the PV reconstruction. The PFA associates with each track a Gaussian weighted p_T sum (Eq. 9.3, Section 9.1) while FastHisto weights each track by their p_T .

Track-to-vertex association in the PFA could be easily implemented. For each point where we evaluate the figure of merits of the algorithm, tracks scoring Gaussian weight p_T sums above a certain threshold are saved. When the PV is reconstructed, the tracks selected from this point are then associated. In this way, the PFA could perform the primary vertex reconstruction and the track-to-vertex association step at the same time. The following plots show the comparison between the PFA and the FastHisto performances over different MC samples.

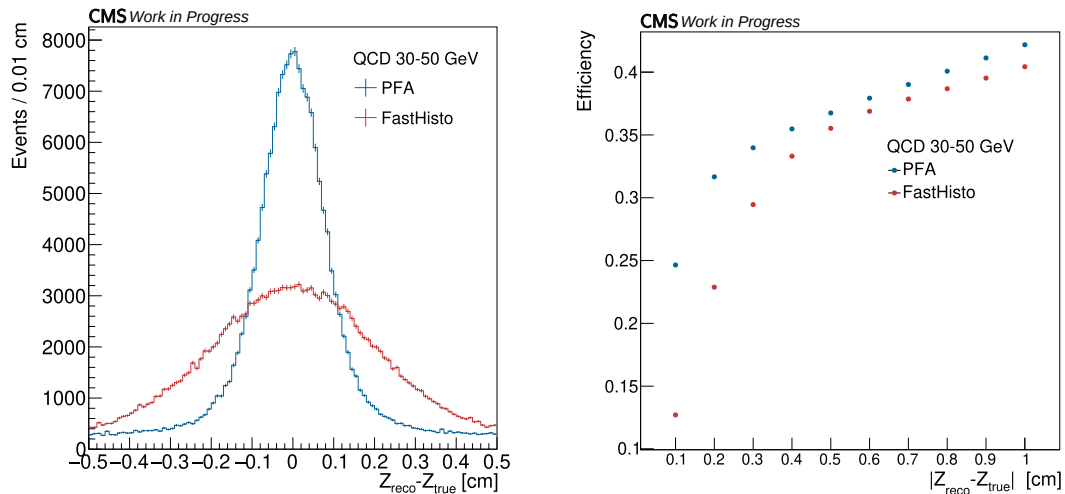


FIGURE 9.29: Residual plot on the left and efficiency plot on the right, comparing the PFA and the FastHisto algorithm using the QCD 30-50 GeV sample.

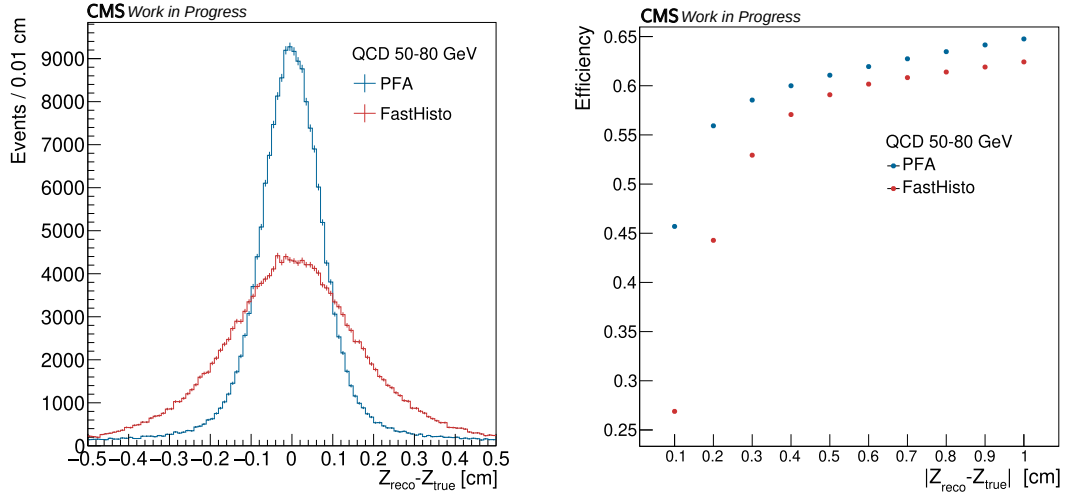


FIGURE 9.30: Residual plot on the left and efficiency plot on the right, comparing the PFA and the FastHisto algorithm using the QCD 50-80 GeV sample.

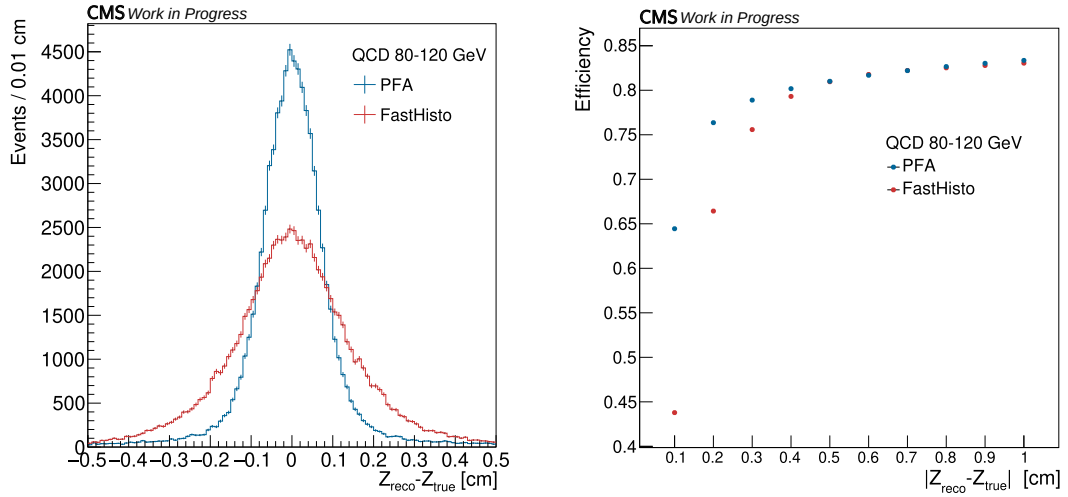


FIGURE 9.31: Residual plot on the left and efficiency plot on the right, comparing the PFA and the FastHisto algorithm using the QCD 80-120 GeV sample.

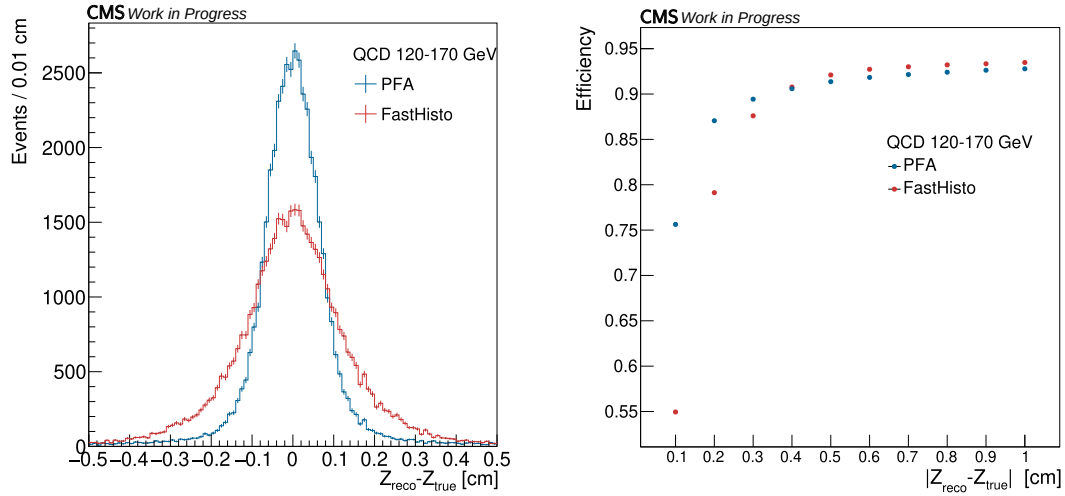


FIGURE 9.32: Residual plot on the left and efficiency plot on the right, comparing the PFA and the FastHisto algorithm using the QCD 120-170 GeV sample.

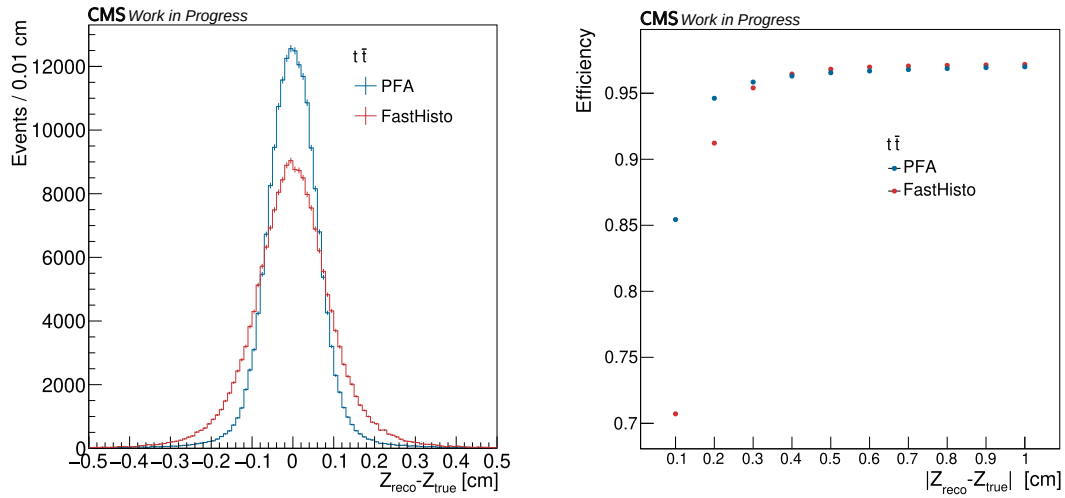


FIGURE 9.33: Residual plot on the left and efficiency plot on the right, comparing the PFA and the FastHisto algorithm using the $t\bar{t}$ sample.

It is visible from Fig. 9.29-Fig. 9.33 that the PFA outperforms FastHisto both in resolution and efficiency. Within the range of interest, which is $\sim Z_{reco} - Z_{true} < 0.5$ cm, PFA always shows better efficiency in PV reconstruction, especially in the low energy samples, where the use of η information in the reconstruction process is crucial to include low-resolution tracks pointing towards the endcaps.

9.3 The implementation of the algorithm on an FPGA

In this section, we will give an overview of FPGAs (Field-Programmable Gate Arrays) and describe their future use in the High-Lumi L1 Tracker Trigger of the CMS experiment. We will show that the Peak Finding Algorithm is implementable on FPGAs, using very low resources of the chip and fits well within the High-Lumi L1 trigger constraints. This compatibility makes this algorithm a good candidate for incorporation into the L1 Trigger Phase-2 upgrade.

9.3.1 Field-Programmable Gate Array

An FPGA is a type of integrated circuit that falls into the category of configurable hardware, allowing for programming or reprogramming after manufacturing [87]. While CPUs or GPUs may feature clock speeds ten times higher, measured in GHz compared to FPGAs which typically operate in the hundreds of MHz range, FPGAs offer crucial advantages: high data throughput (of the order of Tbits/s) and fixed and low latency (on the order of 100 ns). These features are often challenging to achieve with CPUs or GPUs. The low latency ensures rapid response times critical for application in high energy physics trigger and data acquisition systems. Moreover, FPGAs excel in parallel processing capabilities. This means that all data flowing through the FPGA can be processed concurrently, subject to the limitations imposed by the resources available on the chip.

FPGAs excel in projects demanding adaptability, flexibility, and high-frequency operations. Given their ability to accommodate evolving needs, FPGAs stand out as a prime choice for integration into triggering systems [17]. For all these reasons, the CMS experiment uses FPGAs for its L1 triggering system.

In an FPGA, there are several resources available that can be configured and utilized to implement various digital designs. The most important are here briefly described.

- Logic elements are the fundamental building blocks of an FPGA. They typically consist of lookup tables (**LUTs**) and flip-flops (**FFs**) or registers. LUTs are programmable memory units that can implement combinatorial logic functions such as AND, OR, NOT gates. FFs or registers are used to store intermediate or final results in sequential logic circuits.
- Block RAM, also known as memory blocks or **BRAM**, provides dedicated memory resources within the FPGA. BRAMs are used for storing data arrays, lookup tables, or other data structures. They offer fast access times and can be configured to various sizes depending on the FPGA's architecture.

- Digital Signal Processor blocks (**DSPs**) accelerate integer arithmetic. They include hardware multipliers, accumulators, and other arithmetic logic units customized for efficient signal-processing operations.

Additionally, several significant considerations must be taken into account when programming FPGAs, some of which are outlined below.

Performing floating-point arithmetic operations on FPGAs uses a lot of resources, so whenever possible it is preferable to use integers instead.

The range of the floating variables we want to reproduce should be known in advance because it determines the number of bits that will be used. Determining the necessary bit count for every internal signal is essential, as extra bits mean more resources are used.

For basic FPGA components it is inefficient to perform some mathematical operations, such as division. To get the needed functionality, these operations must be carefully broken down into sets of additions, multiplications, bit shifts and lookup tables.

Timing is very important. Certain actions need a certain amount of clock cycles; thus, to guarantee correct functioning, meticulous signal synchronisation is required. A clock in an FPGA system is responsible for driving the FPGA design and determines how fast it can run and process data. Higher clock frequencies mean shorter time for each task, limiting complexity per task. Lower frequencies allow more complex tasks to be done in each cycle. In order to achieve the appropriate clock speed and prevent timing problems and inaccurate outputs in the design, it is essential to manage the path lengths and internal complexity of FPGA logic.

The most popular hardware description languages (HDLs) for programming FPGAs are VHDL and Verilog. These languages are needed for describing algorithms on FPGA. They are recognized as low-level languages, designed to be readable by both machines and humans. High-level synthesis (HLS) tool, instead allows designers to describe the functionality of an algorithm in a high-level language, such as C++ and automatically generate the corresponding low-level code.

After writing an algorithm for the FPGA, referred to as firmware, the next step is *synthesis*, which translates the algorithm to the hardware components. The algorithm can undergo *simulation*, where the data flow across the algorithm is simulated one clock cycle at a time. These simulations serve to validate the synchronization of the algorithm and verify that the output aligns with expectations. Concurrently with firmware development, an *emulation* of the firmware is also created, usually in C++. This emulation facilitates rapid algorithm development and serves as a reference for comparing against the firmware simulation and hardware implementation. Emulation enables the assessment of physics performance with simulated events or real collision data.

Once the synthesized mapping is considered accurate, it is transferred onto a target FPGA, where the actual resources specified in the synthesis are routed on the chip. Additionally, during the implementation phase, a target clock frequency is determined, dictating the operating speed of the FPGA.

The duration required for data to traverse the algorithm from start to finish is termed the algorithm's latency, which can be quantified in clock cycles or total time (clock cycles divided by clock frequency). It is important to note that the latency of the algorithm does not prevent the FPGA from accepting new data. If the algorithm permits, the FPGA can accept new data before the previous computation has finished, as the acceptance of new events is decoupled from the algorithm's latency.

When developing firmware for FPGAs, there are a few potential problems and limitations that must be addressed. These are power consumption, frequency, and resource utilization, which are interconnected. For the L1 trigger the power budget is 100 W for an FPGA and ~ 200 W for a board (which can host an FPGA chip along with other necessary components such as power supplies, input/output interfaces, and connectors).

9.3.2 FPGA at LHC

FPGAs have long been utilized at the LHC in the L1 trigger. Over time the technology has significantly advanced and the FPGA that will be used in HL-LHC will be Virtex Ultrascale+ VU13P [88, 89].

As we described in Section 8.1.3, the outer tracker identifies stubs that are the pairs of tracker hits in the two-layered modules which correspond to high-pT tracks. The OT DTC (Outer Tracker Data, Trigger and Control) system reformats these stubs and sends them to the Track Finder (TF) system which reconstructs tracks from the stubs.

As is visible from Fig. 9.34, in the Track Finder system there are 18 sets of boards, each receiving data in a different time slice. A time slice corresponds to the time interval between two bunch crosses which is 25 ns. Each event is processed in a different time slice, so after 18 events that means after 450 ns (18×25 ns) the first sets of boards, corresponding to Time Slice 0, is filled again with a new event.

Each time slice is associated with various boards housing FPGAs and additional modules. The boards are connected by optical links. These optical links are capable of transmitting data at a rate of 25 Gigabits per second (Gbps). The data transmitted across the optical links corresponds to a rate that is 9 times the frequency of the LHC clock (40 MHz), which is 360 MHz. Each unit of data sent across the optical links has a width of 64 bits. To calculate the effective data transmission rate, we multiply the number of bits per clock cycle (64) by the frequency of data transmission (360 MHz).

In total, there are six GTT time slices, with each one receiving data from three TF slices. There are links between the TF and the GTT sending track information, links between

GTT and the Correlator sending vertices and between GTT and GT sending vertices, Jet reconstruction and E_T miss information. The Correlator merges the information it receives from the L1 tracker, the L1 calo and the L1 muon.

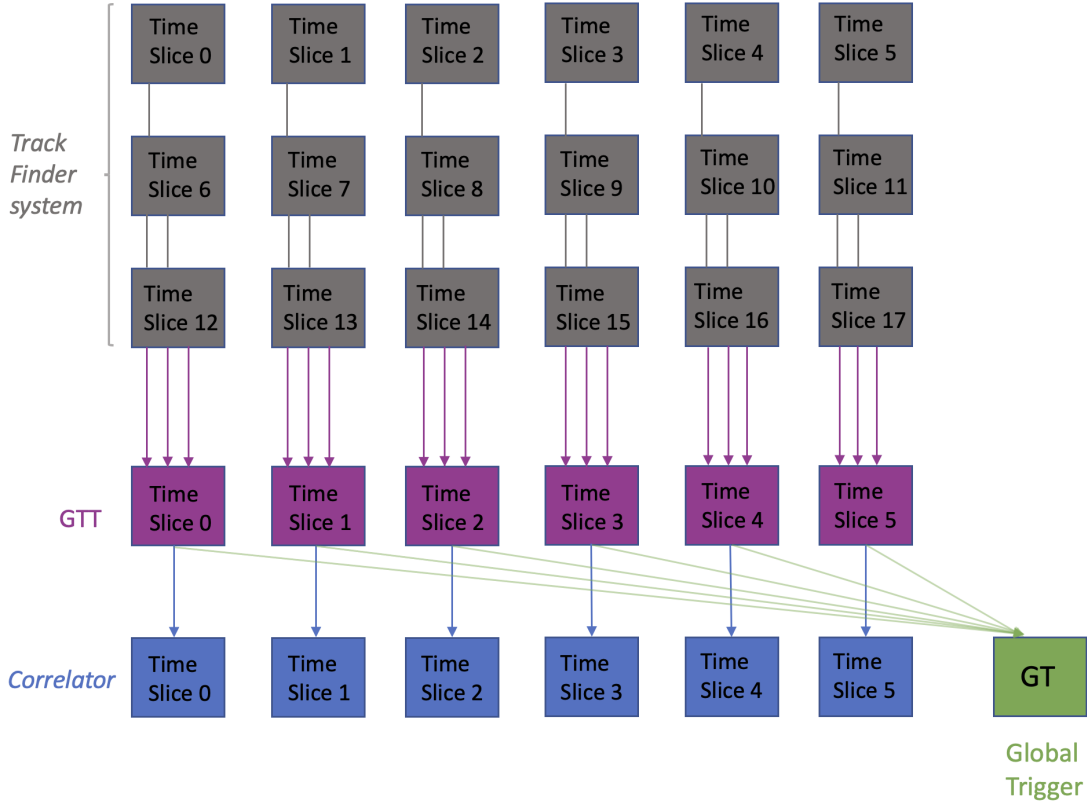


FIGURE 9.34: Illustration of the L1 Track Trigger system. Track Finder system, GTT, Correlator and GT are presented.

The Vertex Finding algorithm will be integrated into the GTT boards alongside two additional algorithms: one for Jet reconstruction and another for calculating E_T miss.

9.3.3 Implementation of Peak Finding Algorithm on an FPGA

In developing the Peak Finding algorithm in VHDL, we prioritized maintaining efficiency and accuracy while adhering to the constraints discussed earlier in Section 9.3.1. As in FPGA, it is a more efficient use of resources using integers over floating-point numbers, we translated continuous floating-point ranges into discrete sets of integers. This process may potentially compromise precision. To optimize performance and resource utilization, we allocated a limited number of bits to represent each variable, determined by both the variable distribution and the resolution needed. This process if appropriately tailored for each variable can have a negligible impact on the algorithm performance. Throughout the process of translating the algorithm into VHDL, we ensured that its efficiency remained consistent with the results obtained through the C++

implementation. To validate this consistency and verify the algorithm's physics performance, we compared the VHDL implementation with a bit-accurate simulation written in C++.

The z-axis range which is taken into account by the algorithm extends from -15 cm to +15 cm. For the implementation of FPGA, this range is divided over a total of 320 bins each of 0.94 mm. Notably, all these bins are processed in parallel. One track per clock tick is propagated through the bins as in the chain depicted in Fig. 9.35. One clock tick corresponds to 360 MHz which corresponds to 2.77 ns.

In this chain implementation, the latency of the algorithm breaks down into three different contributions: the time required to upload the entire event which is 450 ns, the time to propagate tracks through each bin corresponds to $320 \text{ bins} \times \text{clock tick}$ (2.77 ns) and the computation time needed to perform all the operations within the algorithm which is 5 clock ticks. The total latency appears to be: $L = 450 \text{ ns} + (320 + 5) \times 2.77 \text{ ns} = 1350.25 \text{ ns}$.

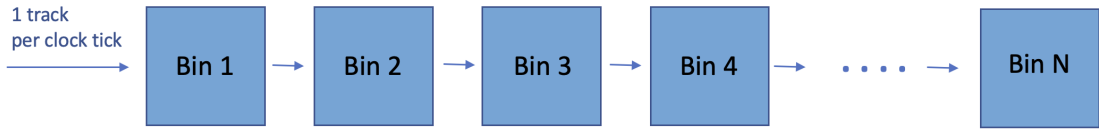


FIGURE 9.35: Chain scheme used to implement the PFA algorithm on FPGA.

Processing the PFA algorithm through this long chain causes two different problems. The first comes from the latency which is slightly higher than the targeted one (1ms), and the second problem arises from the fact that this algorithm can process at most 162 tracks at a time. The number of tracks comes from $450 \text{ ns} / 2.77 \text{ ns}$. Some events have more than 162 tracks. To solve both problems a solution is to use "workers". The implementation of workers in firmware is left for future work but the scheme of how it would work is described below. A sketch of the possible system with workers is shown in Fig. 9.36. A worker comprises multiple bins, each potentially varying in size. In our scenario, we could divide the z-axis range into four distinct regions, each with different dimensions. Two larger regions could cover the extremities of this range, where there are typically fewer tracks, as more distant from the point of highest density of pp interaction. Conversely, smaller bins could be utilized in regions where tracks are concentrated, such as the central region within the Z range.

In the long-chain case implementation, the event is fully propagated through each bin, allowing for the bin a comprehensive view of the data. Conversely, when using workers, each bin will be allocated to a specific worker and will only be able to analyze a portion of each event. As a result, we effectively reduce the latency to:

$$L = 450 \text{ ns} + \left(\frac{320}{N_{\text{workers}}} + 5 \right) \times 2.77 \text{ ns}. \quad (9.6)$$

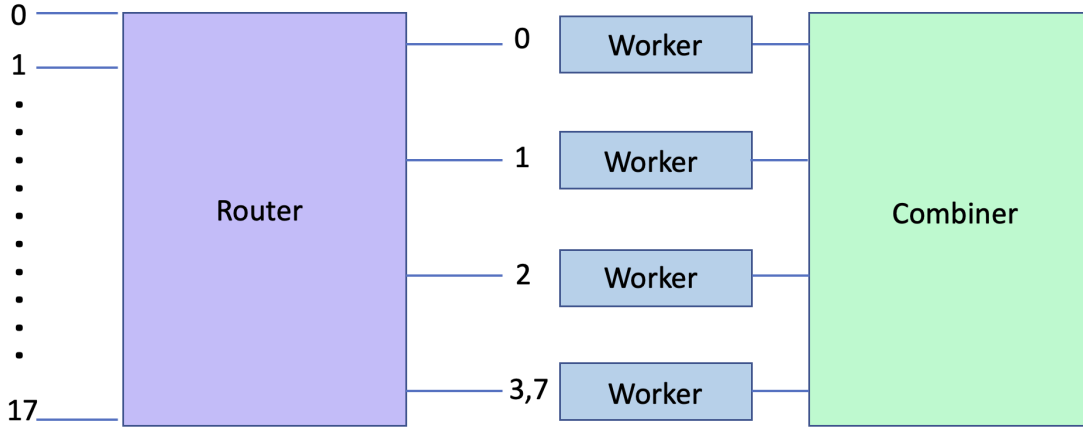


FIGURE 9.36: Scheme of a possible Router, several workers and a final Combiner.

Dividing the workload among multiple shorter chains allows for faster processing. On the other hand, different workers will compute different vertices relevant to the range that they cover. To come out with a single vertex decision, vertices computed by different workers need to be aggregated. The most appropriate vertex needs to be selected based on predefined criteria that will ensure consistency and accuracy in determining the final vertex location. This is the Combiner stage. Having the worker's regions overlapping, could remove the edge effect that otherwise may potentially reduce the efficiency of the algorithm. The Combiner should make sure that this process does not bring to vertices double counting.

As mentioned before, the implementation of workers and the Combiner is not done yet in firmware but is left for future work. Based on our estimates, employing a number of workers ranging from 4 to 8 should keep the latency comfortably within the targeted limit.

Different logic elements are necessary for the algorithm implementation. For example, the sum of the Gaussian weights is accomplished using logic elements like lookup tables and flip-flops. These components handle the task of summing up the weights. On the other hand, Digital Signal Processors are employed to multiply the Gaussian weights by the transverse momentum p_T . The DSPs perform both the multiplication of individual Gaussian weights with p_T and the subsequent summation of these products.

As mentioned in section 9.3.2, each GTT board has implemented three algorithms (Vertex Finding, Jet reconstruction and ET miss) and processes three events at a time, this reduces by a factor 1/9 the available resources. Approximately 11.11% of the space can be maximally occupied by the Vertex Finding algorithm to ensure an equal allocation of resources for each algorithm on the chip.

The FPGA overall resource utilization is depicted in Tab. 9.1. The total resources used by the PFA comfortably fit within all resource constraints.

Resource	Utilization	Available	Utilization%
LUT	46345	1728000	2.68
FFs	35436	3456000	1.03
DSP	320	12288	2.60
BRAM	0	2688	0

TABLE 9.1: Resource utilization of the PFA algorithm.

We can have a rough comparison with the FastHisto resources which are instead shown in Tab.9.2.

Resource	Utilization	Available	Utilization%
LUT	134784	1728000	7.8
FFs	114048	3456000	3.3
DSP	0	12288	0
BRAM	129.024	2688	4.8

TABLE 9.2: Resource utilization of the FastHisto algorithm [17].

It is noteworthy that both comfortably adhere to our resource constraints. Moreover, they both have an acceptable latency. FastHisto uses 38 clock cycles [17], so its latency is equal to $450 \text{ ns} + 38 \times 2.77 \text{ ns} = 555.26 \text{ ns}$. Using 8 workers in PFA we could have a latency of 574.65 ns which is perfectly within the constraint and comparable with the FastHisto latency.

After the FPGA implementation, In Fig. 9.37-Fig. 9.41, we compare again the two algorithm performances to show the impact of the new bin size of 0.94 mm coming from the implementation of the algorithm on FPGAs. The “PFA.FPGA” label in these plots refers to the new bin size implementation. We can notice that even if the performances slightly decrease, they are still better than the FastHisto ones.

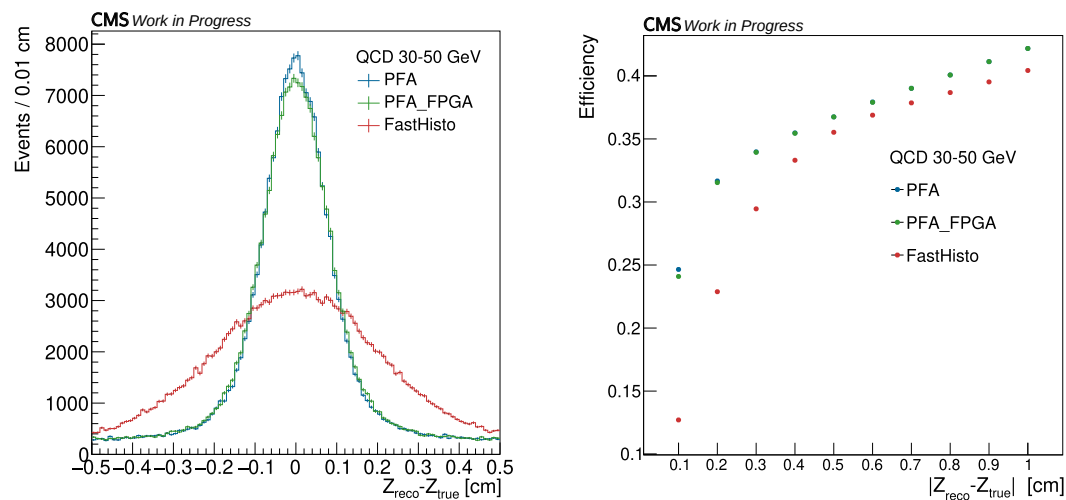


FIGURE 9.37: Residual plot on the left and efficiency plot on the right, comparing the PFA algorithm, its version implemented on FPGA (with a different bin size of 0.94 mm) and the FastHisto algorithm using the QCD 30-50 GeV sample.

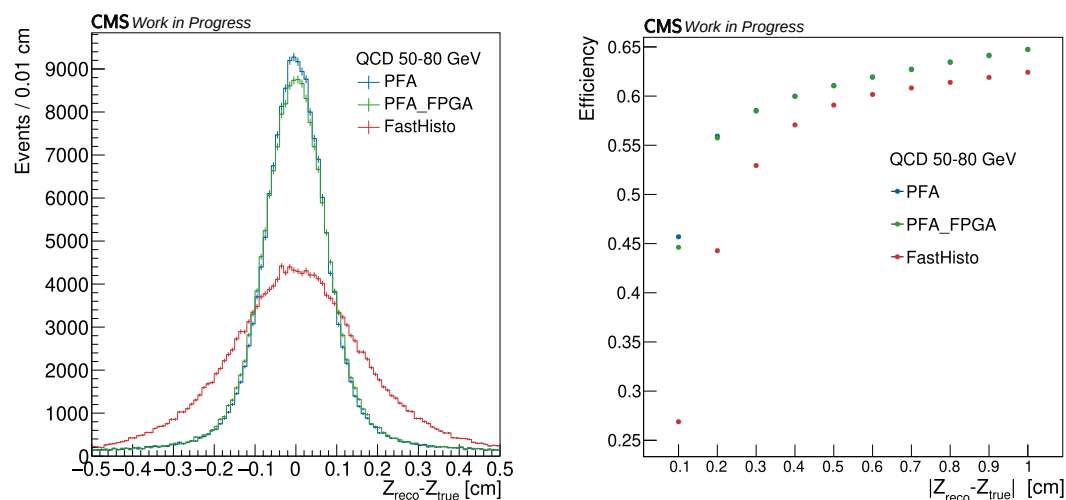


FIGURE 9.38: Residual plot on the left and efficiency plot on the right, comparing the PFA algorithm, its version implemented on FPGA (with a different bin size of 0.94 mm) and the FastHisto algorithm using the QCD 50-80 GeV sample.

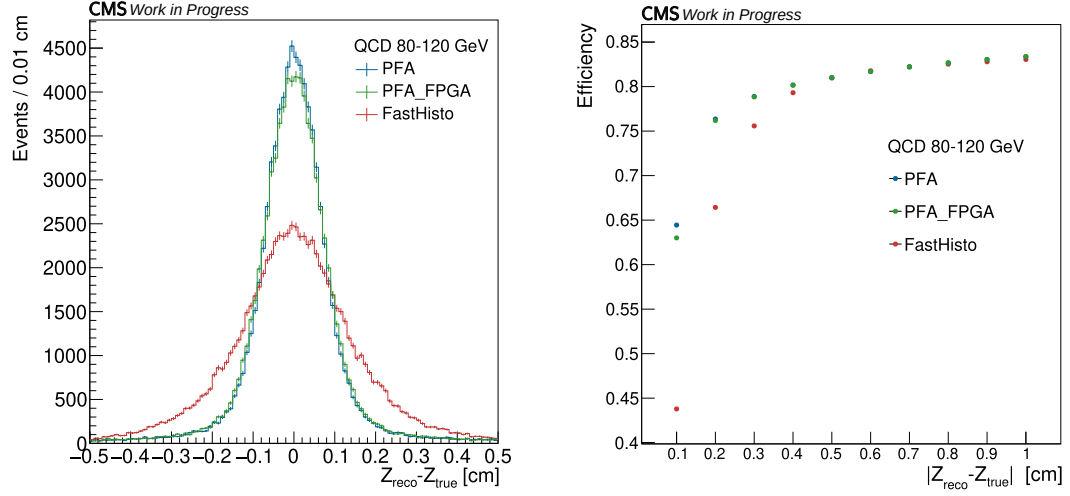


FIGURE 9.39: Residual plot on the left and efficiency plot on the right, comparing the PFA algorithm, its version implemented on FPGA (with a different bin size of 0.94 mm) and the FastHisto algorithm using the QCD 80-120 GeV sample.

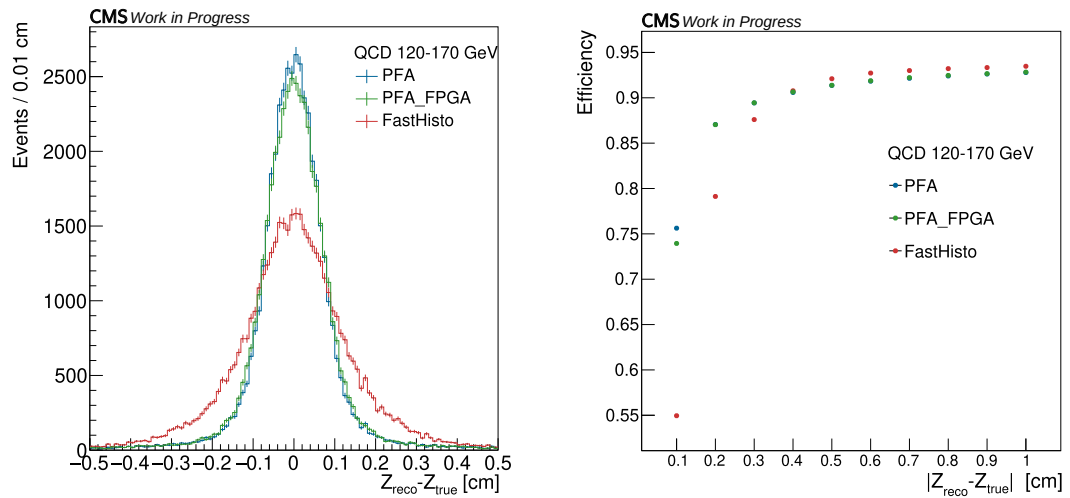


FIGURE 9.40: Residual plot on the left and efficiency plot on the right, comparing the PFA algorithm, its version implemented on FPGA (with a different bin size of 0.94 mm) and the FastHisto algorithm using the QCD 120-170 GeV sample.

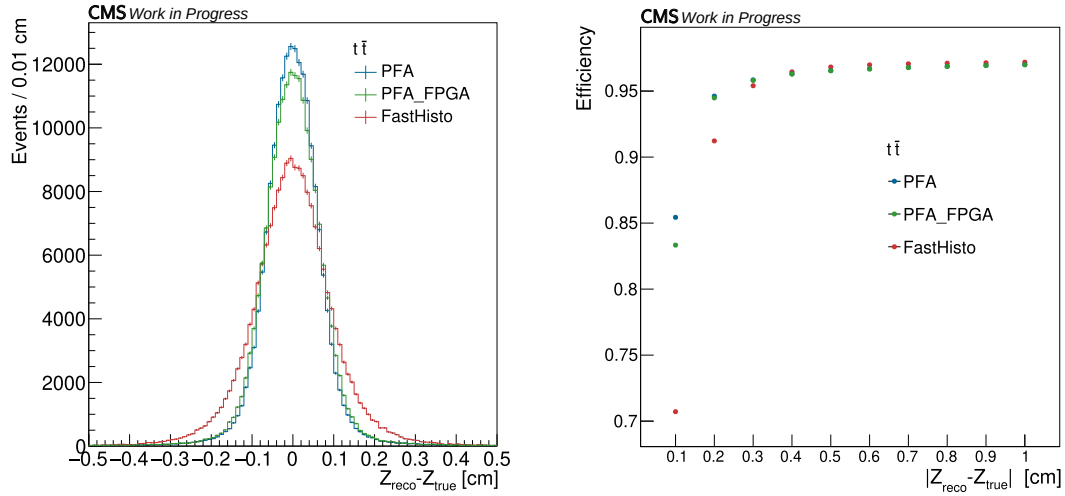


FIGURE 9.41: Residual plot on the left and efficiency plot on the right, comparing the PFA algorithm, its version implemented on FPGA (with a different bin size of 0.94 mm) and the FastHisto algorithm using the $t\bar{t}$ sample.

9.4 Conclusion of the study

In this chapter, we introduced a novel vertex finding algorithm called the Peak Finding Algorithm (PFA) that is tailored for possible implementation at the Level-1 trigger system of the CMS experiment at the HL-LHC.

The first sections delineated the algorithm's design, elucidated the criteria guiding the selection of key variables, and described the improvement achieved by studying some variable optimizations. We showed that using the η variable in the reconstruction process yields noticeable improvements to the algorithm in resolution and efficiency, especially when using low-energy samples.

We compared the proposed algorithm with the baseline approach (called FastHisto), showing superior performances in the reconstruction efficiency and resolution. These advancements can be attributed to the Gaussian weighting methodology and the choice of integrating the η variable as a key variable in the primary vertex reconstruction. FastHisto uses the η information only at the track-to-vertex association step.

Additionally, we described the possibility of reducing the FastHisto two-step process of primary vertex reconstruction and track-to-vertex association into a unique step in the PFA.

The first version of the PFA algorithm has been implemented on FPGAs and it demonstrated that this algorithm fits well within all resource constraints. Both the PFA and the FastHisto are able to satisfy the targeted latency and are implementable as vertex-finding algorithms. The superior efficiency and resolution performances of the PFA

and the possibility of reducing the vertex reconstruction and the tracks to vertex association to a unique step could make this algorithm preferable to the baseline one.

To facilitate the adoption of this algorithm, future steps involve: implementing the track-to-vertex association mechanism outlined in Section 9.2, studying low multiplicity samples to evaluate the potential benefits of the additional cuts discussed in Section 9.1.3, deploying the algorithm on FPGA with the use of workers and a final combiner to minimize latency as described in Section 9.3.3.

Part V

Conclusions

Chapter 10

Conclusions

In the first part of this thesis, an effect present in the scintillating crystals of the Electromagnetic Calorimeter of the CMS experiment at CERN was discussed. These crystals undergo radiation damage with time causing a loss of light transmission, causing a decrease in the amount of collected light. Correction needs to be applied to the transparency lost of each crystal and this is essential to measure the correct ratio of electrons or photons in our detector. Up to the end of Run 2 this correction was performed twice-per-week. In this thesis, we studied the effect of updating these corrections for every LHC fill. Rates from different Level-1 and HLT reconstructed objects were tested and they showed an advantage in increasing the frequency of the transparency correction both in rate stability and energy resolution. This procedure was tested and validated in early 2022 and is now routinely performed in the CMS detector.

The second part of this thesis discusses a particular benchmark model within the NMSSM theory that can be probed by the CMS experiment at the LHC. The Next-to-Minimal Supersymmetric Standard Model (NMSSM) extends beyond the Standard Model, providing resolutions to certain unresolved questions posed by the latter. This analysis focuses on a final state with negligible missing p_T arising from cascade decays of pair-produced squarks or gluinos to a singlino-like LSP particle and a light Higgs boson.

This study builds up on a well-studied theoretical benchmark [61] and on a previously published analysis performed with Run 2 data [14] based on reconstructing the two boosted $H_1 \rightarrow b\bar{b}$ decays as two large-radius jets. This work excluded masses of the light Higgs between 40–120 GeV arising from the decays of squarks or gluinos with a mass of 1200 to 2500 GeV at 95% confidence level for a particular benchmark model. The work discussed in this thesis is preparatory to a new combined analysis that will be performed on Run 2 + Run 3 data. This new analysis extends the squarks or gluinos mass range up to 3200 GeV and includes a new scenario for the benchmark model to which the already measured limits do not apply. Moreover, it introduces the ParticleNet [58] tool to be used in this analysis for the first time. It will be used to both tag the presence of two b quarks and to reconstruct the mass of the large-radius jets.

These new features motivate the need for a new Monte Carlo signal sample production described in Section 5.

Significant improvements in the analysis methodology arise from the introduction of the ParticleNet algorithms. These outperform the b-tagger btagHbb and the mass reconstruction algorithm Softdrop used in the published paper. The enormous improvement in reducing the QCD background performed by ParticleNet changes the background relative importance in the analysis. The $t\bar{t}$ background increases in relative importance compared with the previous search requiring a new reduction strategy which is described in Section 7.2.

After studying all the possible $t\bar{t}$ decay channels, different selections are applied in order to reduce the $t\bar{t}$ background as much as possible, leaving the signal yield almost unchanged. Ultimately, achieving a five-factor reduction in the $t\bar{t}$ background restores its relative importance, once again making the QCD background the predominant background for this analysis in every signal region.

The last part of this thesis discusses a vertex-finding algorithm called the Peak Finding Algorithm (PFA) proposed to be implemented at the Level-1 trigger system of the CMS experiment at the HL-LHC.

The Phase-2 upgrade of the Trigger system of the CMS detector will enable for the first time the track information to be used at the Level-1 trigger. The Level-1 trigger will have a vertex-finding algorithm to reconstruct the primary vertex of each event. The baseline approach described in the technical design report [16] is called FastHisto (FH).

This thesis describes the PFA algorithm from its initial design to the final proposed algorithm. It describes the PFA performances, showing how they outperform the ones of FastHisto in both efficiency and resolution reconstruction. Moreover, while the FastHisto performs the primary vertex reconstruction and the tracks to vertex association in two different steps, it would be possible for the PFA to perform both in a single step. This feature together with the superior efficiency and resolution performances of the PFA, could make this algorithm preferable to the baseline one. Furthermore, the implementation of the PFA on FPGA shows that the algorithm fits well within all the resource constraints of the Level-1 trigger.

Appendix A

DJR plots

This appendix presents some example of DJR plots, described in Section 5.3, fixing XQCUT to 60 GeV and modifying QCUT value, in order to achieve a smooth plot shape. For illustrative purposes three mass points are reported, showing low (800 GeV), medium (2000 GeV) and high masses (2800 GeV).

A.1 First scenario

This scenario is the one in which the gluino is non-decoupled, i.e. $m_{\tilde{g}} = (1 + 1\%)m_{SUSY}$.

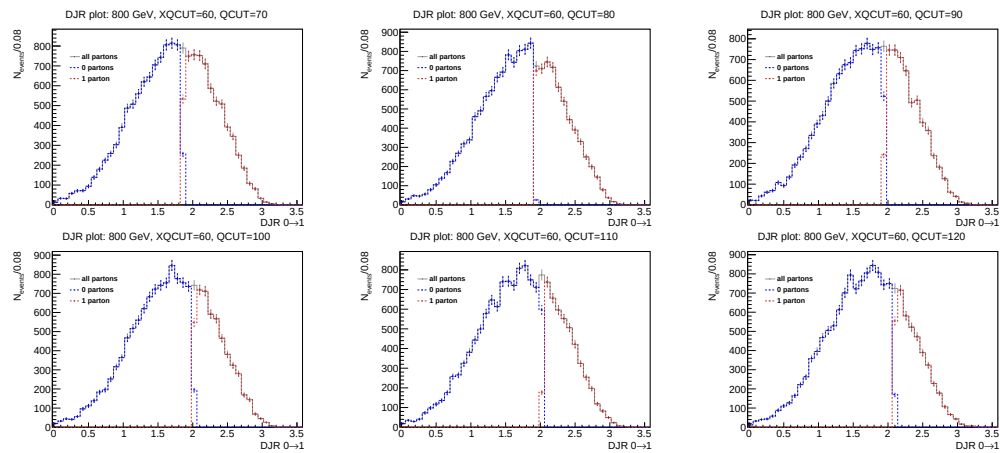


FIGURE A.1: DJR plots for $m_{\tilde{g}} = (1 + 1\%)m_{SUSY}$ where $m_{SUSY} = 800$ GeV, XQCUT=60, QCUT varying

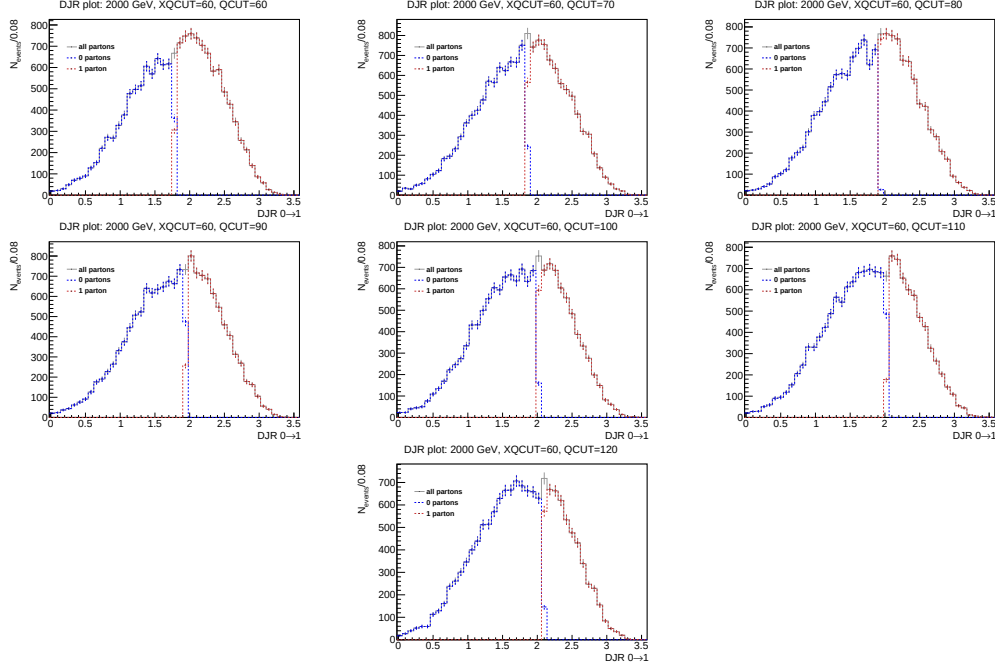


FIGURE A.2: DJR plots for $m_{\tilde{g}} = (1 + 1\%)m_{SUSY}$ where $m_{SUSY} = 2000$ GeV, XQCUT=60, QCUT varying

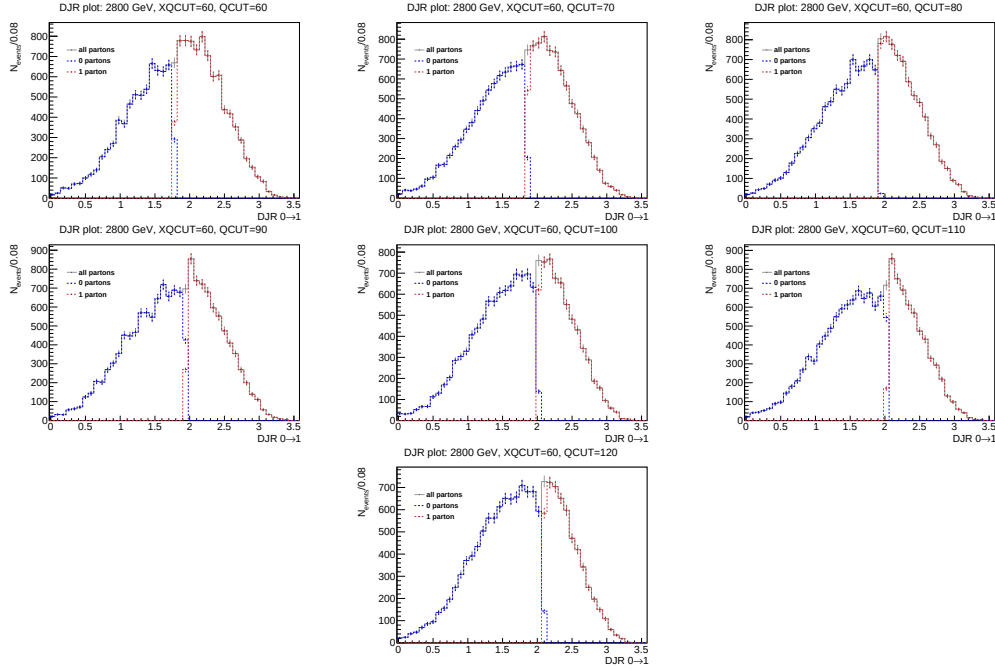


FIGURE A.3: DJR plots for $m_{\tilde{g}} = (1 + 1\%)m_{SUSY}$ where $m_{SUSY} = 2800$ GeV, XQCUT=60, QCUT varying

A.2 Second scenario

In this case, the gluino is decoupled and its mass is fixed to 10^6 GeV.

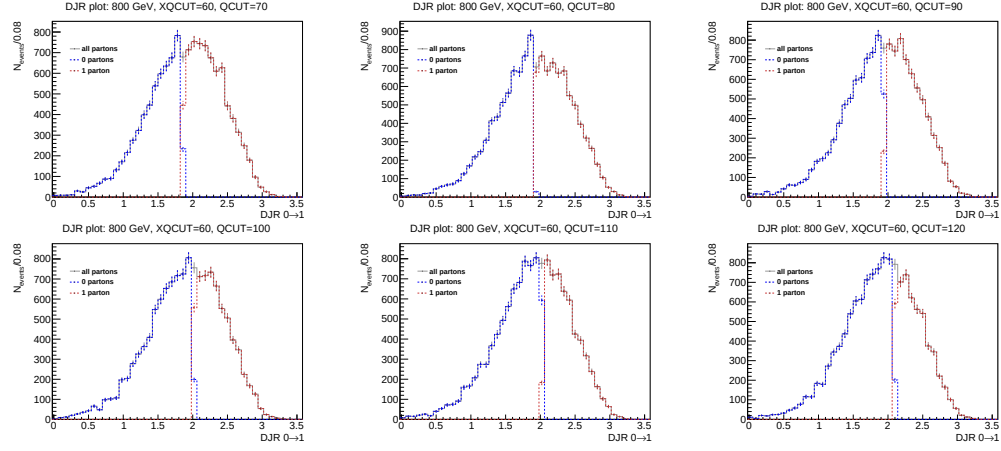


FIGURE A.4: DJR plots for gluino decoupled, $m_{Squark} = 800$ GeV, XQCUT=60, QCUT varying

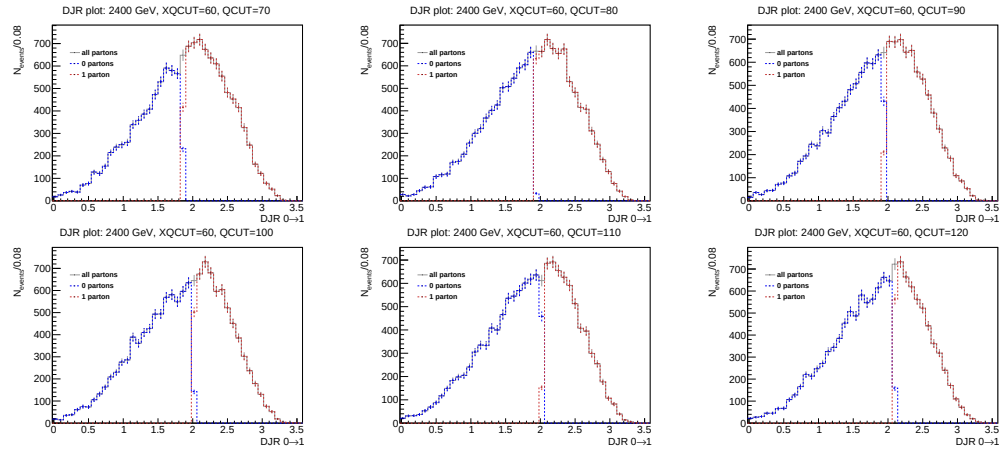


FIGURE A.5: DJR plots for gluino decoupled $m_{Squark} = 2400$ GeV, XQCUT=60, QCUT varying

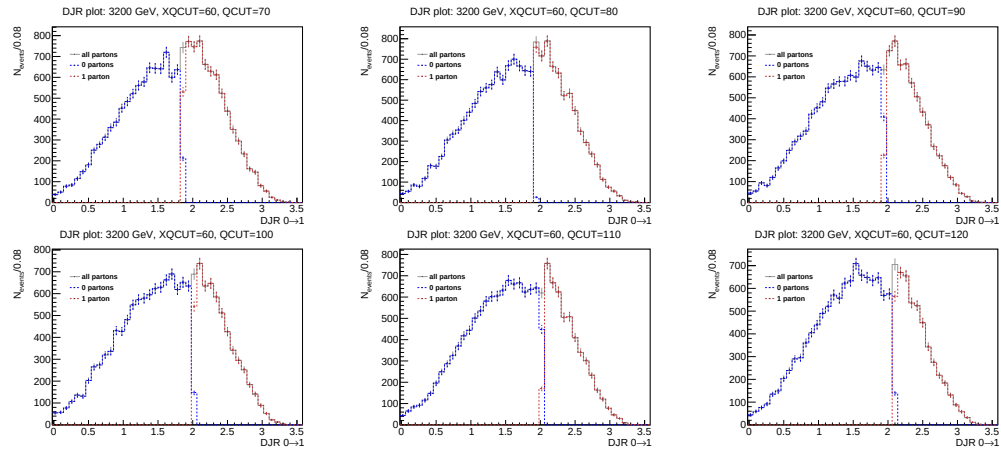


FIGURE A.6: DJR plots for gluino decoupled, $m_{Squark} = 3200$ GeV, XQCUT=60, QCUT varying

Appendix B

Supplementary discrimination factor for PFA

This Appendix reports the results of the same studies performed in Section 9.1.3 but on the QCD samples in a low (30-50 GeV) and high (80-120 GeV) energy range.

B.0.1 QCD 30-50 GeV

All the following plots are performed using the sample QCD 30-50 GeV.

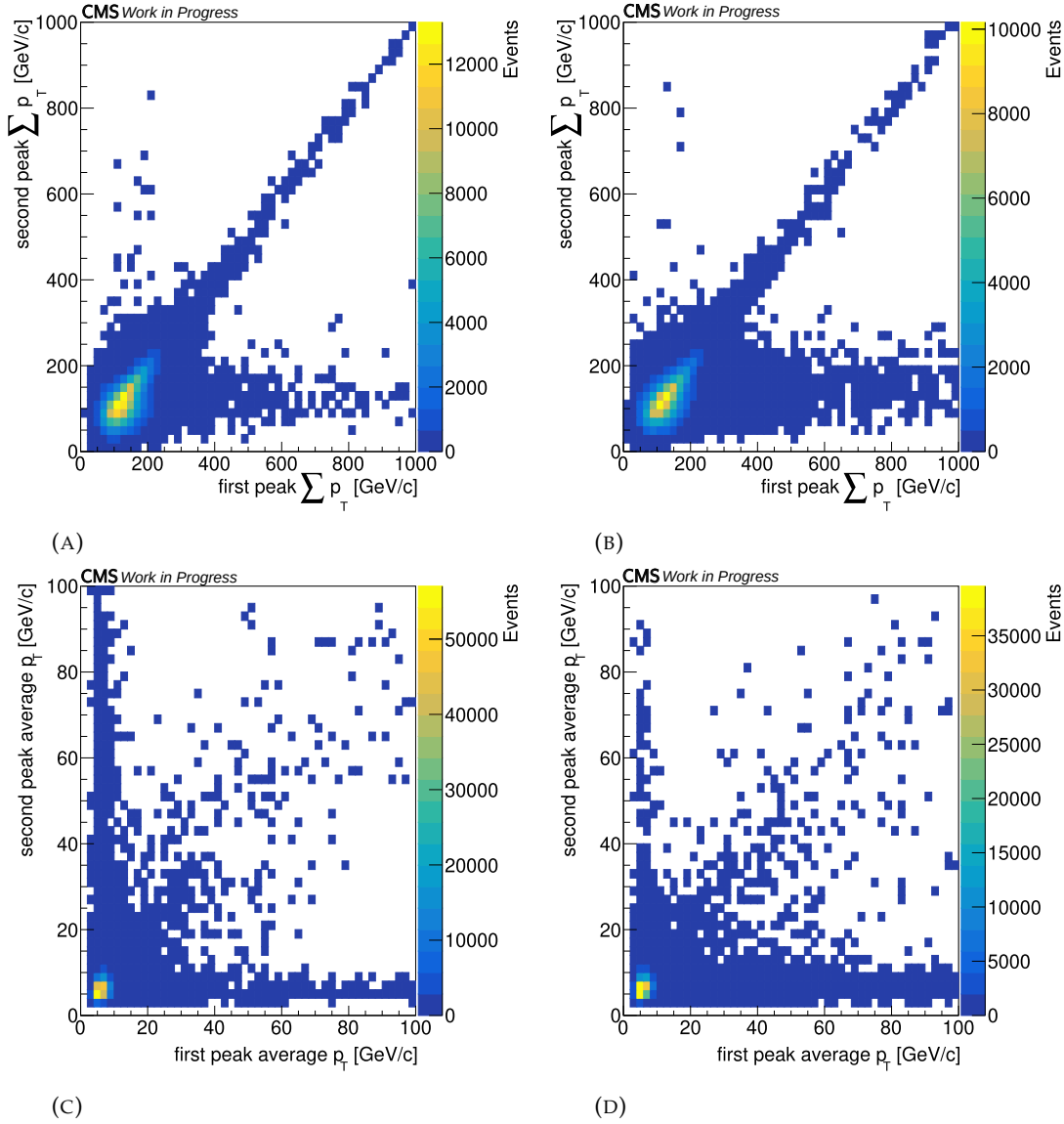


FIGURE B.1: In Fig. (A) and (B) the sum of the transverse momentum ($\sum p_T$) of tracks reconstructing the second peak vs the first peak is shown. Figure (A) represents the case in which the first peak is closer to the true MC PV than the second peak. While Fig. (B) describes the opposite scenario. The same concept is depicted in Figures (C) and (D) for the average p_T quantity.

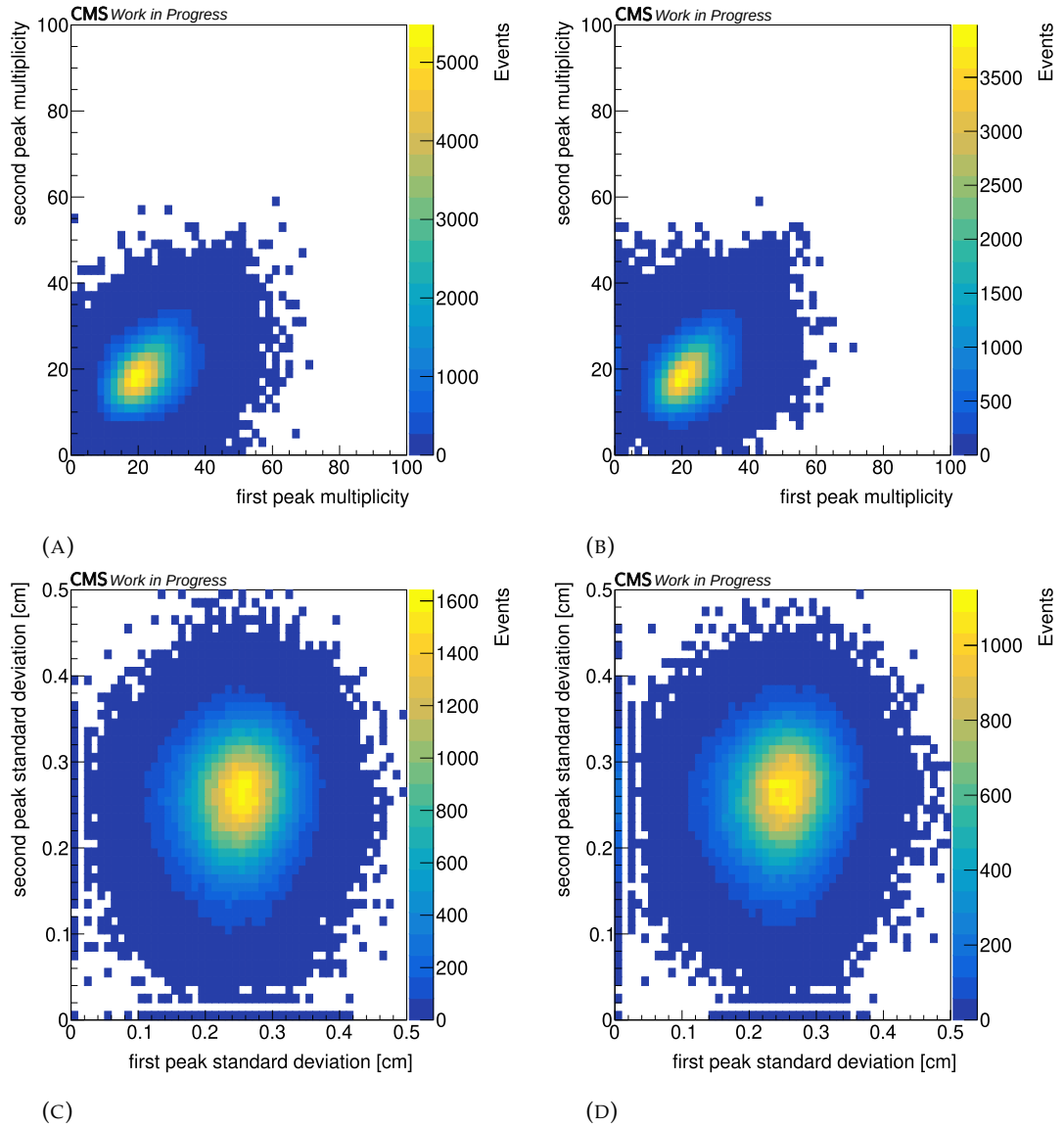


FIGURE B.2: In Fig. (A) and (B) the multiplicity of tracks reconstructing the second peak vs the first peak is shown. Figure (A) represents the case in which the first peak is closer to the true MC PV than the second peak. While Fig. (B) describes the opposite scenario. The same concept is depicted in Figures (C) and (D) for the average standard deviation quantity.

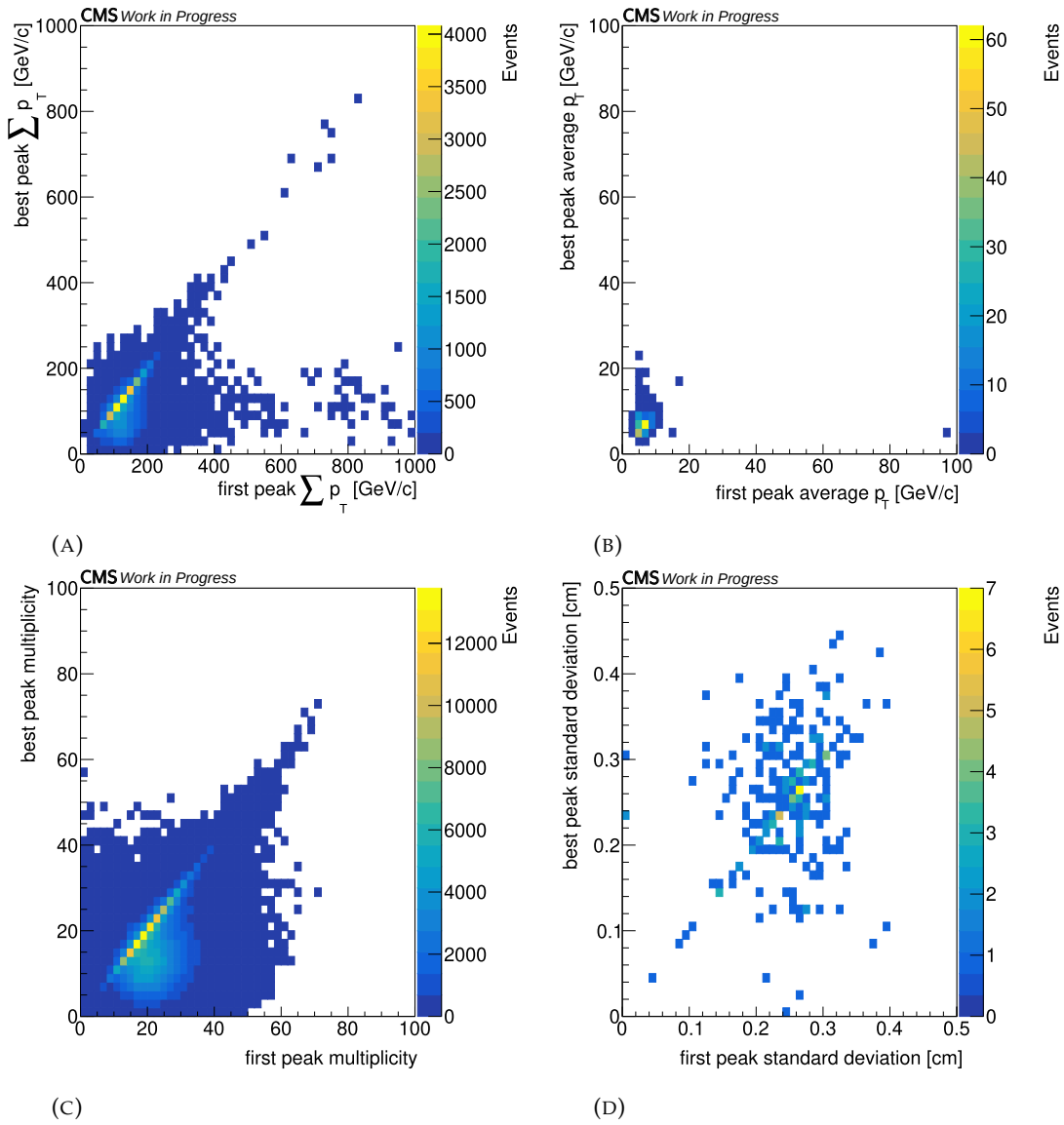


FIGURE B.3: The sum of transverse momenta ($\sum p_T$), average transverse momentum (p_T), multiplicity, and standard deviation are depicted for the best reconstructable peak compared to the first one.

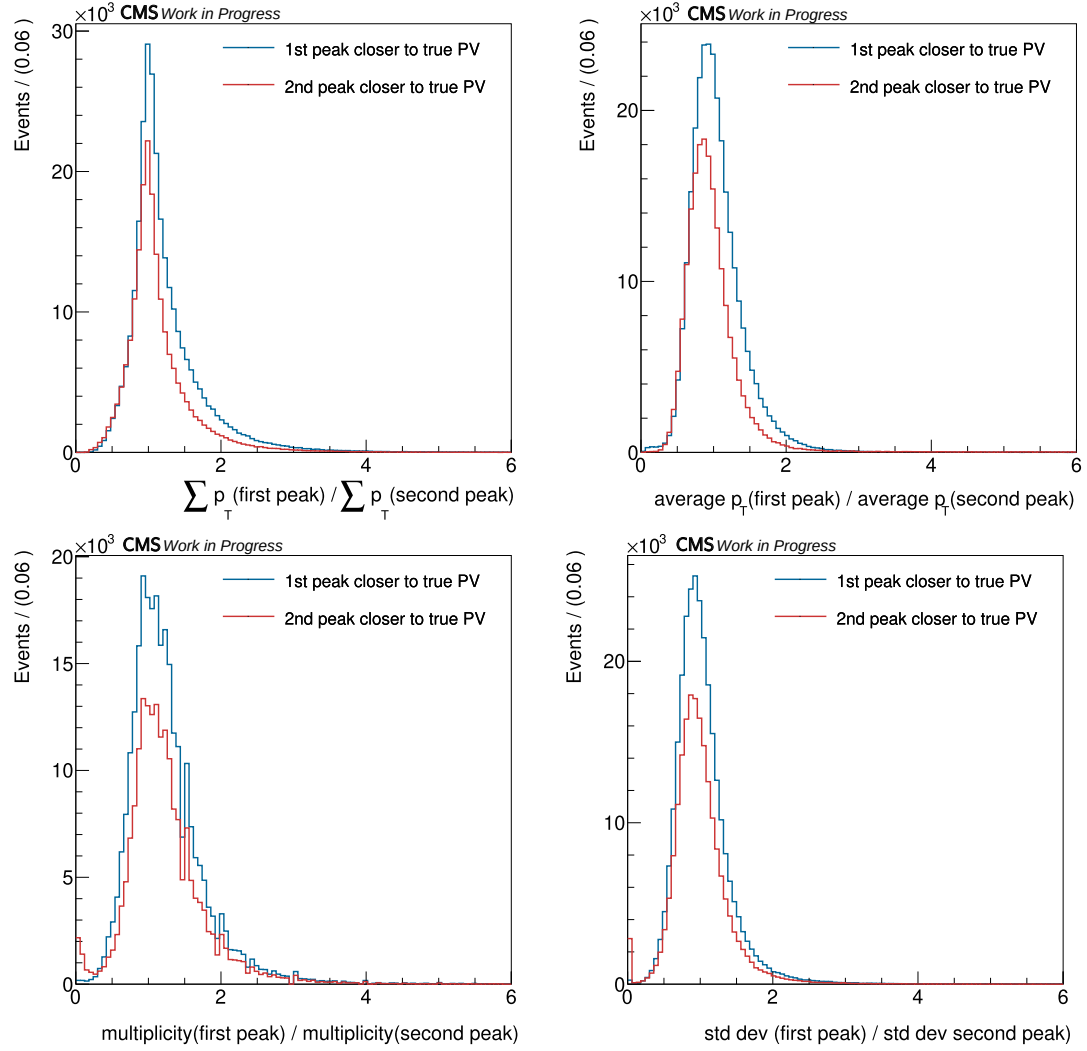


FIGURE B.4: The four plots represent the ratio between four different quantities (sum of transverse momenta (p_T), average p_T , standard deviation, and multiplicity) calculated for the first and the second peak. The blue curve showcases situations where the first peak is closer to the MC PV than the second peak, while the red curve illustrates cases where the second peak is closer than the first to the true MC primary vertex.

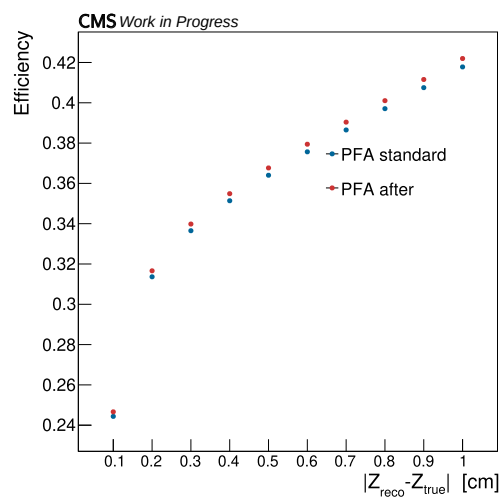


FIGURE B.5: Efficiency curve before and after implementing the described cuts and swaps described in Section 9.1.3.

B.0.2 QCD 80-120 GeV

All the following plots are performed using the sample QCD 80-120 GeV.

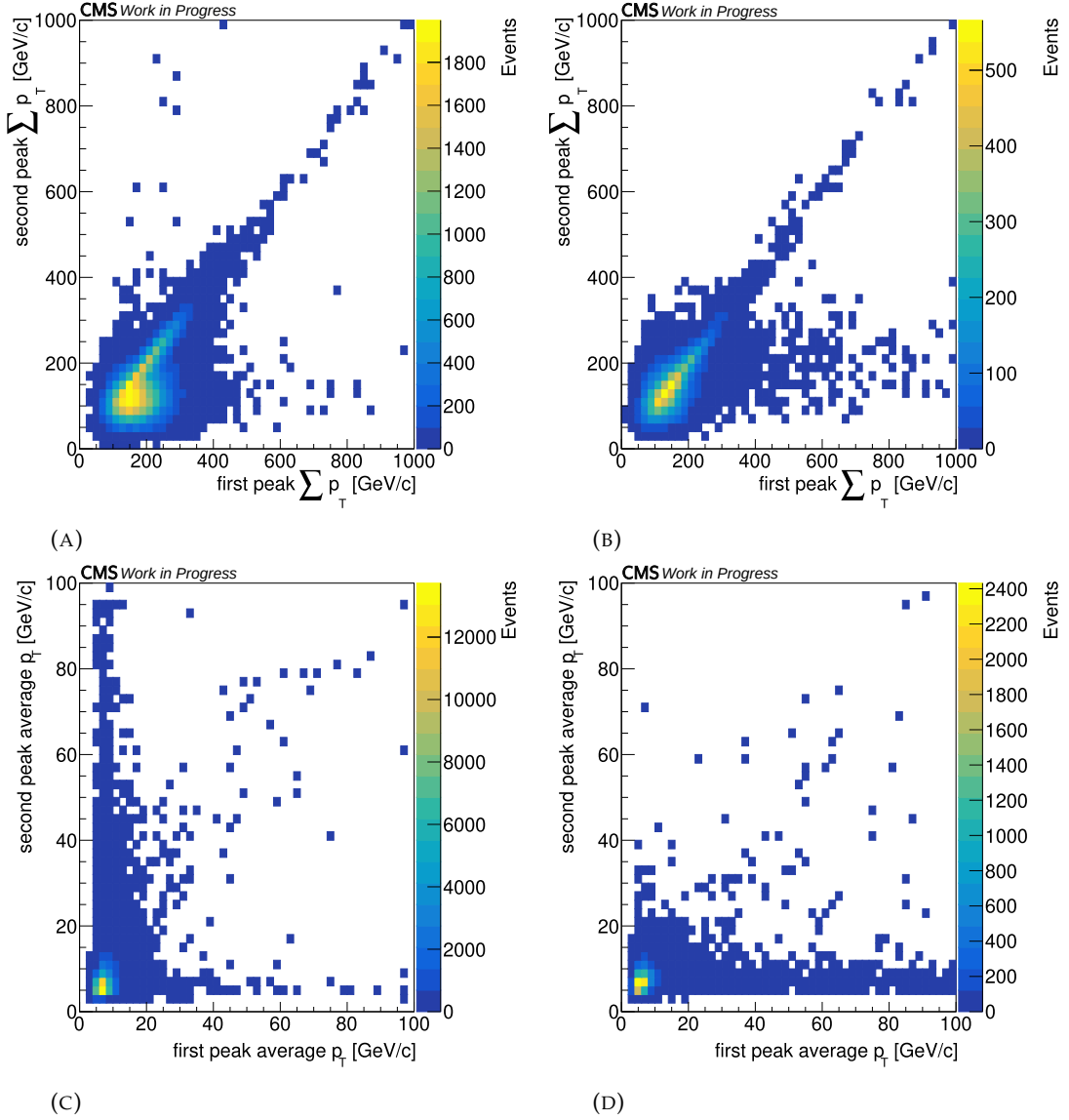


FIGURE B.6: In Fig. (A) and (B) the sum of the transverse momentum (Σp_T) of tracks reconstructing the second peak vs the first peak is shown. Figure (A) represents the case in which the first peak is closer to the true MC PV than the second peak. While Fig. (B) describes the opposite scenario. The same concept is depicted in Figures (C) and (D) for the average p_T quantity.

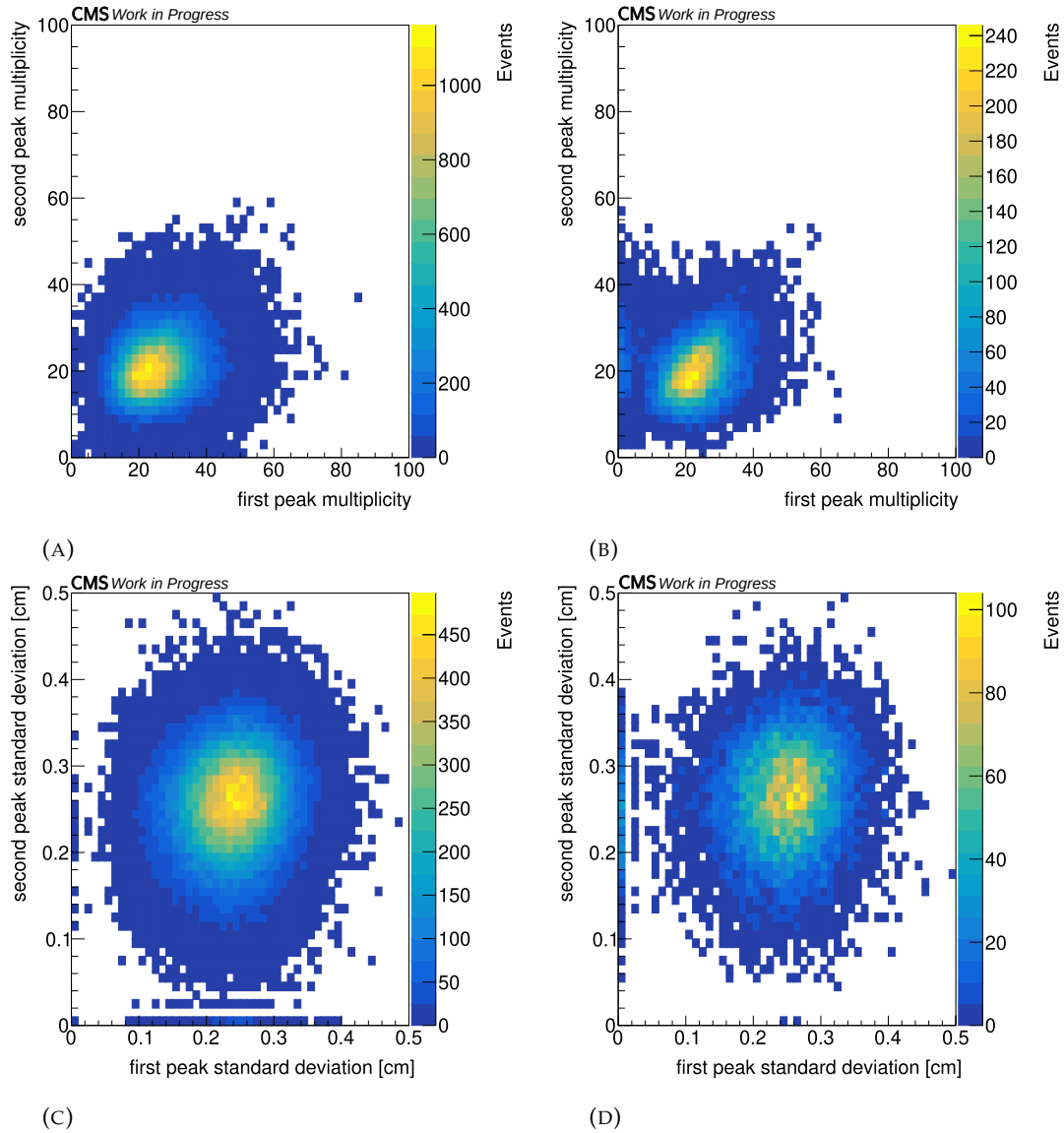


FIGURE B.7: In Fig. (A) and (B) the multiplicity of tracks reconstructing the second peak vs the first peak is shown. Figure (A) represents the case in which the first peak is closer to the true MC PV than the second peak. While Fig. (B) describes the opposite scenario. The same concept is depicted in Figures (C) and (D) for the average standard deviation quantity.

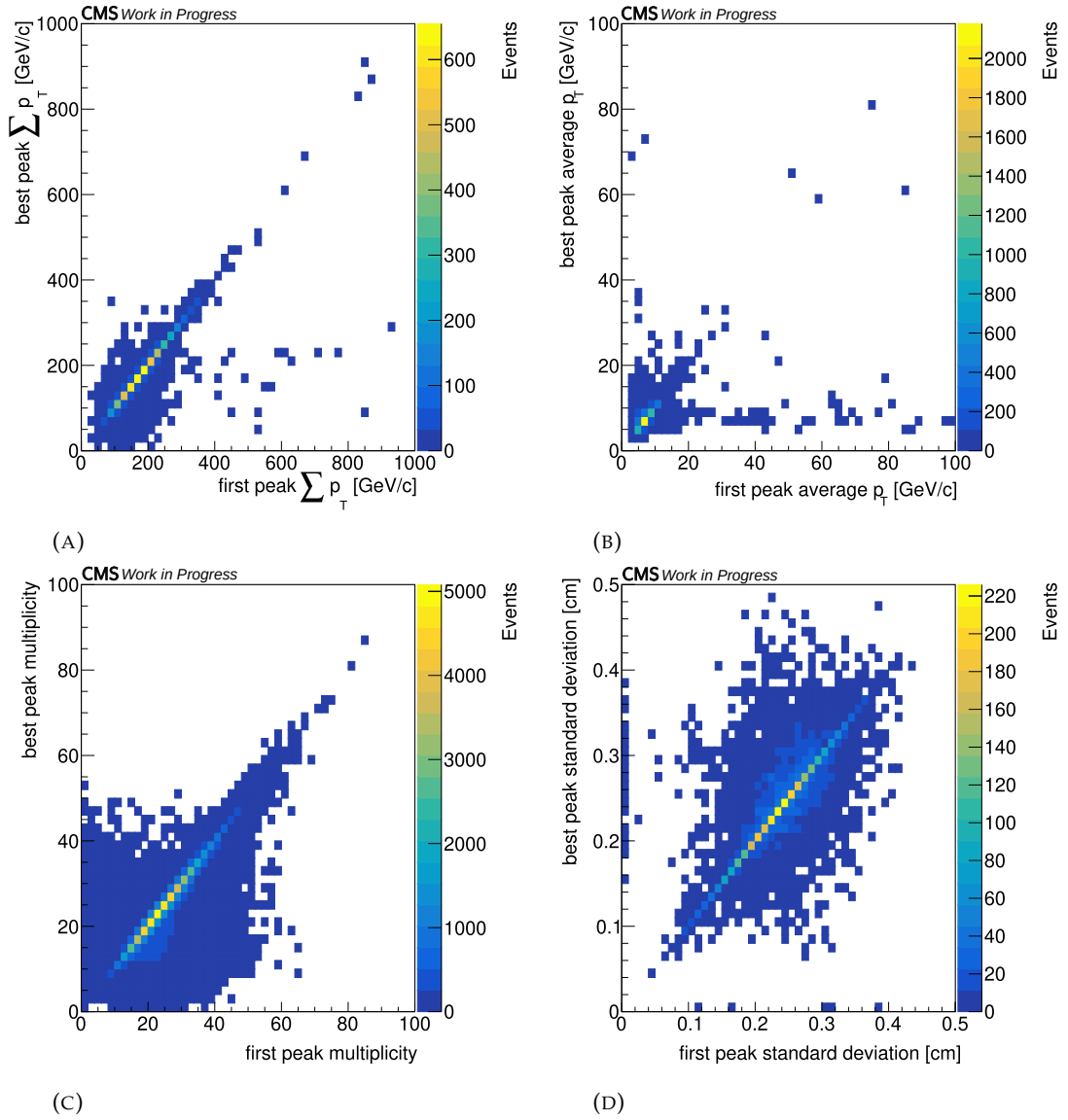


FIGURE B.8: The sum of transverse momenta ($\sum p_T$), average transverse momentum (p_T), multiplicity, and standard deviation are depicted for the best reconstructable peak compared to the first one.

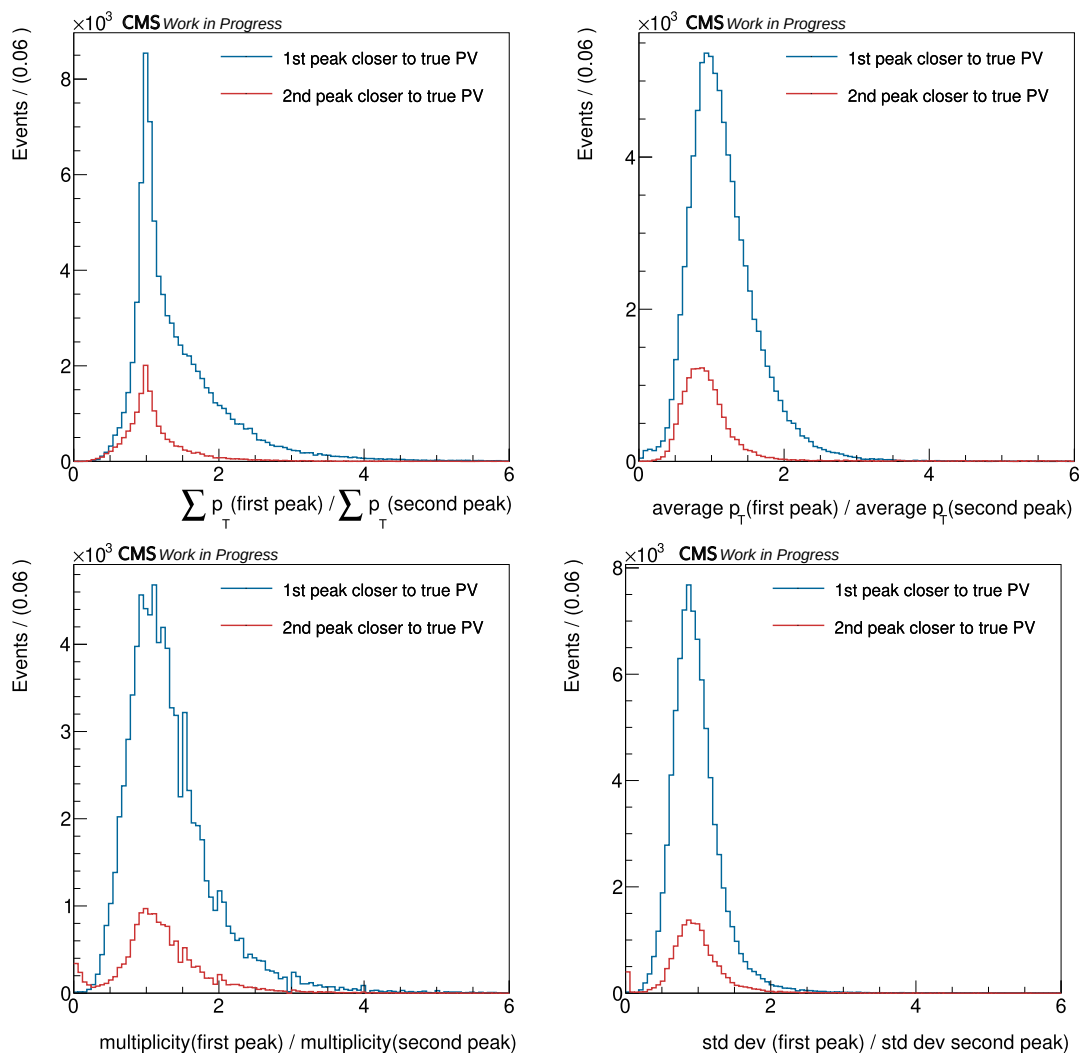


FIGURE B.9: The four plots represent the ratio between four different quantities (sum of transverse momenta (p_T), average p_T , standard deviation, and multiplicity) calculated for the first and the second peak. The blue curve showcases situations where the first peak is closer to the MC PV than the second peak, while the red curve illustrates cases where the second peak is closer than the first to the true MC primary vertex.

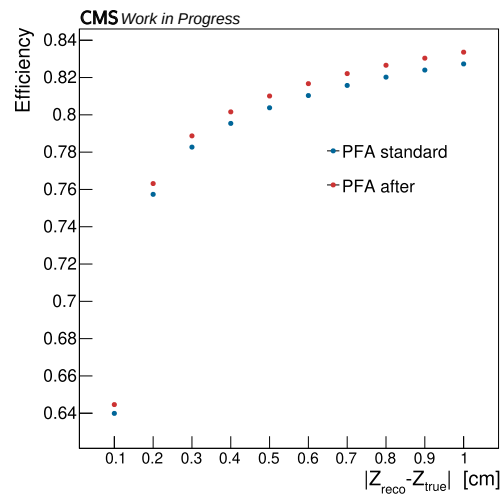


FIGURE B.10: Efficiency curve before and after implementing the described cuts and swaps described in Section 9.1.3.

Bibliography

- [1] I. Neutelings, The CMS coordinate system, accessed on 11-02-2024, https://tikz.net/axis3d_cms/.
- [2] J. Ellis, M. K. Gaillard, and D. V. Nanopoulos, “A Historical Profile of the Higgs Boson”, p. 255 (2016), 1504.07217.
- [3] V. Barger, P. Langacker, H.-S. Lee, and G. Shaughnessy, “Higgs Sector in Extensions of the MSSM”, Phys. Rev. D **73**, 115010 (2006), hep-ph/0603247.
- [4] E. Lopienska, The CERN accelerator complex, layout in 2022, accessed 11-02-2024, <https://cds.cern.ch/images/CERN-GRAPHICS-2022-001-1>.
- [5] Updated LHC Schedule, 2022, <https://hilumilhc.web.cern.ch/article/l3-schedule-change>.
- [6] CMS Collaboration, 2023 proton-proton collisions at 13.6 TeV, accessed 11-02-2024, <https://bit.ly/43GPhjr>.
- [7] S. Tai, Cutaway diagrams of CMS detector, 2019, <https://bit.ly/3VGOa1q>.
- [8] CMS Collaboration, A. Tumasyan *et al.*, “The Phase-2 Upgrade of the CMS Tracker”, 10.1088/1748-0221/13/05/P05011.
- [9] CMS Collaboration, T. CMS, “The CMS electromagnetic calorimeter project: Technical Design Report”, CERN-LHCC-97-033, CMS-TDR-4 (1997).
- [10] CMS Collaboration, A. M. Sirunyan *et al.*, “Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV”, JINST **13**, P05011 (2018), 1712.07158.
- [11] CMS Collaboration, S. Chatrchyan *et al.*, “The CMS Experiment at the CERN LHC”, JINST **3**, S08004 (2008).
- [12] CMS Collaboration, The CMS ECAL Response to Laser Light, 2019, <https://bit.ly/3IVc41w>.
- [13] T. Robens, T. Stefaniak, and J. Wittbrodt, “Two-real-scalar-singlet extension of the SM: LHC phenomenology and benchmark scenarios”, Eur. Phys. J. C **80**, 151 (2020), 1908.08554.

- [14] CMS Collaboration, A. Tumasyan *et al.*, Search for light Higgs bosons from supersymmetric cascade decays in pp collisions at $\sqrt{s} = 13$ TeV, 2023, 2204.13532.
- [15] Valente, Marco, ATLAS Object Reconstruction, 2022, 10.1007/978-3-030-94047-8_5.
- [16] CMS Collaboration, A. Zabi, J. W. Berryhill, E. Perez, and A. D. Tapper, “*The Phase-2 Upgrade of the CMS Level-1 Trigger*”, (2020), 10.1088/1748-0221/12/01/P01020.
- [17] C. Brown, *Fast Machine Learning in the CMS Level-1 Trigger for the High-Luminosity LHC*, PhD thesis, Imperial College, 2023.
- [18] I. Melzer-Pellmann and P. Pralavorio, “*Lessons for SUSY from the LHC after the first run.*”, Eur. Phys. J. C **74**, 2801 (2014) (2014).
- [19] M. Maniatis, “*The Next-to-Minimal Supersymmetric extension of the Standard Model reviewed*”, Int. J. Mod. Phys. A **25**, 3505 (2010), 0906.0777.
- [20] W. Beenakker, C. Borschensky, M. Krämer, A. Kulesza, and E. Laenen, “*NNLL-fast: predictions for coloured supersymmetric particle production at the LHC with threshold and Coulomb resummation*”, JHEP **12**, 133 (2016), 1607.07741.
- [21] Djouadi, A. and Kalinowski, J. and Spira, M., “*HDECAY: A Program for Higgs boson decays in the standard model and its supersymmetric extension*”, Comput. Phys. Commun. **108**, 56 (1998), hep-ph/9704448.
- [22] Djouadi, Abdelhak and Kalinowski, Jan and Muehlleitner, Margarete and Spira, Michael, “*HDECAY: Twenty₊₊ years after*”, Comput. Phys. Commun. **238**, 214 (2019), 1801.09506.
- [23] Particle Data Group, R. L. Workman and Others, “*Review of Particle Physics*”, PTEP **2022**, 083C01 (2022).
- [24] B. Rosenstein and A. Kovner, “*Masslessness of photon and Goldstone theorem*”, Int. J. Mod. Phys. A **6**, 3559 (1991).
- [25] ATLAS Collaboration, G. Aad *et al.*, “*Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*”, Phys. Lett. B **716**, 1 (2012), 1207.7214.
- [26] CMS Collaboration, S. Chatrchyan *et al.*, “*Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC*”, Phys. Lett. B **716**, 30 (2012), 1207.7235.
- [27] ATLAS Collaboration, G. Aad *et al.*, “*Study of the spin and parity of the Higgs boson in diboson decays with the ATLAS detector*”, Eur. Phys. J. C **75**, 476 (2015), 1506.05669, [Erratum: Eur. Phys. J. C **76**, 152 (2016)].

- [28] CMS Collaboration, E. Di Marco, “*Studies of the Higgs boson spin and parity using the $\gamma\gamma$, ZZ , and WW decay channels with the CMS detector*”, Nucl. Part. Phys. Proc. **273-275**, 746 (2016).
- [29] Super-Kamiokande Collaboration, Y. Fukuda *et al.*, “*Evidence for oscillation of atmospheric neutrinos*”, Phys. Rev. Lett. **81**, 1562 (1998), hep-ex/9807003.
- [30] OPERA Collaboration, N. Agafonova *et al.*, “*Final Results of the OPERA Experiment on ν_τ Appearance in the CNGS Neutrino Beam*”, Phys. Rev. Lett. **120**, 211801 (2018), 1804.04912, [Erratum: Phys. Rev. Lett. **121**, 139901 (2018)].
- [31] M. J. Rees, “*Dark matter: Introduction*”, Phil. Trans. Roy. Soc. Lond. A **361**, 2427 (2003), astro-ph/0402045.
- [32] J. H. Oort, “*Observational evidence confirming Lindblad’s hypothesis of a rotation of the galactic system*”, Bulletin of the Astronomical Institutes of the Netherlands **Vol. 3**, Pag. 275 (1927).
- [33] F. Zwicky, “*Die Rotverschiebung von extragalaktischen Nebeln*”, Helv. Phys. Acta **6**, 110 (1933).
- [34] V. C. Rubin and J. Ford, W. Kent, “*Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions*”, **159**, 379 (1970).
- [35] J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turlay, “*Evidence for the 2π Decay of the K_2^0 Meson*”, Phys. Rev. Lett. **13**, 138 (1964).
- [36] S. Khalil and S. Moretti, *Supersymmetry Beyond Minimality: From Theory to Experiment* (CRC Press, 2019).
- [37] S. R. Coleman and J. Mandula, “*All Possible Symmetries of the S Matrix*”, Phys. Rev. **159**, 1251 (1967).
- [38] R. Haag, J. T. Lopuszanski, and M. Sohnius, “*All Possible Generators of Supersymmetries of the s Matrix*”, Nucl. Phys. B **88**, 257 (1975).
- [39] U. Ellwanger, C. Hugonie, and A. M. Teixeira, “*The Next-to-Minimal Supersymmetric Standard Model*”, Phys. Rept. **496**, 1 (2010), 0910.1785.
- [40] W. Wang, “*A comparative study of dark matter in the MSSM and its singlet extensions: a mini review*”, Adv. High Energy Phys. **2012**, 216941 (2012), 1205.5081.
- [41] A. Djouadi *et al.*, “*Benchmark scenarios for the NMSSM*”, JHEP **07**, 002 (2008), 0801.4321.
- [42] H.-C. Cheng and D. E. Kaplan, “*Axions and a gauged Peccei-Quinn symmetry*”, (2001), hep-ph/0103346.

- [43] ATLAS Collaboration, G. Aad *et al.*, “*The ATLAS Experiment at the CERN Large Hadron Collider*”, JINST **3**, S08003 (2008).
- [44] CMS Collaboration, *The CMS magnet project: Technical Design Report* (CERN, Geneva, 1997), CERN-LHCC-97-010, CMS-TDR-1.
- [45] CMS Collaboration, V. Karimäki *et al.*, *The CMS tracker system project: Technical Design Report* (CERN, Geneva, 1998), CERN-LHCC-98-006, CMS-TDR-5.
- [46] CMS Collaboration, “*The CMS tracker: addendum to the Technical Design Report*”, (2000), CERN-LHCC-2000-016, CMS-TDR-5-add-1.
- [47] A. Dominguez *et al.*, “*CMS Technical Design Report for the Pixel Detector Upgrade*”, (2012), CERN-LHCC-2012-016, CMS-TDR-11.
- [48] C. Collaboration, “*CMS electromagnetic calorimeter performance*”, Journal of Instrumentation **5**, T03010 (2010), 10.1088/1748-0221/5/03/T03010.
- [49] CMS Collaboration, “*The CMS hadron calorimeter project: Technical Design Report*”, (1997), CERN-LHCC-97-031, CMS-TDR-2.
- [50] CMS, J. G. Layter, “*The CMS muon project: Technical Design Report*”, (1997), CERN-LHCC-97-032, CMS-TDR-3.
- [51] S. D. Ellis and D. E. Soper, “*Successive combination jet algorithm for hadron collisions*”, Phys. Rev. D **48**, 3160 (1993), hep-ph/9305266.
- [52] Y. L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber, “*Better jet clustering algorithms*”, JHEP **08**, 001 (1997), hep-ph/9707323.
- [53] M. Cacciari, G. P. Salam, and G. Soyez, “*The anti- k_t jet clustering algorithm*”, JHEP **04**, 063 (2008), 0802.1189.
- [54] CMS, “*The CMS electromagnetic calorimeter project: Technical Design Report*”, (1997), CERN-LHCC-97-033, CMS-TDR-4.
- [55] S. Bimbot *et al.*, “*Tests of the CMS-ECAL Trigger Primitive Generation*”, (2004), CMS Note: CMS-CR-2004-027.
- [56] Jenkins Handbook, 2022, <https://www.jenkins.io/doc/book/>.
- [57] A. Titterton, U. Ellwanger, H. U. Flaecher, S. Moretti, and C. H. Shepherd-Themistocleous, “*Exploring Sensitivity to NMSSM Signatures with Low Missing Transverse Energy at the LHC*”, JHEP **10**, 064 (2018), 1807.10672.
- [58] H. Qu and L. Gouskos, “*Jet tagging via particle clouds*”, Phys. Rev. D **101**, 056019 (2020).

- [59] Mei-li Holmberg, *Tracker Readout Development for the CMS Upgrade and a Search for Light Higgs Bosons from SUSY Decay Cascades*, PhD thesis, University of Bristol, 2023.
- [60] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, “Soft Drop”, JHEP **05**, 146 (2014), 1402.2657.
- [61] U. Ellwanger and A. M. Teixeira, “NMSSM with a singlino LSP: possible challenges for searches for supersymmetry at the LHC”, JHEP **10**, 113 (2014), 1406.7221.
- [62] J. Taylor, *A search for light Higgs bosons in supersymmetric decay cascades with the CMS detector*, PhD thesis, University of Bristol, 2019.
- [63] U. Ellwanger, A. Florent, and D. Zerwas, “Discovering the constrained NMSSM with tau leptons at the LHC”, JHEP **01**, 103 (2011), 1011.0931.
- [64] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer, “MadGraph 5 : Going Beyond”, JHEP **06**, 128 (2011), 1106.0522.
- [65] C. Bierlich *et al.*, “A comprehensive guide to the physics and usage of PYTHIA 8.3”, SciPost Phys. Codeb. **2022**, 8 (2022), 2203.11601.
- [66] Agostinelli, S. and others, “GEANT4—a simulation toolkit”, Nucl. Instrum. Meth. A **506**, 250 (2003), SLAC-PUB-9350, FERMILAB-PUB-03-339, CERN-IT-2002-003.
- [67] M. L. Mangano, M. Moretti, F. Piccinini, and M. Treccani, “Matching matrix elements and shower evolution for top-quark production in hadronic collisions”, JHEP **01**, 013 (2007), hep-ph/0611129.
- [68] The NanoAOD-Tools, accessed on 12-20-2023, <https://github.com/cms-nanoAOD/nanoAOD-tools>.
- [69] coffea - columnar object framework for effective analysis, accessed 2023-20-12, <https://coffeateam.github.io/coffea/>.
- [70] CMS Collaboration, “Identification of double-b quark jets in boosted event topologies”, (2016), CMS-PAS-BTV-15-002.
- [71] CMS Collaboration, “Mass regression of highly-boosted jets using graph neural networks”, (2021), CMS-DP-2021-017, CERN-CMS-DP-2021-017.
- [72] W. Buttinger, Background estimation with the abcd method, accessed 27-12-2023, https://twiki.cern.ch/twiki/pub/Main/ABCDMethod/ABCDGuide_draft18Oct18.pdf.
- [73] Lepton mva, accessed 05-01-2024, <https://bit.ly/3IXcHaM>.
- [74] G. Punzi, “Sensitivity of searches for new signals and its optimization”, eConf **C030908**, MODT002 (2003), physics/0308063.

- [75] Cut Based Muon Identification, accessed on 09-01-2024, <https://twiki.cern.ch/twiki/bin/viewauth/CMS/LeptonMVA>.
- [76] Baseline muon selections for Run-II, accessed 09-01-2024, <https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideMuonSelection>.
- [77] Tau ID Recommendation For Run 2, accessed 10-01-2024, <https://twiki.cern.ch/twiki/bin/viewauth/CMS/TauIDRecommendationForRun2>.
- [78] CMS Collaboration, Tag and Probe, <https://bit.ly/3VGM9SU>, Accessed on 05-01-2024.
- [79] CMS Collaboration, A. Hayrapetyan *et al.*, “Search for physics beyond the standard model in top quark production with additional leptons in the context of effective field theory”, JHEP **12**, 068 (2023), 2307.15761.
- [80] CMS Collaboration, “Identification of prompt and isolated muons using multivariate techniques at the CMS experiment in proton-proton collisions at $\sqrt{s} = 13$ TeV”, (2023), CMS-PAS-MUO-22-001.
- [81] CMS Collaboration, A. M. Sirunyan *et al.*, “Measurement of the Higgs boson production rate in association with top quarks in final states with electrons, muons, and hadronically decaying tau leptons at $\sqrt{s} = 13$ TeV”, Eur. Phys. J. C **81**, 378 (2021), 2011.03652.
- [82] CMS Collaboration, “Calibration of the mass-decorrelated ParticleNet tagger for boosted $b\bar{b}$ and $c\bar{c}$ jets using LHC Run 2 data”, (2022), CMS-DP-2022-005, CERN-CMS-DP-2022-005.
- [83] Wolfram Mathematica, 2022, <https://www.wolfram.com/mathematica/>.
- [84] P. D. Group, Review of particle physics: Statistics, 2020, <https://pdg.lbl.gov/2020/reviews/rpp2020-rev-statistics.pdf>, Accessed: 2024-07-04.
- [85] CMS Collaboration, M. O. Wiehe, “The CMS High Granularity Calorimeter for the High Luminosity LHC”, (2022), CMS-CR-2022-047.
- [86] CMS Collaboration, Y. Haddad, “Upgrade of the CSC Muon System for the CMS Detector at the HL-LHC”, (2022), CMS-CR-2021-182.
- [87] R. Merrick, “Getting Started With FPGAs”, (2023), No Starch Press, ISBN: 9781718502949.
- [88] Amd xilinx, ultrascale, architecture and product data sheet: Overview, <https://docs.xilinx.com/v/u/en-US/ds890-ultrascale-overview>.
- [89] T. Williams, “System design and prototyping of the CMS Level-1 Trigger at the High-Luminosity LHC”, Journal of Instrumentation **19**, C03016 (2024).