Imperial College of Science, Technology and Medicine
Department of Physics

Flavour-universal search for heavy neutral leptons with a deep neural network-based displaced jet tagger with the CMS experiment

Vilius Čepaitis

A dissertation submitted in partial fulfilment of the requirements for the degree of
Doctor of Philosophy from Imperial College London, February 2022

Copyright Declaration

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial 4.0 International Licence (CC BY-NC).

Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose.

When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes.

Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Abstract

This thesis describes a search for long-lived heavy neutral leptons using a dataset of 137 fb^{-1} collected during the 2016–2018 proton-proton runs with the CMS detector. The search uses a final state containing two leptons and at least one hadronic jet. This is the first analysis at the Large Hadron Collider which considers universal mixing between the Standard Model and heavy neutral lepton species. The search makes heavy use of a deep neural network-based displaced jet tagging algorithm, originally developed to target heavy long-lived gluino decays. The tagger was trained on both simulation and proton-proton collision data using the domain adaptation technique, which significantly improved the modelling of its output in simulation. The tagger has excellent performance for a range of long-lived particle lifetimes and generalises well to various flavours of displaced jets. In this analysis, the backgrounds are estimated in an entirely data-driven manner. No evidence for heavy neutral leptons is observed, and upper limits are set for a wide range of heavy neutral lepton mass, lifetime, and mixing scenarios. This is the most sensitive search for heavy neutral leptons in the 1–12 GeV mass range to date.

Statement of Originality

I, the author of this thesis, hereby declare the work contained in this document to be my own. All figures labelled "CMS" are sourced directly from CMS publications, including those produced by the author, and referenced in the figure caption. All figures and studies taken from external sources are referenced appropriately throughout this document.

In Chapter 4, I carried out displaced muon, electron, and jet reconstruction and identification efficiency studies.

In Chapter 6, I was involved in training the displaced jet tagging algorithm with a CMS collaborator. My primary responsibility was hyperparameter optimisation and performance studies. The result of this work was published in Ref. [1].

In Chapter 7, I contributed to all areas of the analysis strategy. My main focus area was physics object and event selection, event categorisation, and the implementation of systematic uncertainties.

In Chapter 8, I implemented the background estimation technique and was responsible for the statistical interpretation of the results.

The results presented in Chapters 7 and 8 are preliminary and currently undergoing internal review by the CMS collaboration.

Vilius Čepaitis

Acknowledgements

First of all, I would like to thank my supervisor, Alex Tapper, for his guidance and feedback throughout my PhD. I am incredibly grateful for his supervision's relaxed, informal, and pragmatic nature and his help in my academic endeavours.

I am deeply indebted to Matthias Komm for his patience, enthusiasm, and attention to detail in his mentoring. It helped me quickly get up to speed on the intricacies of CMS and grow as a physicist, which I appreciate immensely. Many thanks to all Imperial College colleagues: Rob Bainbridge, Matthew Citron, Aaron Bundock, Nick Wardle, and Mikael Mieskolainen, for the many engaging and fruitful discussions, practical advice, and for keeping me on the right track.

Thanks also to Paula Brown and Gavin Davies for helping me address the many bureaucratic challenges quickly and efficiently.

Special thanks to the LTA crew of Sergey, Shameena, Vukašin, Robert, Ryan, and others. It has been a pleasure going through this experience with you.

To Julia, thank you very much for the many beautiful moments shared throughout this PhD, and for your care and support in keeping my sanity in check throughout this pandemic. Without our collaboration, this work would not have been possible.

I would also like to acknowledge the funding from STFC supporting my studies and allowing me to spend time at CERN.

Norėčiau padėkoti savo tėvams, broliui ir seneliams už visokeriopą pagalbą studijų metu ir nesibaigiantį tikėjimą manimi. Be Jūsų paramos ši disertacija nebūtų įmanoma. Ačiū Henrikui, Martynui ir Lukui už mūsų besitesiančią draugystę, aistringas diskusijas, įtemptus žaidimus bei bendrus žygius.

Contents

Introduction **2**

1 The Standard Model of particle physics **5**

1.1 Overview of the Standard Model 5

1.2 Gauge theories 6

1.2.1 Gauge invariance and quantum electrodynamics 8

1.3 Strong interaction 10

1.4 Electroweak sector 13

1.5 Electroweak symmetry breaking 15

1.5.1 Fermion mass generation 17

2 Physics beyond the Standard Model **20**

2.1 Introduction 20

2.1.1 The nature of neutrino masses 21

2.1.2 Dark matter 23

2.1.3 Baryon asymmetry 23

2.1.4 The hierarchy problem 23

2.2 Supersymmetry 24

2.3 The seesaw mechanism 25

2.4 Long-lived particles 31

2.4.1 Introduction 31

2.4.2 Long-lived supersymmetry 33

2.4.3 Long-lived heavy neutral leptons 35

2.5 Summary 36

3 Experimental setup 38

3.1 The Large Hadron Collider 38

3.2 The CMS experiment 40

3.2.1 The inner tracking system 41

3.2.2 The electromagnetic calorimeter 43

3.2.3 The hadron calorimeter 44

3.2.4 The muon system 45

3.2.5 The data acquisition and trigger system 46

3.2.6 Data-taking performance 48

4 Event reconstruction 51

4.1 Particle flow event reconstruction 52

4.1.1 Track reconstruction 52

4.1.2 Vertex reconstruction 53

4.1.3 Cluster reconstruction 54

4.1.4 Muon tracking 55

4.2 Physics object reconstruction 55

4.2.1 Muon reconstruction 56

4.2.2 Electron reconstruction 58

4.2.3 Jet reconstruction 61

4.2.4 Missing energy reconstruction 63

5 Analysis techniques 67

5.1 Monte Carlo event generation 67

5.1.1 Parton-level event generation 67

5.1.2 Parton showering and hadronisation 69

5.1.3 Detector simulation 70

5.1.4 Background samples 70

5.2 Machine learning 72

5.2.1 Neural networks 73

5.2.2 Decision trees 81

5.2.3 Domain adaptation 83

5.3 Statistical inference 86

5.3.1 Likelihood model 86
5.3.2 Nuisance parameters 87
5.3.3 Profile likelihood ratio 87
5.3.4 Hypothesis testing 88

6 Deep neural network-based displaced jet tagger 94
6.1 Introduction 94
6.2 Signal samples 95
6.3 Definition of jet labels 96
6.3.1 Ghost labelling 97
6.3.2 Displaced jet labelling 99
6.4 Training setup 101
6.4.1 Input features 101
6.4.2 Network architecture 104
6.4.3 Parametric network 105
6.4.4 Control regions 107
6.4.5 Training samples 107
6.4.6 Technical implementation 108
6.4.7 Hyperparameter optimisation 109
6.5 Tagger performance 112
6.5.1 Inclusive performance 112
6.5.2 Performance as a function of jet kinematic quantities 114
6.5.3 Parameterisation performance 114
6.5.4 Performance versus lifetime 115
6.5.5 Wrong lifetime test 115
6.6 Tagger validation 116
6.7 Summary 118

7 Search for heavy neutral leptons 120
7.1 Heavy neutral lepton signatures at CMS 120
7.2 Simulated signal samples 122
7.3 Event selection and categorisation 127
7.3.1 Trigger strategy 128

7.3.2 Physics objects 129
7.3.3 Preselection 135
7.3.4 Event categorisation 135
7.3.5 Corrections 137
7.4 Event-level boosted decision tree 139
7.4.1 Input features 139
7.4.2 Training setup 139
7.4.3 Performance 140
7.4.4 Validation 141
7.5 Displaced jet tagger 142
7.5.1 Displaced jet subclasses 142
7.5.2 Background jet classes 144
7.5.3 Tagger architecture 144
7.5.4 Displacement parameterisation 144
7.5.5 Tagger profiling 145
7.5.6 Performance and validation 147
7.6 Summary 147

8 Background estimation and results 151
8.1 Background estimation 151
8.1.1 The ABCD technique 151
8.1.2 Candidate variable studies 152
8.1.3 Simulation checks 153
8.1.4 Threshold optimisation 156
8.1.5 Closure tests 157
8.2 Systematic uncertainties 162
8.2.1 Signal systematic uncertainties 162
8.2.2 Background systematic uncertainties 163
8.3 Results 164
8.3.1 Interpretation 164
8.3.2 Mass-coupling plane limits 164
8.3.3 Flavour-generic limits 164
8.3.4 Discussion 166
8.4 Summary 168

Conclusions **171**

Abbreviations **173**

Bibliography **179**

A Control region validation plots **194**

B Extended displaced jet tagger inputs **202**

List of Tables

1.1 The three generations of leptons of the SM. The charged lepton masses are obtained to very high precision from direct measurements [3]. The neutrino masses are known to be non-zero but orders of magnitude smaller than the quark and charged lepton masses. 6

1.2 The three quark generations of the SM. The quark masses are reported in the $\overline{\text{MS}}$-scheme, except for the top quark mass, which corresponds to the pole mass [3]. 6

1.3 The force mediators, spin-1 gauge bosons of the SM. The EW gauge boson masses are obtained from Ref. [3]. 6

4.1 Cut-based loose electron identification criteria. 60

5.1 Background samples for statistical inference. 71

5.2 Background samples for MVA discriminant training. 72

6.1 Charged PF candidate features. 102

6.2 Neutral PF candidate features. 103

6.3 SV features. 103

6.4 Global jet features. 104

6.5 Overview of selection requirements in data CRs for training and validating DA. 107

7.1 Benchmark HNL coupling scenarios given in barycentric coordinates. . . 124

7.2 BRs for the tree-level decay modes of a Dirac HNL at various masses as calculated by MadGraph. The BRs are identical for Majorana HNLs if the charge-conjugated decay modes are considered in addition. 125

7.3 Transverse momentum thresholds for single, isolated lepton triggers used in the analysis for the three years of data taking (2016–2018). 128

7.4 Overview of selection requirements in SR and CR for the dilepton categories. The prompt lepton requirements are shown per year (2016, 2017, 2018) of data-taking. 136

7.5 Cumulative efficiency of various object- and event-level preselection requirements shown for a benchmark signal model of $m_\mathrm{N} = 10\,\mathrm{GeV}$, $c\tau_0 = 1\,\mathrm{mm}$, $V_\mathrm{e} = V_\mu = V_\tau$ and the major SM background (W+jets, Z/γ^*+jets) processes. 136

7.6 BDT input features. 139

8.1 The MI per category between the jet tagger score P and different candidate variables per category. 155

8.2 Signal samples used to optimise the tagger and BDT thresholds. For each model, coupling scenarios summarised in Table 7.1 were considered in the optimisation. 157

8.3 Best-fit values κ and ratios of predicted to observed yields for a combined closure test performed per year of data taking and event topology (boosted, resolved). In addition, the ratio results are shown when varying the lower DNN tagger threshold P_low by ±10% from its nominal value. 158

8.4 Impacts of systematic uncertainties on the signal cross-section. For the background prediction sideband yield impacts, an average over 48 categories is provided. 169

B.1 Global jet features. 203

B.2 Features used to train the CSV b-tagging algorithm. These are added to the global jet features. 203

B.3 Features of neutral PF candidates. 204

B.4 Features of secondary vertices. 204

B.5 Features of charged PF features. 205

B.6 Features of PF muons. 206

B.7 Features of PF electrons (part 1). 207

B.8 Features of PF electrons (part 2). 208

List of Figures

1.1 Summary of measurements of α_s as a function of the energy scale Q. Figure taken from Ref. [3]. 11

1.2 Proton PDFs in the NNPDF3.0 set. The valence (u_v, d_v) quark, sea quark, and gluon (scaled down by a factor of 10) distributions are shown at the energy scale of $Q^2 = 10^2$ GeV (left) and $Q^2 = 10^4$ GeV (right). 13

2.1 Mass limits for a simplified model of gluino pair production and decay to pairs of light quarks and the LSP. Figure taken from Ref. [37] 25

2.2 Exclusion limits obtained from a range of measurements on the mixing between active neutrinos (electron, muon, tau) and a single HNL in the mass range of 0.1–500 GeV. Figures taken from Ref. [46]. 27

2.3 Allowed parameter space as a fraction of the HNL coupling to a given active neutrino generation (in percentage) for different HNL ULs for normal (left) and inverted (right) active neutrino mass hierarchy. The allowed parameter space, resembling an ellipse, is delineated by solid (dashed) lines marking the 1σ (2σ) contours. Figure taken from Ref. [57]. 30

2.4 Selection of elementary and composite SM particle masses and lifetimes. The shaded regions indicate regions where the particles are either prompt or stable in a general-purpose detector like CMS. Figure taken from Ref. [58]. 31

2.5 Pair production of long-lived gluinos and their subsequent decay via a squark to a pair of light-flavour quarks and the neutralino LSP. 33

2.6 The 95% CL observed exclusion limit on the pair production cross-section of a long-lived gluino in a split SUSY model. The gluino $\tilde{g}$ and neutralino $\tilde{\chi}_1^0$ masses are assumed to be 2400 GeV and 100 GeV, respectively and the gluino is treated as a neutral, non-interacting particle [63]. 34

2.7 Contour lines indicate the HNL decay length $c\tau_0$ in the mass-coupling plane. Shaded areas correspond approximately to HNL scenarios excluded by various experiments. Figure taken from Ref. [69]. 35

3.1 A schematic representation of the CERN accelerator complex. Image credit to CERN. 39

3.2 A schematic of the CMS detector, with the physical dimensions, weight, and the number of components in various subdetectors specified. Image credit to CERN. 41

3.3 A schematic cross-section of the CMS tracker in the *rz* plane. Strip tracker modules that provide 2-D hits are shown by thin, black lines, while those permitting the reconstruction of hit positions in 3-D are shown by thick, blue lines. The pixel modules, shown by the red lines, also provide 3-D hits. Within a given layer, each module is shifted slightly in r or z with respect to its neighbouring modules, which allows them to overlap, thereby avoiding gaps in the acceptance. Figure taken from Ref. [75]. . . 43

3.4 A schematic layout of the CMS ECAL, presenting the arrangement of crystal modules, supermodules, endcaps and the ES in front (left). Geometric view of one quarter of the ECAL (right). 44

3.5 A schematic representation of the CMS HCAL. Figure taken from Ref. [80]. 46

3.6 A quadrant of the CMS detector. The IP is at the lower-left corner. The locations of the various muon stations and the steel flux-return disks (dark areas) are shown. The DTs stations are labelled MB and the CSCs are marked as ME. RPCs are mounted in both the barrel and endcaps of CMS, labelled RB and RE, respectively. Figure taken from Ref. [82]. . . . 47

3.7 Run 2 pp collision data (2015–2018) delivered by the LHC and recorded by CMS. 49

4.1 Track reconstruction efficiency for each of the twelve tracking iterations (denoted in the legend) as a function of simulated track p_T (left) and production vertex radius (right). Figures taken from Ref. [91]. 53

4.2 Efficiency for the tight PF muon isolation and identification working point versus p_T (left) and versus η for muons with $p_T > 20$ GeV (right) for the Run 2 data-taking conditions. Figure taken from Ref. [82]. 57

4.3 Muon from HNL decay reconstruction and identification efficiency as a function of generator-level lab frame displacement (left) and generator-level p_T (right). 58

4.4 Electron identification efficiency measured in data (upper panels) and data-to- simulation efficiency ratios (lower panels), as a function of the electron energy, for the cut-based identification veto working point (left) and the BDT-based (without isolation) loosest working point (right). Figure taken from Ref. [94]. 59

4.5 Electron from HNL decay reconstruction and identification efficiency as a function of generator-level lab frame displacement (left) and generator-level p_T (right). 60

4.6 JEC uncertainty breakdown for anti-k_t, $R = 0.4$ charged hadron-subtracted central jets as a function of p_T (left) and for jets with $p_T = 30\,\text{GeV}$ as a function of η (right). Figure taken from Ref. [98]. 63

4.7 Jet reconstruction and identification efficiency as a function of generator-level lab frame displacement (left) and generator-level p_T (right). 63

4.8 Average charged and neutral PF candidate energy fraction as a function of jet generator-level lab frame displacement. 64

4.9 Distributions of p_T^{miss} in $Z \longrightarrow \mu^+\mu^-$ (left) and $Z \longrightarrow e^+e^-$ (right) samples. Figure taken from Ref. [99]. 64

5.1 Schematic representation of event simulation using MC generators. The incoming partons are shown as blue lines. Stages of the simulation are displayed: the hard process (red circle), PS (red gluon lines), hadronisation (light green circles), the UE interactions (purple circle), and hadron decays (in dark green). Figure taken from Ref. [104]. 69

5.2 An example of a simple NN model that could solve binary classification problems. The network has three input features, one hidden layer, and outputs probabilities for the two classes. The activation function of the hidden layer is not shown explicitly. 75

5.3 Commonly used activation functions (a) sigmoid $\sigma(x) = 1/(1+e^{-x})$; (b) hyperbolic tangent $\sigma(x) = (e^x - e^{-x})/(e^x + e^{-x})$; (c) ReLU $\sigma(x) = \max(0, x)$ [120]; (d) LeakyReLU $\sigma(x) = \max(\alpha x, x)$ where $\alpha = 0.1$ is the leakage factor. 80

5.4 The progression of the training and validation data set loss as the NN is trained using backpropagation and SGD. At some point, the validation loss starts to increase again, which signals that the network is overfitting the training data set. 81

5.5 A sketch of several decision trees combined into a BDT. 83

5.6 Sketch of a NN architecture to perform DA by backpropagation. The feature extractor (blue) and a label predictor (red) together form a standard feed-forward architecture and predict class labels. A domain classifier (green) is connected to the feature extractor via a GRL that multiplies the gradient by a certain negative constant during the training via backpropagation. 85

5.7 The CL_s criterion construction in case the signal+background and background hypotheses, denoted as $f\left(\tilde{q}_\mu \mid 1\right)$ and $f\left(\tilde{q}_\mu \mid 0\right)$, respectively, for well separated (left) and strongly overlapping (right) scenarios. 91

6.1 Split SUSY samples used for training the DNN jet tagger algorithm. The number of generated events as a function of the gluino and $\tilde{\chi}_1^0$ mass is shown for lifetimes of $c\tau_0 = 1\,\text{mm}$ (left) and $1\,\text{m}$ (right). 96

6.2 Example of a decay of a long-lived gluino R-hadron (in red) to a pair of d quarks (in green) and the $\tilde{\chi}_1^0$ (also in red) is shown. Several additional partons (in green) result from hard gluon radiation before hadronisation. The R-hadron or its daughters might not follow the direction of the final particles and thus cannot be reliably used for ghost labelling. The particle p_{T} is shown in brackets. 98

6.3 Two examples of long-lived gluino decay chains, constructed from information provided by MC event generators. The positions of various particles in the $\eta\phi$ plane are shown: the LLP gluino and its daughter particles are shown in the lower and middle planes, respectively; the upper plane depicts the location of the stable particles after hadronisation, with shaded ellipses overlaid to indicate the reconstructed jets. Each quark and its decay is assigned a unique colour. The dotted lines indicate the links between parent and daughter particles. Figure taken from Ref. [1]. . 99

6.4 Various tagger input features for SM background jets and signal split SUSY jets for $c\tau_0$ values of 1 mm and 1 m. 102

6.5 An overview of the DNN architecture, which comprises convolutional and dense layers; the numbers of filters and nodes are indicated. Dropout layers and activation functions are not shown. The input features are grouped by object type; ($m \times n$) indicates the maximum number of objects (m) and the number of features per object (n). The gradients of the class and domain losses with respect to the weights $\boldsymbol{\omega}$ used during backpropagation are shown. Figure taken from Ref. [1]. 105

6.6 Training jet (a) transverse momentum, (b) pseudorapidity, and (c) proper lifetime distributions for signal and background classes after kinematic resampling. 108

6.7 An example of probability distributions in a validation data set for the five jet classes obtained after training the NN. The $c\tau_0$ values used to evaluate the probability were generated in the same way as for the training data set. 110

6.8 Hyperparameter optimisation. In (a), the optimisation of the learning rate decay parameter η_0 with no DA is shown. A value of $\eta_0 = 0.1$ is chosen as the best performing scenario. In (b), DA hyperparameter optimisation is shown. The mean signal efficiency and the mean JSD score is investigated as a function of the hyperparameter λ_0. The value of $\lambda_0 = 30$ is chosen as optimal. 111

6.9 The evolution of class and domain validation loss during training is shown with (a) $\lambda_0 = 30$ and (b) $\lambda_0 = 60$. 111

6.10 ROC curves showing the tagger performance for uncompressed split SUSY, GMSB and RPV SUSY scenarios: (a) LLP jets vs light-flavour jets for $c\tau_0 = 1$ mm; (b) LLP jets vs light-flavour jets for $c\tau_0 = 1$ m; (c) LLP jets vs c jets for $c\tau_0 = 1$ mm; (d) LLP jets vs b jets for $c\tau_0 = 1$ mm. For the uncompressed scenario, a reference no DA training is also included for comparison. Figures (a) and (b) taken from Ref. [1]. 113

6.11 ε(LLP) as a function of the jet p_{T}, η, and N_{SV}. The efficiency curves are shown separately for the split (circular marker), GMSB (triangle marker), and RPV (square marker) SUSY benchmark models, assuming $c\tau_0$ values of 1 m (upper row) and 1 mm (lower row). Figure taken from Ref. [1]. . . . 114

6.12 In (a), ε(LLP) as a function of the split SUSY $c\tau_0$ for compressed and uncompressed split SUSY scenarios is shown. In (b), ε(LLP) versus $c\tau_0$ for a shallow network, trained only on global jet features, is shown. Figure (a) taken from Ref. [1]. 115

6.13 The DNN tagger efficiency ε(LLP) as a function of the lifetime used to evaluate the network. The purple (solid filling) and green (hollow filling) lines correspond to uncompressed split SUSY scenarios generated assuming a gluino lifetime of 1 m and 1 mm, respectively. Figure taken from Ref. [1]. 116

6.14 Distributions of the maximum probability for the LLP jet class obtained from all selected jets in each event, $P_{\mathrm{max}}(\mathrm{LLP}|c\tau_0)$. The distributions from data (circular marker) and simulated events (histograms) are compared in the μ+jets (upper row) and $\mu\mu$+jets (lower row) CRs, using a DNN trained without (left column) and with (right column) DA. All probabilities are evaluated with $c\tau_0$ = 1 mm. The statistical (hatched bands) and systematic (solid bands) uncertainties due to the finite-size simulation samples and the simulation mismodelling of the mistag rate, respectively, are shown. Figures taken from Ref. [1]. 117

7.1 CC HNL production in association with a prompt trigger lepton ℓ_1 and decay to a displaced lepton (ℓ_2) and a pair of quarks. In (a), an OS dilepton final state is shown, possible for both Dirac and Majorana HNLs. In (b), a SS dilepton final state is shown, possible only for Majorana HNLs. The flavours of $\ell_{1,2}$ can differ. 121

7.2 The considered signal scenarios with $|V_{\mathrm{e}}| = |V_{\mu}| = |V_{\tau}|$ in the mass-coupling plane for Majorana (left) and Dirac (right) HNLs. Scenarios with the same proper lifetime, $c\tau_0$, as calculated by MadGraph, are colour-coded and connected by dashed lines. 123

7.3 The considered 66 relative couplings points in barycentric coordinates ($f_{\mathrm{e}}+f_{\mu}+f_{\tau} = 1$) for reweighting each signal sample that has been generated with $|V_{\mathrm{e}}| = |V_{\mu}| = |V_{\tau}|$, i.e. $f_{\mathrm{e}} = f_{\mu} = f_{\tau} = 1/3$. 124

7.4 (a) HNL transverse momentum; (b) HNL pseudorapidity; (c) HNL boost; (d) HNL transverse displacement; (e) ℓ_1 transverse momentum; (f) ℓ_1 pseudorapidity; (g) ℓ_2 transverse momentum; (h) ℓ_2 pseudorapidity for various benchmark models and coupling scenarios. For the pure-τ coupling, ℓ_2 corresponds to the leptonic tau daughter. 126

7.5 The distributions of $p_T(\ell_1)$ (top row), $p_T(\ell_2)$ (middle row), and $m(\ell_1, \ell_2)$ (bottom row), shown for benchmark OS $\mu\mu$ (left column) and SS $e\mu$ (right column) final states for the major simulated SM backgrounds after the preselection. A benchmark HNL scenario $m_N = 10\,\text{GeV}, c\tau_0 = 1\,\text{mm}, V_e = V_\mu$ is overlaid. 130

7.6 The distributions of (a) p_T; (b) $\Delta R(j^\star, \ell_2)$ of reconstructed displaced jets from HNL decays. The jets have the ℓ_2 from HNL decay boosted inside the jet or are resolved. 131

7.7 The distributions of the number of selected jets (top row), $p_T(j^\star)$ (second row), the angular separation $\Delta R(\ell_2, j^\star)$ (third row), and $m(\ell_1, \ell_2, j^\star)$ (bottom row) shown for benchmark OS $\mu\mu$ (left column) and SS $e\mu$ (right column) categories for the major simulated SM backgrounds after the preselection. A benchmark HNL scenario $m_N = 10\,\text{GeV}, c\tau_0 = 1\,\text{mm}, V_e = V_\mu$ is overlaid. 133

7.8 The distributions of p_T^{miss} shown for benchmark OS $\mu\mu$ (left column) and SS $e\mu$ (right column) final states for the major simulated SM backgrounds. A benchmark HNL scenario $m_N = 10\,\text{GeV}, c\tau_0 = 1\,\text{mm}, V_e = V_\mu$ is overlaid. 134

7.9 The distributions of d_{xy}^{sig} shown for benchmark OS $\mu\mu$ boosted (left) and SS $e\mu$ resolved (right) categories for the major simulated SM backgrounds in the SR. Two benchmark HNL scenarios, $m_N = 10\,\text{GeV}, c\tau_0 = 1\,\text{mm}$, and $m_N = 2\,\text{GeV}, c\tau_0 = 1\,\text{m}$, are overlaid. 138

7.10 ROC curve showing the BDT performance for the 2016 scenario. The signal efficiency for three benchmark HNL scenarios, indicated in the legend, is compared against the expected SM backgrounds in the SR. The efficiency corresponding to a working point threshold of 0.4 is also indicated by dashed lines. Additional reference ROC curves obtained by training a logistic regression discriminant are also shown as dotted lines. 140

7.11 Data/MC comparison of the BDT signal class probability for OS $\mu\mu$ (top) and $e\mu$ (bottom) categories, broken down by boosted (left) and resolved (right) topology, shown for simulation and CR data collected in 2016. . . 141

7.12 Displaced jet subclass fractions for 3 GeV (left) 10 GeV (right) HNL models, assuming equal couplings to all three active neutrino generations. 143

7.13 Probability matrix of reconstructing a truth-level quark-lepton or quark HNL jet in the boosted or resolved categories. 143

7.14 The architecture of the NN for identifying displaced jets from the HNL decay. Compared with Figure 6.5, additional PF muon and electron input features have been introduced. In addition, the tagger is parameterised as a function of the jet displaced instead of the LLP $c\tau_0$. 145

7.15 True vs profiled displacement for signal (a) quark-muon; (b) quark-electron; (c) quark jet classes. 146

7.16 Tagger ROC curves for signal (a) quark-muon; (b) quark-electron; (c) quark jets versus the equivalent SM background jet class obtained from a W+jets sample, measured for an HNL with $m_{\mathrm{N}} = 10\,\mathrm{GeV}$ with a range of proper lifetimes. 148

7.17 Expected yields in simulation in the SR (left, blinded) and the data/MC comparison in the CR (right) for the 2016 data-taking scenario of the profiled ratio likelihood (a) $P_{q\mu}$ in OS $\mu\mu$ boosted events; (b) P_{qe} in OS ee boosted events; (c) P_q in OS $\mu\mu$ resolved events; (d) P_q in SS $e\mu$ resolved events. 149

8.1 A sketch of the ABCD region definitions. The region populated below P_{low} is discarded to ensure only signal-like jets are selected. 153

8.2 Two-dimensional histograms (heatmaps) of P and the BDT score for simulated background events in the SR for boosted (top-row) and resolved (bottom row) categories, OS $\mu\mu$ (left) and SS $e\mu$ (right) final states. The MI and the Pearson correlation coefficient between the two variables are indicated in the plots. 154

8.3 Background composition by $\mathrm{j}^{\star}$ truth label for (a) boosted and (b) resolved categories, broken down by dilepton flavour and sign final states. 155

8.4 Distributions of $P_{\mathrm{q}\mu}$ in low- and high-BDT score regions for the dimuon OS prompt boosted category, shown with differing P_{min} thresholds. . . . 156

8.5 Closure test in VR1 ($m(\ell_1, \ell_2, \mathrm{j}^{\star}) < 70\,\mathrm{GeV}$) for boosted (left) and resolved (right) categories for 2016 (top row), 2017 (middle row), and 2018 (bottom row) data-taking scenarios. 159

8.6 Closure test in VR2 (sideband AB) for boosted (left) and resolved (right) categories for 2016 (top row), 2017 (middle row), and 2018 (bottom row) data-taking scenarios. 160

8.7 The predicted saturated likelihood test statistic distribution and the observed value (in blue) for VR1 (left) and (b) VR2 (right) GOF test. 161

8.8 Observed and predicted yields in the SR for boosted (left) and resolved (right) categories for 2016 (top row), 2017 (middle row), and 2018 (bottom row) data-taking scenarios. 165

8.9 The predicted saturated likelihood test statistic distribution and the observed value (in blue) for the SR GOF test. 166

8.10 Expected and observed limits for Dirac and Majorana HNLs for different benchmark coupling scenarios (listed in Table 7.1). 167

8.11 Maximum excluded mass for each flavour combination for Dirac HNLs with (a) $c\tau_0 = 1\,\mathrm{mm}$ and (b) $c\tau_0 = 10\,\mathrm{mm}$. 168

A.1 Distribution of $\mathrm{j}^{\star}$ p_{T} in the CR. The distributions are shown for 2016, 2017, and 2018 scenarios in the left, middle, and right columns, respectively. The distributions are shown for the OSSF categories in the first two rows, and the SS μe and $e\mu$ categories in the bottom two rows. 195

A.2 Distribution of the ℓ_1 p_{T} in the CR. The distributions are shown for 2016, 2017, and 2018 scenarios in the left, middle, and right columns, respectively. The distributions are shown for the OSSF categories in the first two rows, and the SS μe and $e\mu$ categories in the bottom two rows. . 196

A.3 Distribution of the ℓ_2 p_{T} in the CR. The distributions are shown for 2016, 2017, and 2018 scenarios in the left, middle, and right columns, respectively. The distributions are shown for the OSSF categories in the first two rows, and the SS μe and $e\mu$ categories in the bottom two rows. . 197

A.4 Distribution of the $p_{\mathrm{T}}^{\mathrm{miss}}$ in the CR. The distributions are shown for 2016, 2017, and 2018 scenarios in the left, middle, and right columns, respectively. The distributions are shown for the OSSF categories in the first two rows, and the SS μe and $e\mu$ categories in the bottom two rows. . 198

A.5 Modelling of the tagger output in the CR, resolved scenario. The distributions are shown for 2016, 2017, and 2018 scenarios in the left, middle, and right columns, respectively. The distributions are shown for the OS dimuon category in the first row, and the SS $e\mu$ category in the bottom row. 199

A.6 Modelling of the tagger output in the CR, boosted muon scenario. The distributions are shown for 2016, 2017, and 2018 scenarios in the left, middle, and right columns, respectively. The distributions are shown for the OS dimuon category in the first row, and the SS $e\mu$ category in the bottom row. 199

A.7 Modelling of the tagger output in the CR, boosted electron scenario. The distributions are shown for 2016, 2017, and 2018 scenarios in the left, middle, and right columns, respectively. The distributions are shown for the OS dielectron category in the first row, and the SS μe category in the bottom row. 200

A.8 Distribution of the BDT score in the CR. The distributions are shown for 2016, 2017, and 2018 scenarios in the left, middle, and right columns, respectively. The distributions are shown for the OSSF categories in the first two rows, and the SS μe and $e\mu$ in the bottom two rows. 201

Introduction

Several models of physics beyond the Standard Model (SM) predict the existence of new long-lived particles (LLPs). Hadron collider experiments, such as CMS, were mostly designed to search for particles that decay promptly after being produced during hadron collisions. A substantial effort has been invested at the Large Hadron Collider (LHC) for such searches; however, no new particle has been observed yet. On the other hand, a large amount of phase space remains uncovered when considering various models of LLPs. Searches for LLPs are difficult because experiments like CMS were not explicitly designed to detect signatures of significantly displaced decays.

Machine learning (ML) methods are excellent tools to bridge the performance gap of new LLP searches, considering the recent advances in deep learning (DL). This thesis presents a search for long-lived heavy neutral leptons (HNLs) decaying to displaced jets and leptons. The search makes use of a novel deep neural network (DNN) displaced jet tagging algorithm. This is the first search at the LHC of HNLs with arbitrary coupling to all three SM lepton generations simultaneously.

The layout of this thesis is as follows: first, an introduction to the SM of particle physics is provided in Chapter 1. Several unresolved questions within the SM are addressed in Chapter 2. Possible resolutions of these questions require extensions to the SM, and two such extensions, called supersymmetry (SUSY) and HNLs, are considered within this thesis. Crucially, both theories can give rise to LLPs.

An overview of the LHC and the CMS experiment with its subdetectors is given in Chapter 3. The reconstruction of particles, based on the signatures they leave in the CMS detector, called Particle Flow (PF), is the subject of Chapter 4. Reconstruction algorithms, that ensure sufficient efficiency of reconstructing displaced particles, are of crucial importance for LLP searches, and are also discussed in this Chapter.

Chapter 5 covers different techniques and methodologies relevant for the data analysis performed within this thesis. This includes the generation of the simulated signal and background events; the foundations of various ML methods such as DNNs and boosted decision trees (BDTs) used extensively within this analysis; the statistical framework for hypothesis testing used to set upper limits (ULs).

Decays of LLPs to displaced jets are often the best detection channels due to typically high branching ratios (BRs). A displaced jet tagger was initially developed as a proof-of-concept algorithm to search for energetic displaced jets originating from the decays of heavy long-lived SUSY gluinos [1]. A crucial development included domain adaptation (DA), a transfer learning technique that facilitates the neural network (NN) training directly on data by penalising the use of mismodelled NN input features. The development of the algorithm is described in Chapter 6.

A search for long-lived HNLs, performed using a proton-proton (pp) collision data set of 137 fb^{-1} collected in 2016–2018 with the CMS experiment, is described in Chapter 7. A distinguishing feature of this search is that it considers HNL couplings to all three lepton generations simultaneously. For this, the search heavily relies on using the displaced jet tagging algorithm. Several adjustments have since been made to the tagging algorithm to accommodate the decays of boosted, low-mass LLPs, such as long-lived HNLs.

The backgrounds are estimated using a data-driven technique which is the subject of Chapter 8. Based on the predicted signal and background yields, and the observed counts in data, ULs are set on various HNL scenarios.

Chapter 1

The Standard Model of particle physics

In this Chapter, a brief introduction to the SM of particle physics is provided. An overview of the fundamental particles and the relationship between these particles and the forces is given. Natural units with $\hbar = c = 1$ are used throughout this thesis, unless stated otherwise.

1.1 Overview of the Standard Model

The review presented here is based on Ref. [2]. The SM is a theory describing all known interactions and classifying all known elementary particles. Each interaction is characterised by a quantum field theory (QFT), whose excitations correspond to force mediators. The SM only includes fundamental particles, i.e. particles that have not been observed to have an internal structure. These can be divided based on spin: fermions have half-integer spin, while bosons have integer spin. The fermions, which always have spin-1/2 in the SM, are further subdivided into quarks and leptons. The quarks participate in the strong interaction, while the leptons do not.

Almost all the ordinary matter comprises only several particles: the electron, the electron neutrino, the up quark, and the down quark. Together, these four particles make up the first fermionic generation of the SM. The SM contains three generations, where the second and third generations are almost exact copies of the first one, the only difference being in the particle mass. In contrast, the other quantum numbers remain the same. The properties of the leptons are described in Table 1.1, while the quarks are listed in Table 1.2. The interactions between particles are mediated by spin-1 bosons, summarised in Table 1.3.

The gauge bosons acquire their mass via the Higgs mechanism, which additionally implies the presence of a spin-0 Higgs boson with a mass of $m_{\mathrm{H}} = (125.10 \pm 0.14)\,\mathrm{GeV}$ [3].

Table 1.1: The three generations of leptons of the SM. The charged lepton masses are obtained to very high precision from direct measurements [3]. The neutrino masses are known to be non-zero but orders of magnitude smaller than the quark and charged lepton masses.

	Leptons		
Generation	Particle	Mass	Charge
First	e^-	511 keV	-1
	ν_e	~ 0	0
Second	μ^-	106 MeV	-1
	ν_μ	~ 0	0
Third	τ^-	1.78 GeV	-1
	ν_τ	~ 0	0

Table 1.2: The three quark generations of the SM. The quark masses are reported in the $\overline{\mathrm{MS}}$-scheme, except for the top quark mass, which corresponds to the pole mass [3].

	Quarks		
Generation	Particle	Mass	Charge
First	u	$2.16^{+0.49}_{-0.26}$ MeV	+2/3
	d	$4.67^{+0.48}_{-0.17}$ MeV	−1/3
Second	c	1.27 ± 0.02 GeV	+2/3
	s	93^{+11}_{-5} MeV	−1/3
Third	t	172.76 ± 0.30 GeV	+2/3
	b	$4.18^{+0.03}_{-0.02}$ GeV	−1/3

Table 1.3: The force mediators, spin-1 gauge bosons of the SM. The electroweak (EW) gauge boson masses are obtained from Ref. [3].

Force	Particle	Mass	Charge
Electromagnetic	γ	0	0
Weak	$\mathrm{W}^\pm$	80.4 ± 0.012 GeV	±1
	Z	91.1876 ± 0.0021 GeV	0
Strong	g	0	0

1.2 Gauge theories

In QFTs, particles are described as excitations of fields, $\phi(x) = \phi(t, \vec{x})$, which are continuous functions of the space-time coordinates. The Lagrangian density is a quantity that describes the dynamics of the physical system in terms of the field and its spacetime derivative, i.e. $\mathscr{L} = \mathscr{L}(\phi, \partial_\mu \phi)$. The fields follow the principle of least action, which states that the action, defined as:

$$S = \iint \mathscr{L}(\phi, \partial_\mu \phi) \mathrm{d}^3 \vec{x} \, \mathrm{d}t = \int \mathscr{L}(\phi, \partial_\mu \phi) \mathrm{d}^4 x, \tag{1.1}$$

is minimised, which gives rise to the Euler-Lagrange equation for the fields:

$$\partial_\mu \left(\frac{\partial \mathscr{L}}{\partial \left(\partial_\mu \phi \right)} \right) - \frac{\partial \mathscr{L}}{\partial \phi} = 0. \tag{1.2}$$

The quantitative description of particles and their interactions (forces) comes from the perturbative QFT approach, where the Lagrangian is decomposed into a free field and an interaction part:

$$\mathscr{L} = \mathscr{L}_0 + \mathscr{L}_\mathrm{I}. \tag{1.3}$$

The free field $\mathscr{L}_0$ describes freely propagating particles, whereas the interaction terms $\mathscr{L}_I$ describe processes involving several particles. In the interaction picture of quantum mechanics, the free-field model can be solved exactly, and solutions to the complete model are expressed as perturbations of the free-field solutions. Each interaction can be visually represented as Feynman diagrams, obtained using Feynman rules from $\mathscr{L}_I$. The calculation allows to determine a desired physical quantity a to order n as a power series:

$$a = \underbrace{\alpha a_1}_{\mathrm{LO}} + \underbrace{\alpha^2 a_2}_{\mathrm{NLO}} + \cdots + \underbrace{\alpha^n a_n}_{\mathrm{N^nLO}}, \tag{1.4}$$

where α denotes the strength of the interaction. Given $\alpha < 1$, the series is convergent. Practically, the calculation in perturbation theory is cut-off at some fixed order, which describes its accuracy: leading-order (LO), next-to-leading-order (NLO), etc.. The number of Feynman diagrams used to construct the matrix element (ME) increases factorially with the number of particles involved in the interaction.

As an example of a basic QFT, let us consider the Lagrangian density for a free non-interacting scalar field:

$$\mathscr{L}_\mathrm{S}^0 = \frac{1}{2} (\partial_\mu \phi)(\partial^\mu \phi) - \frac{1}{2} m^2 \phi^2. \tag{1.5}$$

Substituting this Lagrangian density into the Euler-Lagrange equation returns the Klein-Gordon equation for a free scalar field:

$$\partial_\mu(\partial^\mu\phi) + m^2\phi = 0. \tag{1.6}$$

The excitations of the field are spin-0 scalar particles of mass m. In the SM, the Klein-Gordon equation describes the free-field part of the Higgs boson. For free relativistic spin-half fermion fields, the Lagrangian density is:

$$\mathscr{L}_D = i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi, \tag{1.7}$$

where ψ is a four-component complex spinor, with the four degrees of freedom corresponding to particle, antiparticle, and left- and right-handed chiral states; γ^μ are the four Dirac (gamma) matrices. In this case, substituting the Lagrangian density into the Euler-Lagrange equation gives the famous Dirac equation:

$$i\gamma^\mu\left(\partial_\mu\psi\right) - m\psi = 0. \tag{1.8}$$

1.2.1 Gauge invariance and quantum electrodynamics

The Lagrangians of the SM are required to satisfy a principle called local gauge invariance. In other words, the Lagrangian must be invariant under a local phase transformation of the fields. Curiously, this requirement induces interaction terms in the system. For example, the theory of electromagnetism can be obtained by imposing a local U(1) symmetry of the Lagrangian for a particle satisfying the Dirac equation:

$$\psi(x) \longrightarrow \psi'(x) = e^{iq\chi(x)}\psi(x), \tag{1.9}$$

where q is the electric charge of the particle and $\chi(x)$ is a space-time dependent phase. To denote the electromagnetic coupling strength, the fine-structure constant, $\alpha_{\text{QED}} = e^2/(4\pi) \approx 1/137$, is sometimes used. The Lagrangian $\mathscr{L}_D$ of Equation 1.7 then transforms as:

$$\begin{aligned}\mathscr{L}_D \longrightarrow \mathscr{L}'_D &= ie^{-iq\chi}\bar{\psi}\gamma^\mu\left[e^{iq\chi}\partial_\mu\psi + iq\left(\partial_\mu\chi\right)e^{iq\chi}\psi\right] - me^{-iq\chi}\bar{\psi}e^{iq\chi}\psi \\ &= \mathscr{L}_D - q\bar{\psi}\gamma^\mu\left(\partial_\mu\chi\right)\psi.\end{aligned} \tag{1.10}$$

As it stands, the Lagrangian is not invariant under a local U(1) phase transformation. The required gauge invariance is restored by replacing the derivative ∂_μ with the covariant derivative:

$$\partial_\mu \longrightarrow D_\mu = \partial_\mu + iqA_\mu, \tag{1.11}$$

and by introducing the photon field, A_μ, which transforms as:

$$A_\mu \longrightarrow A'_\mu = A_\mu - \partial_\mu \chi. \tag{1.12}$$

The photon field satisfies the wave equation $\partial_\mu \partial^\mu A^\nu = 0$. The complete, gauge-invariant Lagrangian density can be expressed as:

$$\mathscr{L}_{\text{QED}} = \mathscr{L}_{\text{D}} - q\bar{\psi}\gamma^\mu \psi A_\mu - \frac{1}{4} F_{\mu\nu} F^{\mu\nu}, \tag{1.13}$$

where $F_{\mu\nu}$ is the photon (gauge) kinetic energy term:

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu, \tag{1.14}$$

which is also invariant under the U(1) local phase transformation. The resulting QFT is called quantum electrodynamics (QED). Thus, a gauge boson, the photon, which interacts with a matter field, i.e. the electron, emerges naturally by requiring local gauge invariance. Furthermore, it has been shown that only QFTs with local gauge invariance are renormalisable, i.e. predictive beyond a specific (low-energy) regime [4]. All other interactions in the SM are generated identically, the only difference being that the corresponding gauge symmetries are based on higher-dimensional (special) unitary groups.

A significant result, known as Nöther's theorem [5], states that there exists a conservation law associated with each differentiable symmetry of the Lagrangian. The U(1) symmetry of the QED Lagrangian results in the conservation of the electric charge, q. It is important to stress that even though a global phase invariance would still lead to charge conservation, it would not result in the photon field. Finally, the global Poincaré symmetry group, consisting of translational symmetries, rotational symmetries, and Lorentz boosts, results in conservation laws of energy, momentum, and angular momentum.

1.3 Strong interaction

The strong interaction, described by the theory of quantum chromodynamics (QCD), corresponds to an SU(3) gauge symmetry:

$$\psi(x) \longrightarrow \psi'(x) = \exp\left[ig_s\boldsymbol{\alpha}(x)\cdot\hat{\mathbf{T}}\right]\psi(x), \tag{1.15}$$

where $\boldsymbol{\alpha}(x)$ is a three-dimensional phase, $\hat{\mathbf{T}}$ are the generators of the SU(3) symmetry group, related to the 3 × 3 Gell-Mann matrices $T^a = \lambda^a/2, a = 1, \ldots, 8$, and g_s encodes the coupling strength of the strong interaction. The wavefunction has three components:

$$\psi = \left(\psi_{\text{red}}, \psi_{\text{blue}}, \psi_{\text{green}}\right), \tag{1.16}$$

called "colours", corresponding to the three conserved colour charges of the strong interaction. A gauge-invariant Dirac Lagrangian density is constructed by introducing eight gluon fields, one for each of the generators, $G^a_\mu(x)$, which act as the strong force mediators:

$$\mathscr{L}_{\text{QCD}} = \sum_{j=\text{u,d,s,c,b,t}} \overline{\psi_j}\left(i\gamma^\mu D_\mu - m_j\right)\psi_j - \frac{1}{4}F^a_{\mu\nu}F^{a,\mu\nu}, \tag{1.17}$$

with the gauge-covariant derivative:

$$D_\mu = \partial_\mu + ig_s G^a_\mu \lambda^a/2. \tag{1.18}$$

Unlike QED, QCD is a non-Abelian gauge theory — the generators of the SU(3) group do not commute, instead $[T^a, T^b] = if^{abc}T^c$, where f^{abc} are the structure constants of the SU(3) symmetry group. The field strength tensor then becomes:

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g_s f^{abc} A^b_\mu A^c_\nu, \tag{1.19}$$

where the additional term is necessary to uphold gauge invariance. The non-Abelian structure has significant consequences for the theory, as it gives rise to gluon self-interactions in the gauge term $F^a_{\mu\nu}F^{a,\mu\nu}$.

QCD has a property called asymptotic freedom, meaning the strength of the interaction, conventionally denoted as $\alpha_s \sim g_s^2$, decreases with increasing energy transfer between

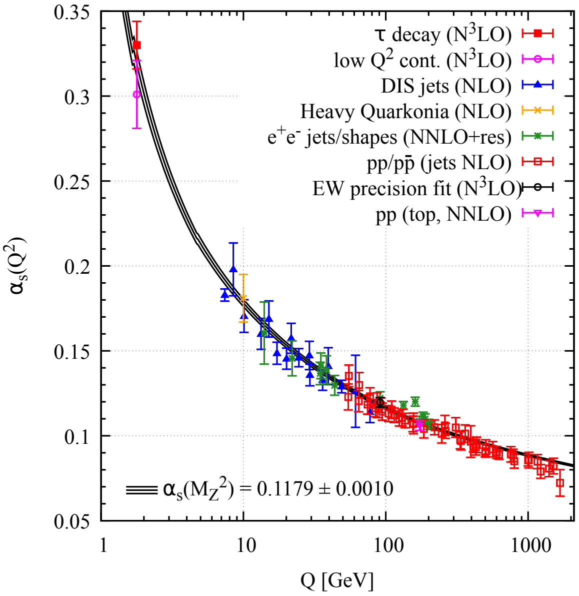


Figure 1.1: Summary of measurements of α_s as a function of the energy scale Q. Figure taken from Ref. [3].

interacting particles, Q, in particular, $\alpha_s \longrightarrow 0$, as $Q \longrightarrow \infty$. Various experimental measurements of α_s as a function of Q are shown in Figure 1.1. At the confinement scale, $\Lambda_{\mathrm{QCD}} \cong 300\,\mathrm{MeV}$, where $\alpha_s(\Lambda_{\mathrm{QCD}}) \geq 1$, the theory becomes non-perturbative, as can be seen from Equation 1.4. On the other hand, perturbation theory can still be employed at higher energies, e.g. $\alpha_s(100\,\mathrm{GeV}) = \mathcal{O}(0.1)$. Such energies are readily accessible at hadron colliders such as the LHC. Perturbative calculations in QCD are done by fixing the physical coupling constant at the renormalisation scale, μ_{R}. For instance, at the Z boson mass, the coupling constant is $\alpha_s(m_{\mathrm{Z}}^2) = 0.1179$. The contribution of quarks and gluons to the "bare" coupling constant is absorbed into the physical constant at the scale μ_{R}.

Although experimental evidence is abundant for quarks, they have not been observed as free particles. In the non-perturbative regime $Q < \Lambda_{\mathrm{QCD}}$, lattice field theory must be used to make predictions. In this regime, quarks form bound colourless states, called hadrons. The principle of colour confinement states that coloured objects only exist as singlet states [6]. On the contrary, objects with a non-zero colour charge cannot propagate as free particles. Colour confinement restricts the possible hadronic states, which can be characterised based on valence quark content as mesons ($\mathrm{q\bar{q}}$) and baryons (qqq, $\mathrm{\bar{q}\bar{q}\bar{q}}$). For example, the quark content of a proton is uud. The LHCb collaboration has also recently observed exotic tetraquark ($\mathrm{qq\bar{q}\bar{q}}$) and pentaquark ($\mathrm{qqqq\bar{q}}$) states [7], [8]. Apart from the valence quarks, a sizeable fraction of the hadron energy is made of virtual gluons and quarks (sea quarks). Collectively, all strongly-interacting constituents of a hadron are referred to as partons.

The energy content of a hadron is given by parton distribution functions (PDFs), describing the momentum fraction distribution of a given type of parton. The PDFs describe the internal structure of hadrons. They must be measured experimentally at a given energy transfer Q^2, as perturbation theory is not applicable in the low-energy regime of bound quark states. However, the evolution of PDFs to the relevant experimental energy scale can still be calculated from the DGLAP[1] equation [9]–[11]. PDFs are indispensable when calculating interaction cross-section in hadron colliders, such as the LHC, as interactions between hadrons are modelled as interactions between constituent partons in the hard scattering process, where a parton of flavour f_i carries a momentum fraction x of the total hadron energy. The momentum contributions from all parton flavours result in the following normalisation equation:

$$\int_0^1 \sum_i^{\text{partons}} \mathrm{PDF}\left(x, f_i, \mu_\mathrm{F}^2\right) \cdot x \mathrm{d}x = 1, \tag{1.20}$$

The factorisation scale μ_F is a key concept in perturbative QCD, and it defines the boundary between short- and long-distance physics. Collinear emissions of partons with energy transfers below μ_F are implicitly absorbed into the PDFs themselves, i.e. they are treated as part of the internal proton structure. Emissions of partons above μ_F are treated as part of the hard process. The renormalisation and factorisation scales are typically chosen to match the expected energy transfer in that process $\mu_\mathrm{F,R}^2 \approx Q^2$. It is important to stress that renormalisation and factorisation scales are not physical quantities and arise due to calculations being performed at some fixed order in perturbation theory. The higher-order calculations result in decreasing dependence on the choice of the scale. Uncertainties in the calculations are conventionally determined by varying the factorisation and renormalisation scales by a factor of two.

The data analysis performed within this thesis uses pp collision data. As a result, only the proton PDFs are directly relevant. Proton PDFs have been measured in various experiments, such as fixed-target electron-proton scattering, high-energy electron-proton collider experiments, high-energy proton-antiproton collision data from Tevatron, and high-energy pp collision data from the LHC itself. Most of these measurements are complimentary, and a combined fit provides tight constraints on proton PDFs. The analysis described in this thesis uses the neural network parton distribution function (NNPDF) set [12], [13], which is obtained using NNs as basic interpolating functions and a Monte Carlo (MC) method to estimate the PDF uncertainties. The proton PDFs for two different scales are shown in Figure 1.2. The contribution from valence u and d quarks is seen to peak at $x \sim 0.1$–0.2, the u quark contribution is approximately twice as large as that of the d quark, as expected from the proton composition. On the other hand, at low x values the proton is dominated by a “gluon sea” instead.

---

[1] Dokshitzer–Gribov–Lipatov–Altarelli–Parisi

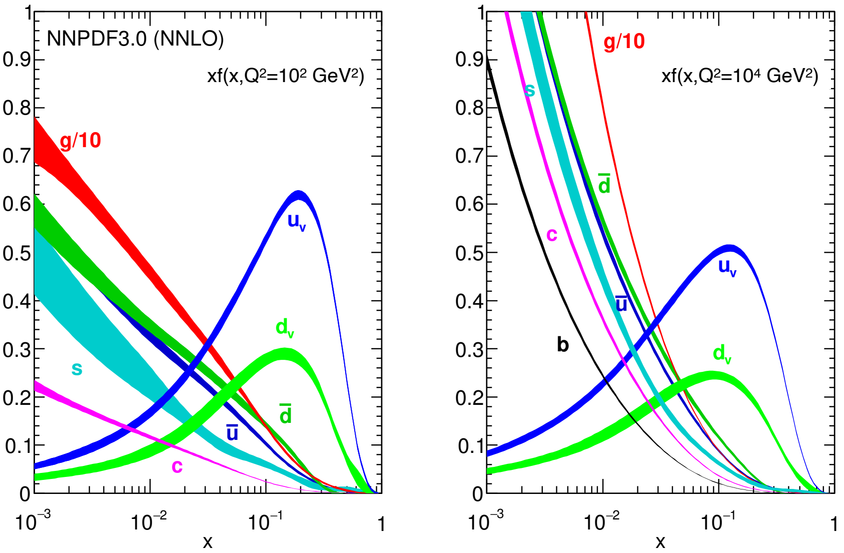


Figure 1.2: Proton PDFs in the NNPDF3.0 set. The valence (u_V, d_V) quark, sea quark, and gluon (scaled down by a factor of 10) distributions are shown at the energy scale of $Q^2 = 10^2$ GeV (left) and $Q^2 = 10^4$ GeV (right).

Another crucial experimental implication of quark confinement is that high-energy quarks, such as those produced at the LHC, form collimated bunches of hadrons. This process is known as hadronisation, and the resulting hadron jets follow the direction of the original quarks.

1.4 Electroweak sector

The unified Glashow-Salam-Weinberg (GSW) model [14]–[16] combines electromagnetic and weak forces. The EW sector is described by an internal SU(2) ⊗ U(1) symmetry:

$$\begin{aligned}\mathscr{L}_{\text{EW}} = \bar{\psi}\gamma^{\mu}\left(i\partial_{\mu} - \underbrace{g'\frac{1}{2}YB_{\mu}}_{\text{U(1)}} - \underbrace{g\frac{1}{2}\boldsymbol{\sigma}\cdot\mathbf{W}_{\mu}}_{\text{SU(2)}}\right)\psi \\ -\frac{1}{4}B_{\mu\nu}B^{\mu\nu} - \frac{1}{2}\,\text{tr}\,W_{\mu\nu}W^{\mu\nu},\end{aligned} \tag{1.21}$$

where $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$ are the three Pauli spin matrices, and $\mathbf{W}_\mu = \left(\mathrm{W}_\mu^{(1)}, \mathrm{W}_\mu^{(2)}, \mathrm{W}_\mu^{(3)}\right)$ are three gauge fields of the SU(2) symmetry group, while g and g' are the weak coupling constants. In the GSW model, the U(1) gauge symmetry of QED is redefined by a different U(1) gauge symmetry, where the electric charge and photon field are replaced by a weak hypercharge and a new gauge field: $q \longrightarrow Y,\ A_\mu \longrightarrow B_\mu$. The relationship between the two sets of quantum numbers is:

$$q = T_3 + \frac{1}{2}Y, \tag{1.22}$$

where T_3 is the third component of the weak isospin, the weak charge of the $\mathbf{W}_\mu$ gauge fields. The physical W bosons, mediators of the charged-current (CC) part of the weak interaction, are identified as the linear combination:

$$\mathrm{W}_\mu^\pm = \frac{1}{\sqrt{2}}\left(\mathrm{W}_\mu^{(1)} \mp i\,\mathrm{W}_\mu^{(2)}\right). \tag{1.23}$$

A vital feature of the weak interaction is that, unlike QED and QCD, it has been experimentally observed to violate parity [17]. To this end, it is helpful to decompose the particle state ψ into left- and right-handed components:

$$\psi = \psi_L + \psi_R, \tag{1.24}$$

by using the chiral projection operators $P_{L,R} = \frac{1}{2}(1 \mp \gamma^5)$, which satisfy:

$$\begin{aligned} \psi_L &= P_L\psi, & \psi_R &= P_R\psi, & &\text{for particles,} \\ \overline{\psi_L} &= \overline{\psi}P_R, & \overline{\psi_R} &= \overline{\psi}P_L, & &\text{for antiparticles.} \end{aligned} \tag{1.25}$$

The CC weak interaction has a V-A (vector minus axial-vector) structure, meaning that only left-handed (right-handed) chiral (anti-) particle states ψ_L $\left(\overline{\psi_R}\right)$ participate in the interaction. The physical W bosons, the left-handed fermion states, and the right-handed fermion antiparticle states all form weak isospin doublets with $(T, T_3) = (1/2, \pm 1/2)$.

The photon and Z boson fields are obtained as linear combinations of the B_μ and $\mathrm{W}_\mu^{(3)}$ fields:

$$\begin{aligned} A_\mu &= +B_\mu \cos\theta_\mathrm{W} + \mathrm{W}_\mu^{(3)} \sin\theta_\mathrm{W}, \\ \mathrm{Z}_\mu &= -B_\mu \sin\theta_\mathrm{W} + \mathrm{W}_\mu^{(3)} \cos\theta_\mathrm{W}, \\ &\text{with } e = g\sin\theta_\mathrm{W} = g'\cos\theta_\mathrm{W}. \end{aligned} \tag{1.26}$$

The resulting Z boson, the mediator of the neutral-current (NC) interaction, violates parity, which has been experimentally established [18]. The photon and Z boson fields, the right-handed fermions, and the left-handed antifermions form weak isospin singlets with eigenvalue $T = T_3 = 0$. As a result, electrically uncharged, right-handed neutrinos (left-handed antineutrinos) are not subject to any interaction, i.e. sterile, in the SM.

The weak mixing (Weinberg) angle has been measured experimentally giving $\sin^2\theta_W \cong 0.22$, and hence the weak coupling strength to be $\alpha_W = \alpha_{QED}/\sin^2\theta_W \approx 0.032$ [19]. The strength of the weak interaction is typically expressed in terms of the Fermi coupling constant $G_F \cong 1.17 \times 10^{-5}\,\text{GeV}^{-2}$:

$$G_F \approx \frac{\sqrt{2}g^2}{8m_W^2}. \tag{1.27}$$

The force carriers, as introduced via gauge symmetry invariance, must have zero mass. Two of the force carriers, the gluon and the photon, are massless, while the W and Z bosons are massive. However, the EW Lagrangian contains no mass terms for the gauge bosons or the fermions, which would break the gauge invariance. Thus, it does not describe the observed universe, where the said particles have non-zero mass. For the EW gauge bosons to acquire mass and the theory to be self-consistent, a further spin-0 boson, the Higgs boson, is required.

1.5 Electroweak symmetry breaking

The Higgs mechanism is an essential part of the SM, as it allows the W and Z bosons to acquire mass without breaking the local gauge symmetries of the SM. The Higgs mechanism embeds the idea of electroweak symmetry breaking (EWSB) within a local gauge symmetry. For the SM Higgs mechanism, at least four degrees of freedom are required, and the simplest Higgs model introduces a set of two complex scalar fields, ϕ^+ and ϕ^0, one charged and one neutral, and both with $Y = 1$:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}, \tag{1.28}$$

which satisfy the following Lagrangian density:

$$\mathscr{L} = \left(D_\mu\phi\right)^\dagger\left(D^\mu\phi\right) - \mu^2\phi^\dagger\phi - \lambda(\phi^\dagger\phi)^2. \tag{1.29}$$

This Lagrangian density possesses a global U(1) symmetry, as it is invariant under the transformation $\phi \longrightarrow \phi' = e^{i\alpha}\phi$. Assuming $\mu^2 < 0, \lambda > 0$, the vacuum state of the field will not be at $\phi = 0$, but instead at:

$$\phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2 = v^2 = \frac{-\mu^2}{\lambda}. \tag{1.30}$$

There is an infinite number of minima, but only one ground state is realised. Thus, in EWSB, the U(1) symmetry of the Lagrangian is violated by the system which it describes around the vacuum state. The excitations of the field, which describe the particle states, are obtained in perturbation theory by considering small deviations from the vacuum state. In the unitary gauge, the vacuum state is chosen to correspond to $\phi^+ = 0, \phi^0 = v$, and the Higgs doublet is expanded around the vacuum state as:

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}. \tag{1.31}$$

The required local gauge invariance can be restored by replacing derivatives in the Lagrangian with the covariant derivative:

$$D_\mu \phi = \frac{1}{2} \left[2\partial_\mu + \left(ig\boldsymbol{\sigma} \cdot \mathbf{W}_\mu + ig' B_\mu \right) \right] \phi, \tag{1.32}$$

such that the Lagrangian respects the SU(2) ⊗ U(1) local gauge symmetry of the EW model. Expanding the kinetic part of the Lagrangian density gives:

$$\begin{aligned} \left(D_\mu \phi\right)^\dagger \left(D^\mu \phi\right) = & \frac{1}{2} \left(\partial_\mu h\right) \left(\partial^\mu h\right) + \frac{1}{8} g^2 \left(\mathrm{W}_\mu^{(1)} + i\mathrm{W}_\mu^{(2)}\right) \left(\mathrm{W}^{(1)\mu} - i\mathrm{W}^{(2)\mu}\right) (v+h)^2 \\ & + \frac{1}{8} \left(g\mathrm{W}_\mu^{(3)} - g' B_\mu\right) \left(g\mathrm{W}^{(3)\mu} - g' B^\mu\right) (v+h)^2. \end{aligned} \tag{1.33}$$

From this, the masses of the gauge bosons can be identified by collecting the terms quadratic in the gauge boson fields. The gauge fields $\mathrm{W}_\mu^{(3)}, B_\mu$ are mixed; however, a diagonalising transformation can be performed, which returns the physical photon and Z boson fields of Equation 1.26. From quadratic terms in $\mathbf{W}_\mu$ and B_μ we obtain:

$$m_\mathrm{W} = \frac{1}{2} g v, m_\mathrm{Z} = \frac{1}{2} v \sqrt{g^2 + g'^2}, m_\mathrm{A} = 0. \tag{1.34}$$

Hence, the gauge bosons acquire mass with the Higgs mechanism, while the Lagrangian still holds a "hidden" local gauge symmetry. The vacuum expectation value of the Higgs field, which carries weak isospin but not electric nor colour charge, is given by:

$$v = \left(\sqrt{2}G_{\mathrm{F}}\right)^{-\frac{1}{2}} \cong 246\,\mathrm{GeV}. \tag{1.35}$$

1.5.1 Fermion mass generation

The mass term in the Dirac Lagrangian (Equation 1.7):

$$-m\bar{\psi}\psi = -m(\overline{\psi_R}\psi_L + \overline{\psi_L}\psi_R) \tag{1.36}$$

does not respect the SU(2) ⊗ U(1) gauge symmetry and cannot be present in the SM Lagrangian. Remarkably, the Higgs field ϕ can generate the SM fermion masses instead. The mass generation is accomplished by placing left-handed chiral fermions in SU(2) doublets, denoted as L, while right-handed chiral fermions are placed in SU(2) singlets, named R. The Yukawa interaction term is introduced to the Lagrangian:

$$\mathscr{L}_f = -g_f(\bar{L}\phi R + \bar{R}\phi^{\dagger}L), \tag{1.37}$$

where g_f is the corresponding Yukawa coupling of the fermion to the Higgs field. For example, the first SM lepton generation is described as $L = (\nu_e, e)_L \equiv (\psi^L_{\nu_e}, \psi^L_e)$ and $R = e_R \equiv \psi^R_e$. After EWSB, the Yukawa Lagrangian becomes:

$$\mathscr{L}_e = -\frac{g_e}{\sqrt{2}}v(\overline{e_L}e_R + \overline{e_R}e_L) - \frac{g_e}{\sqrt{2}}h(\overline{e_L}e_R + \overline{e_R}e_L). \tag{1.38}$$

The first term is precisely the mass term of Equation 1.36, yet introduced in a gauge-invariant fashion. It represents the electron coupling to the Higgs field, while the second term represents the interaction between the electron and the Higgs boson. The Yukawa coupling is not a predicted parameter of the SM but is related to the fermion masses through:

$$g_f = \sqrt{2}\frac{m_f}{v}. \tag{1.39}$$

Curiously, the Yukawa coupling for the top quark is almost unity, $y_{\mathrm{t}} \sim 0.7$. A more general description is employed to accommodate all Dirac fermions, with the following notation introduced:

$$\begin{aligned} Q_{iL} &= \begin{pmatrix} u_{iL} \\ d_{iL} \end{pmatrix}, u_{1,2,3} = (u, c, t), d_{1,2,3} = (d, s, b), \\ L_{iL} &= \begin{pmatrix} \nu_{iL} \\ \ell_{iL} \end{pmatrix}, \nu_{1,2,3} = (\nu_e, \nu_\mu, \nu_\tau), \ell_{1,2,3} = (e, \mu, \tau). \end{aligned} \tag{1.40}$$

The complete Yukawa Lagrangian is:

$$\mathscr{L}_{\mathrm{Y}} = -\overline{Q}_{iL}\mathbf{Y}^d_{ij}\phi d_{jR} + \overline{Q}_{iL}\mathbf{Y}^u_{ij}\phi_c u_{jR} - \overline{L}_{iL}\mathbf{Y}^l_{ij}\phi l_{jR} + \overline{L}_{iL}\mathbf{Y}^\nu_{ij}\phi_c \nu_{jR} + \mathrm{h.c.}, \tag{1.41}$$

where h.c. means Hermitian conjugate and the conjugate Higgs doublet is:

$$\phi_c = -i\sigma_2\phi^* = \begin{pmatrix} -\phi^{0*} \\ \phi^- \end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix} -\phi_3 + i\phi_4 \\ \phi_1 - i\phi_2 \end{pmatrix}. \tag{1.42}$$

After EWSB, the physical mass states become a mixture of flavour states:

$$\mathbf{M}^u = \frac{v}{\sqrt{2}}\mathbf{Y}^u, \mathbf{M}^d = \frac{v}{\sqrt{2}}\mathbf{Y}^d, \mathbf{M}^\ell = \frac{v}{\sqrt{2}}\mathbf{Y}^\ell, \mathbf{M}^\nu = \frac{v}{\sqrt{2}}\mathbf{Y}^\nu, \tag{1.43}$$

parameterised by the 3 × 3 coupling matrices: $\mathbf{Y}^d$ and $\mathbf{Y}^\ell$ are chosen to be diagonal; $\mathbf{Y}^u$ is called the CKM[2] (quark-mixing) matrix, $\mathbf{Y}^\nu$ is called the PMNS[3] (neutrino-mixing) matrix. Their elements are free parameters in the SM and must be measured experimentally.

---

[2] Cabibbo–Kobayashi–Maskawa

[3] Pontecorvo–Maki–Nakagawa–Sakata

Chapter 2

Physics beyond the Standard Model

In this Chapter, several extensions to the SM, that could address some of its open questions, are described. These models complement the SM by extending its particle content and, crucially, can give rise to LLPs.

2.1 Introduction

The SM has been very successful in withstanding various EW precision measurements. However, several theoretical and experimental observations suggest that the SM is an effective theory, which breaks down at some energy, Λ_{UV}, below the Planck scale, $\Lambda_{\mathrm{Pl.}} = 1.22 \cdot 10^{19}$ GeV, a point at which the quantum effects of gravity become strong. For instance, the coupling constants of the fundamental interactions are of a similar order of magnitude, as are the masses of fermions within a given generation. Related to that, the large number of free parameters in the SM (26) appears contrived. Strangely, the SM also possesses four "accidental" U(1) global symmetries, not explicitly used in its construction, which correspond to baryon and lepton number conservation. It is thought that the three gauge interactions of the SM, comprising the electromagnetic, weak, and strong interactions, could be unified into a single Grand Unified Theory (GUT) at high energies, just like the GSW theory unifies the electromagnetic and weak interactions. In addition, several critical observations within the realm of high energy physics (HEP), astrophysics and cosmology, including the smallness of neutrino masses, the existence of dark matter (DM), and the baryon asymmetry in the universe are not explained within the SM.

2.1.1 The nature of neutrino masses

The observation of neutrino flavour oscillations from neutrino experiments conclusively established that at least two SM neutrinos have a non-zero mass. The current results from the oscillation experiments, expressed in terms of the "atmospheric" and the "solar" mass differences, are:

$$\Delta m_{21}^2 = m_2^2 - m_1^2 = (7.53 \pm 0.18) \times 10^{-5} \mathrm{eV}^2, \tag{2.1}$$

$$\left|\Delta m_{32}\right|^2 = \left|m_3^2 - m_2^2\right| = (2.44 \pm 0.12) \times 10^{-3} \mathrm{eV}^2. \tag{2.2}$$

The oscillation experiments only allow determining the differences between mass states; whether the neutrino masses follow the normal ($\Delta m_{32}^2 > 0$) or the inverted mass hierarchy ($\Delta m_{32}^2 < 0$) is at present not known. While the absolute values of neutrino masses are unknown, several measurements place ULs on neutrino masses. One such limit was established from the Planck collaboration measurements of the cosmic microwave background (CMB) anisotropies [20]:

$$\sum_{i=1,2,3} m_\nu < 0.12 \text{ eV}. \tag{2.3}$$

Finally, direct measurements of tritium β-decay from the KATRIN experiment [21] provide an UL on the electron neutrino mass:

$$m_{\nu_e} < 1.1 \text{ eV}. \tag{2.4}$$

Based on the previously mentioned measurements, there should be a corresponding neutrino mass term in the SM Lagrangian. However, the origin of small neutrino masses remains a mystery. As shown in Equation 1.41, one of the ways to generate neutrino masses is by postulating a Dirac mass term:

$$\mathscr{L}_D = -m_D \left(\overline{\nu_R} \nu_L + \overline{\nu_L} \nu_R\right), \tag{2.5}$$

however, this requires a particle to have both a left-handed and a right-handed chiral state. As shown in Section 1.4, right-handed neutrinos would not participate in any interaction in the SM, and there is no direct experimental evidence that they exist. Furthermore, as the neutrino masses are so small compared to other fermions, it suggests another mechanism might be at play. Considering the results of oscillation experiments, the Yukawa coupling would have to be at least ten orders of magnitude smaller than couplings of other particles of the third generation, i.e. $y_\mathrm{t}/y_{\nu_3} > 10^{10}$. Hence, it seems unlikely that the neutrino masses have the same origin as the other fermions in the SM, and thus, another mechanism might be at play.

It turns out that it is also possible to generate their mass via the Majorana mechanism without requiring a right-handed chiral state. The neutrino state can be expressed as a superposition of left- and right-handed chiral projections according to Equation 1.24:

$$\nu = \nu_L + \nu_R = \nu_L + C\overline{\nu_L}^T = \nu_L + \nu_L^C, \tag{2.6}$$

where a charge-conjugate field $\nu_L^C = C\overline{\nu_L}^T$ is defined, and C is the charge-conjugation operator, which reverses the electric charge. This expression is only possible since neutrinos are electrically neutral, and as a result, Majorana neutrinos would be their own antiparticles:

$$\nu^C = \left(\nu_L + \nu_L^C\right)^C = \nu_L^C + \nu_L = \nu. \tag{2.7}$$

The Majorana mass term is then:

$$\mathscr{L}_M = -\frac{1}{2}M\left(\overline{\nu_R^C}\nu_R + \overline{\nu_L^C}\nu_L\right). \tag{2.8}$$

Hence, it is possible to construct a Majorana mass term solely from one-handed neutrino fields. For left-chiral Majorana neutrinos ν_L, the Yukawa term would change weak hypercharge by two units, inconsistent with the standard Higgs doublet, as introduced in Equation 1.28. Thus, it would require the Higgs field to be extended to include an extra triplet with $Y = 2$. On the other hand, no Higgs extensions are necessary for right-handed Majorana neutrinos ν_R.

The question of whether neutrinos are Dirac or Majorana particles is of great theoretical interest. One practical consequence of neutrinos being Majorana particles is lepton number violation (LNV), implying the possibility of observing the phenomenon of neutrinoless double-β decay, which is being actively searched for [22].

2.1.2 Dark matter

The existence of DM provides perhaps the most compelling evidence for beyond Standard Model (BSM) physics. The most direct evidence of DM comes from galaxy rotational curves [23], where the velocity of stars do not correspond to the gravitational force due to the measurable mass content of the galaxy. From this observation, it can be concluded that there exists a non-luminous component of mass in galaxies. Gravitational lensing provides further strong evidence for DM [24] and allows to map DM distributions around galaxy clusters [25]. The existence of DM is also supported by cosmological measurements, such as those of the CMB anisotropies by the Planck collaboration [20].

2.1.3 Baryon asymmetry

The fact that the universe is primarily made of baryonic matter remains a mystery, as it is expected that equal amounts of matter and antimatter were produced during the Big Bang. Sakharov formulated three necessary conditions for baryogenesis to occur: baryon number violation, C- and CP-symmetry violation, and interactions out of the thermal equilibrium [26]. CP-symmetry is violated in the weak sector of the SM, namely in the quark-mixing matrix $\mathbf{Y}^u$ and potentially the neutrino-mixing matrix $\mathbf{Y}^\nu$; however, this could only account for a small fraction of the observed baryon asymmetry [27].

2.1.4 The hierarchy problem

As a generic scalar boson, the Higgs boson mass receives quantum corrections at the loop level from self-interaction, gauge boson, and fermion loops. Crucially, corrections due to fermions via Yukawa coupling diverge quadratically at some cut-off scale Λ_{UV}, where the SM ceases to function as an effective theory:

$$\Delta m_{\mathrm{H}}^2 = -\frac{\left|\lambda_f\right|^2}{8\pi^2}\Lambda_{\mathrm{UV}}^2, \tag{2.9}$$

where λ_f is the Yukawa coupling to a fermion. In the SM, the dominant contribution comes from the top quark, the most massive fermion. The smallness of the Higgs boson mass ($\Lambda_{\mathrm{Pl.}} = 10^{17} m_{\mathrm{H}}$) is problematic, as it means corrections to the Higgs boson mass are many orders of magnitude higher than the mass itself, $\Delta m_{\mathrm{H}}^2 \gg m_{\mathrm{H}}^2$, which could only result from some unnatural fine-tuning of the SM parameters.

2.2 Supersymmetry

SUSY hypothesises a new symmetry that provides a connection between bosons and fermions that seemingly have no apparent relationship in the SM [28]–[31]. The Minimal Supersymmetric Standard Model (MSSM) [32] is a minimal extension of the SM that doubles its particle spectrum: there is a corresponding supersymmetric boson for every SM fermion, and every SM boson has a fermionic superpartner. If SUSY were an exact symmetry, each superpartner would have the same mass as its SM equivalent. Since no such superpartner has been observed experimentally, SUSY would have to be a broken symmetry [33], similarly to the Higgs mechanism. In addition, the MSSM contains two Higgs-doublet fields, instead of one, as well as their superpartners [34]. In the MSSM, neither the baryon number (B) nor the lepton number (L) is conserved, while the difference $B - L$ remains conserved. The conserved R-parity is thus introduced as:

$$P_{\mathrm{R}} = (-1)^{3(B-L)+2s}, \tag{2.10}$$

where s is the particle's spin. All SM particles have an R-parity of +1, while the supersymmetric particles are −1. SUSY is very appealing from a theoretical point of view for several reasons. First, the MSSM could provide a candidate DM particle. A generic candidate class for DM are weakly interacting massive particles (WIMPs), corresponding to a new particle only interacting weakly with a mass in the 100 GeV range. Due to R-parity conservation, the lightest supersymmetric particle (LSP) cannot decay and is a WIMP candidate [35]. Moreover, SUSY could provide a solution to the hierarchy problem, as contributions from supersymmetric scalar partners could cancel out the SM corrections of Equation 2.9:

$$\Delta m_{\mathrm{H}}^2 = \left[2 \times \frac{\lambda_S}{16\pi^2} - \frac{\left|\lambda_f\right|^2}{8\pi^2} \right] \Lambda_{\mathrm{UV}}^2, \tag{2.11}$$

assuming there are two scalars with $\lambda_S = \left|\lambda_f\right|^2$. For this to happen, the symmetry breaking would have to occur around the TeV scale, i.e. accessible at the LHC. In addition, TeV-scale superpartners would facilitate the unification of gauge couplings at high energies [36].

In most models, the pair production of gluinos, supersymmetric partners of gluons, $\mathrm{pp} \longrightarrow \tilde{\mathrm{g}}\tilde{\mathrm{g}}$, is the dominant SUSY signature at the LHC by cross-section, and hence one of the golden channels for SUSY discovery. Figure 2.1 shows a summary of recent searches for gluinos decaying to light quarks (u, d, s) and the LSP, performed by CMS, which exclude gluinos with masses up to 2 TeV. In the compressed mass scenario, where the difference between the LSP and gluino masses is relatively small ($\lesssim$ 100 GeV), the exclusion limits are not as stringent due to the more challenging low-p_T quark final state.

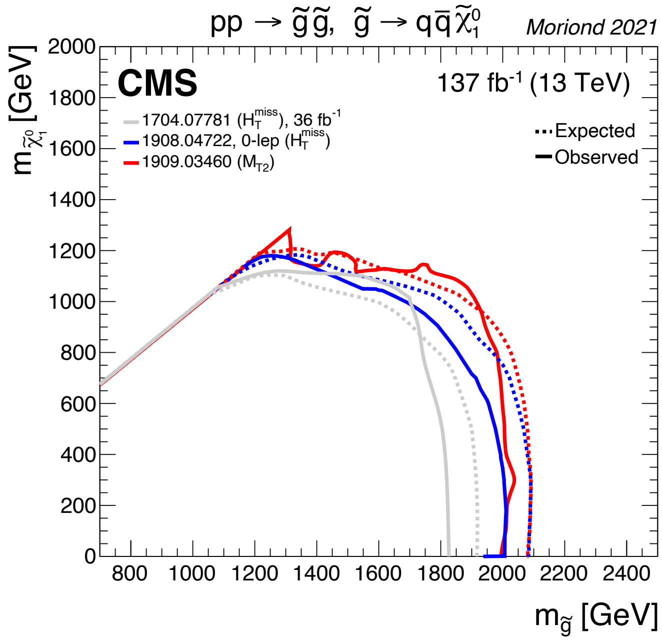


Figure 2.1: Mass limits for a simplified model of gluino pair production and decay to pairs of light quarks and the LSP. Figure taken from Ref. [37]

2.3 The seesaw mechanism

In the most general case, the neutrino mass term can be written as a superposition of Dirac (Equation 2.5) and left- and right-handed Majorana terms (Equation 2.8):

$$\mathscr{L}_{\text{D-M}} = -\frac{1}{2}\left[m_D\overline{\nu_L}\nu_R + m_D\overline{\nu_R^C}\nu_L^C + M\overline{\nu_R^C}\nu_R\right] + \text{h.c.}, \tag{2.12}$$

which alternatively can be expressed as a matrix:

$$\mathscr{L}_{\text{D-M}} = -\frac{1}{2}\begin{pmatrix} \overline{\nu_L} & \overline{\nu_R^C} \end{pmatrix}\begin{pmatrix} 0 & m_D \\ m_D & M \end{pmatrix}\begin{pmatrix} \nu_L^C \\ \nu_R \end{pmatrix} + \text{h.c.} \tag{2.13}$$

Such a term would violate baryon and lepton number but preserve $B - L$ invariance [38]. The physical states of the system can be identified by diagonalising the mass matrix, as was done for the gauge bosons of the EW interaction, which gives the following mass states:

$$m_{1,2} = \frac{M \pm \sqrt{M^2 - 4m_D^2}}{2}. \tag{2.14}$$

The type-I seesaw mechanism assumes the Majorana mass is much greater than the Dirac mass, $M \gg m_D$, which results in a light neutrino state, ν, and a heavy neutrino state, N, with the following masses:

$$m_\nu \approx \frac{m_D^2}{M} \quad \text{and} \quad m_N \approx M. \tag{2.15}$$

Naturally, the Dirac mass term could be of a similar magnitude compared to the other SM fermions, $m_D \sim 1\,\text{GeV}$, but the physical neutrino state is made small by an arbitrarily large Majorana mass, i.e. $M \sim 10^{11}\,\text{GeV}$, as there is no SM principle that would prevent it from being extremely large. The hypothetical Majorana particle, like the light neutrino ν, would only experience weak interactions. However, since the heavy mass eigenstate is almost entirely right-handed, it can effectively be considered sterile under all SM gauge interactions. Hence, it can only be produced via mixing with the active neutrino, ν. To this end, the Neutrino Minimal Standard Model (νSSM) [39] has been proposed as a minimal extension to the SM to accommodate the non-zero neutrino masses. The model proposes the existence of three heavy right-handed Majorana neutrinos, also called HNLs, denoted as $\mathrm{N}_1, \mathrm{N}_2, \mathrm{N}_3$ — one for each SM fermion generation.

The existence of HNLs may not only answer the nature of neutrino masses, but may also provide a possible candidate for DM. The lightest sterile neutrino, N_1, is usually selected as a DM candidate with a mass in the 1–10 keV range. It couples very weakly to active neutrinos, and the estimated restriction on its lifetime exceeds the age of the universe by several orders of magnitude [40]. There are proposals for β-decay experiments to detect a sterile neutrino in this mass range [41], however, searches usually target the other two generations of HNLs ($\mathrm{N}_2, \mathrm{N}_3$) instead. These heavier HNLs, which couple more strongly to the active neutrinos, are necessary to explain the pattern of neutrino oscillations, and could be directly accessible at colliders and other experiments.

HNLs could explain the baryon asymmetry of the universe, as the νSSM can readily accommodate the required CP-symmetry violation. HNLs created in the early universe could be a mechanism of leptogenesis [42], a process producing an excess of leptons over their antiparticles. As examined in Ref. [43], the required matter-antimatter asymmetry in the universe places restrictions on the mass of the two additional sterile neutrinos: they should be in the mass range of $0.15\,\text{GeV} \leq m_{\mathrm{N}_{2,3}} \leq 100\,\text{GeV}$ and be nearly degenerate ($|m_{\mathrm{N}_2} - m_{\mathrm{N}_3}| \ll m_{\mathrm{N}_2,\mathrm{N}_3}$). Finally, the lifetime of the neutrinos should be $\tau_{2,3} < 0.02\,\text{s}$ not to affect the predictions of Big Bang nucleosynthesis (BBN) [44].

In the “inverse” variation of the seesaw mechanism [45], the light neutrino masses are directly proportional to a small lepton-number breaking scale. This scenario can be considered an extension of the type-I seesaw, as it requires two sets of singlet fermions instead of one. The coherent superposition of two quasi-degenerate Majorana fermions behaves like a Dirac fermion, which is the main distinguishing experimental feature of this variation.

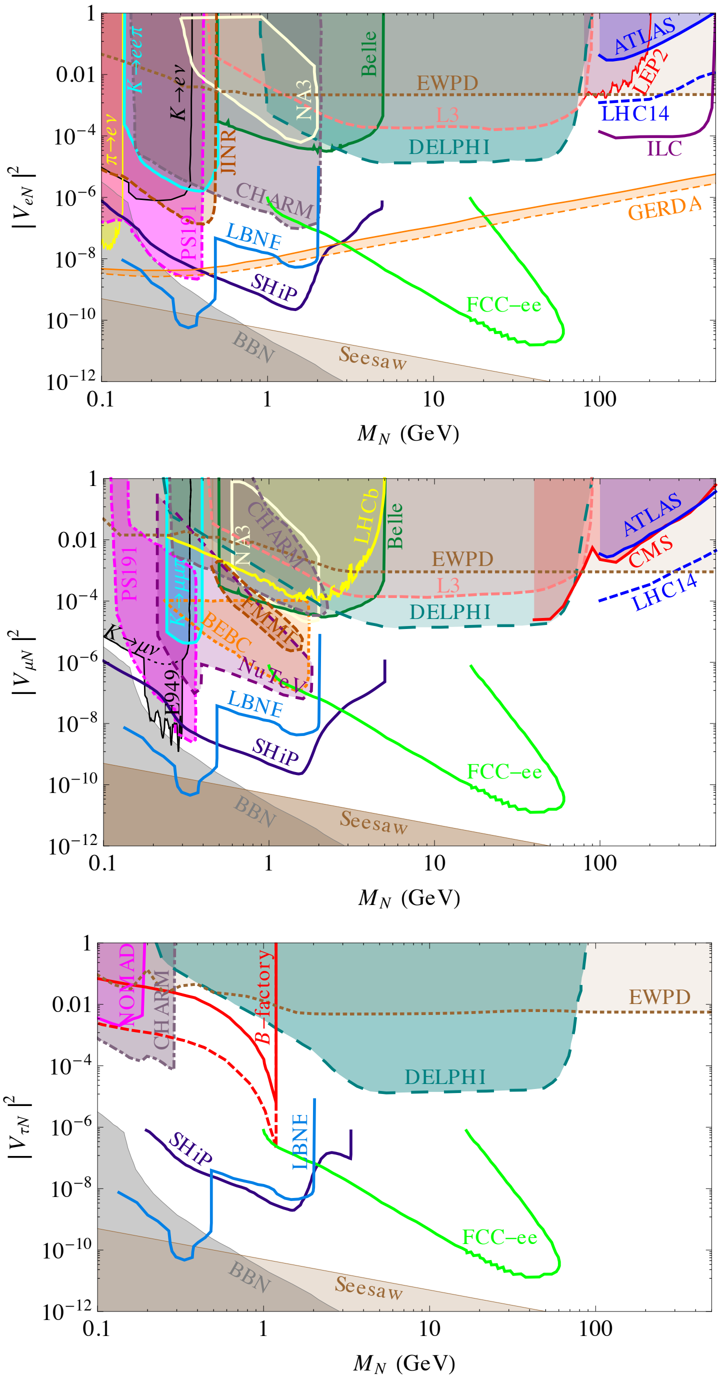


Figure 2.2: Exclusion limits obtained from a range of measurements on the mixing between active neutrinos (electron, muon, tau) and a single HNL in the mass range of 0.1–500 GeV. Figures taken from Ref. [46].

As the HNLs can only be discovered through their coupling with SM neutrinos, the discovery potential is very much dependent on the strength of the coupling between the active and sterile neutrinos. The coupling strength (mixing angle) is denoted by $V_{\ell i}$, where $\ell = e, \mu, \tau$ denotes the active neutrino flavours, and i is the sterile neutrino generation. It is found in Ref. [40] that the coupling of N_2 and N_3 with the active neutrinos must be almost identical. Hence, to simplify the notation, we use V_ℓ is instead. Due to the likely degenerate mass and coupling strengths of $\mathrm{N}_{2,3}$ to the active neutrinos, they are usually considered as a single HNL, denoted as N. Many searches have been performed for HNLs in the mass range ranging from several keV up to the TeV scale, and several others are planned, which are summarised in Figure 2.2. The available HNL phase space is constrained by a variety of complementary measurements and can be subdivided into different regions with different search strategies, some of them highlighted below:

- For $m_\mathrm{Z} < m_\mathrm{N} < 1\,\mathrm{TeV}$, the best sensitivity is derived from direct ATLAS [47] and CMS [48], [49] searches which probe HNL mixing with the first and second lepton generations.
- Above B meson masses, $m_\mathrm{B} < m_\mathrm{N} < m_\mathrm{Z}$, the LHC measurements are complemented with Z boson precision measurements from the Large Electron-Positron Collider (LEP) [50], which, crucially, provide sensitivity for third-generation coupling.
- The heavy meson meson range, $m_\mathrm{K} < m_\mathrm{N} < m_\mathrm{B}$, has been accessed at B-factories such as Belle [51] and the CHARM beam dump experiment [52]. The two experiments utilised B and D meson decays, respectively. In this parameter space, the best future sensitivity is expected from a proposed beam dump experiment, such as SHiP [53], sensitive to HNL mixing with all three lepton generations.
- Finally, the low-mass range ($0.1\,\mathrm{GeV} < m_\mathrm{N} < m_\mathrm{K}$) was investigated using kaon (K) decays to muons in NA62 [54]. In addition, the PS191 experiment used a low-energy neutrino beam, probing coupling to the first two lepton generations [55]. The future long-baseline Deep Underground Neutrino Experiment (DUNE) [56] is expected to provide the best sensitivity in this mass range for all three generations.

Finally, for a given m_N, a lower bound on V_ℓ is provided by BBN constraints and from the type-I seesaw mixing relationship between active and sterile neutrino states:

$$V_\ell \simeq \sqrt{\frac{m_\nu}{m_\mathrm{N}}} \lesssim 10^{-6}\sqrt{\frac{100\,\mathrm{GeV}}{m_\mathrm{N}}}, \tag{2.16}$$

assuming $m_\nu < 0.1\,\mathrm{eV}$. A global fit of the favoured HNL parameter space was performed by the GAMBIT collaboration [57], which employed the previously mentioned results comprising HNL searches, EW precision measurements, and active neutrino mixing experiments. The fit considered HNLs with masses between 60 MeV and 500 GeV and its results for various HNL mixing scenarios are shown in Figure 2.3, assuming normal and inverted active neutrino mass hierarchy. The lines indicate the allowed phase space in a triangle spanned by the flavour ratios $V_\ell^2/(V_e^2 + V_\mu^2 + V_\tau^2)$. If the lightest active neutrino mass is large (black line), there are no strong constraints on the allowed flavour ratios. However, with decreasing lightest active neutrino mass, large HNL mixing with the first lepton generation becomes disfavoured for the normal hierarchy, and with the second and third generation for the inverted hierarchy. In general, simultaneous coupling to multiple generations is preferred by the fit.

To summarise, such a minimal extension to the SM, which explains several critical phenomena in a unified way, is very appealing from a theoretical point of view. So far, no experimental observations have ruled out the possibility of the νSSM [40].

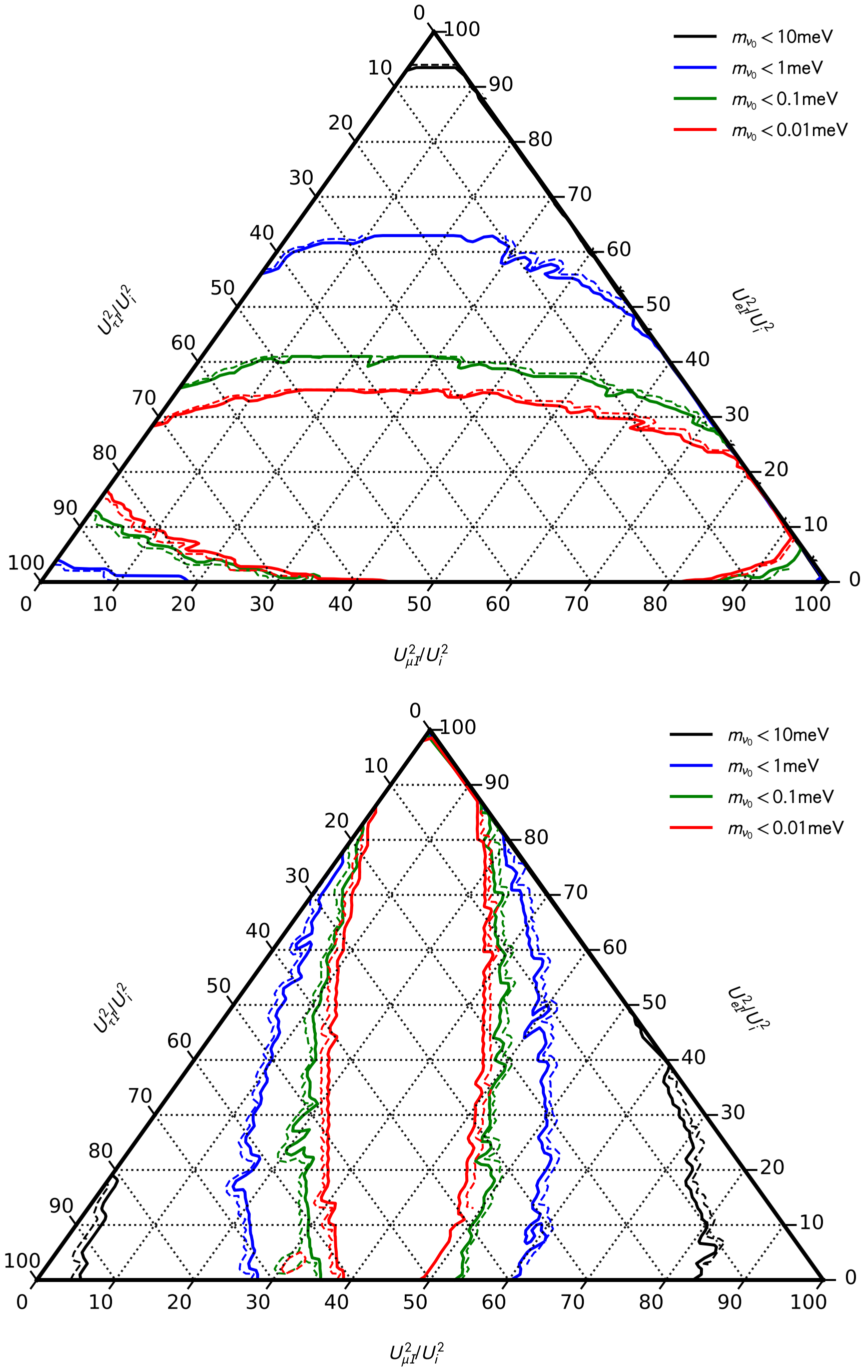


Figure 2.3: Allowed parameter space as a fraction of the HNL coupling to a given active neutrino generation (in percentage) for different HNL ULs for normal (left) and inverted (right) active neutrino mass hierarchy. The allowed parameter space, resembling an ellipse, is delineated by solid (dashed) lines marking the 1σ (2σ) contours. Figure taken from Ref. [57].

2.4 Long-lived particles

In this Section, the properties of LLPs are described. LLPs are introduced as new particles in several BSM scenarios. Traditionally, searches at the LHC have focused on prompt BSM particles, i.e. those with a very short lifetime. Such searches could miss an LLP signature, and hence, dedicated searches for LLPs are required to fully exploit the capabilities of the LHC detectors.

2.4.1 Introduction

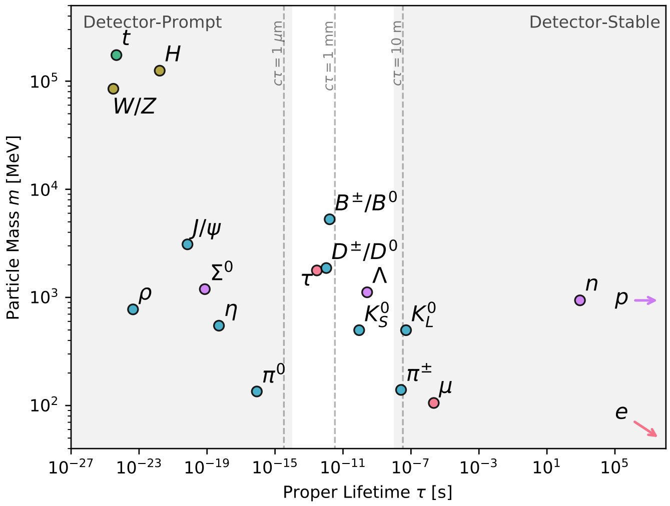


Figure 2.4: Selection of elementary and composite SM particle masses and lifetimes. The shaded regions indicate regions where the particles are either prompt or stable in a general-purpose detector like CMS. Figure taken from Ref. [58].

The partial decay width of a particle can be calculated in perturbation theory using Fermi's golden rule:

$$\Gamma_{fi} = 2\pi |T_{fi}|^2 \rho(E_i), \tag{2.17}$$

where T_{fi} is the transition ME for the process $i \longrightarrow f$, and $\rho(E_i)$ is the density of available decay states at energy E_i. Assuming a particle i decaying to several final-state particles j, the transition element is expressed in terms of the Lorentz-invariant ME:

$$|\mathscr{M}_{fi}|^2 = 2E_i \left(\prod_j 2E_j \right) |T_{fi}|^2. \tag{2.18}$$

The ME can be computed using Feynman rules, obtained from the interaction Lagrangian (Equation 1.3) of a given theory. In general, a particle can decay to several different final states, with a total decay width:

$$\Gamma = \sum_j \Gamma_j. \tag{2.19}$$

From Equation 2.17, it can be deduced that long lifetimes arise from a weak coupling leading to a small transition amplitude or a small available decay phase space (density of states). These conditions are generic features of many BSM models developed to address the open questions of HEP.

The average number of particles remaining after a time t is given by:

$$N(t) = N(0)\exp\left(-\frac{t}{\tau_0}\right), \tag{2.20}$$

where τ_0 is the proper lifetime of the particle (in its rest frame), inversely proportional to its total decay width:

$$\tau_0 = \frac{1}{\Gamma}. \tag{2.21}$$

For a given decay, the BR is defined as:

$$\mathscr{B}(j) = \frac{\Gamma_j}{\Gamma}. \tag{2.22}$$

An LLP with momentum p will travel, on average, the following distance in the transverse plane from its production point:

$$\bar{L}_{xy} = \sqrt{\frac{2}{3}}\beta\gamma c\tau_0, \tag{2.23}$$

where $\beta\gamma = p/m$ is its boost.

In the SM, particle lifetimes span many orders of magnitudes, from the very prompt Z boson $\tau_0 \sim 2 \times 10^{-25}$ s, to the stable electron. In Figure 2.4, the lifetimes and masses of several SM particles are shown. Particles are considered detector prompt if the distance between the particles' decay and production points is smaller than the spatial resolution of the detector. At the other extreme, particles with displacements far exceeding the physical size of the detector are considered detector stable. These two regimes are indicated with shaded areas. In this context, long-lived usually refers to particles producing sufficiently large displacement between the production and decay points that is resolvable by the detector, but not so high that the LLPs decay exclusively outside the detector acceptance. This range is $10\,\mu\mathrm{m} < \beta c\tau_0 < 10\,\mathrm{m}$ for CMS. It is worth noting that long-lived SM particles have a small mass, below 5 GeV. BSM models can also predict LLPs with a wide range of possible masses.

2.4.2 Long-lived supersymmetry

With the recent negative CMS results for "natural" SUSY [59], attention has turned to other variants of SUSY, which relax some of its attractive theoretical properties. This, in turn, often leads to long-lived superpartners. To this end, it is helpful to consider the concept of the next-to-lightest supersymmetric particle (NLSP). The decays from the NLSP to the stable LSP can often be suppressed, making the NLSP long-lived. In the following, some examples of such scenarios are highlighted.

Split supersymmetry

Models of split SUSY [60] solve some of the issues found in natural SUSY models, including the absence of experimental evidence of superpartners, while still providing gauge unification but give up on solving the hierarchy problem. Split SUSY predicts TeV-scale fermionic superpartners, while all other particles are assumed to be ultraheavy, hence the name "split". Split SUSY can give rise to long-lived NLSP gluinos, which decay to the LSP neutralinos $\tilde{\mathrm{g}} \longrightarrow \tilde{\chi}_1^0 \mathrm{q}\bar{\mathrm{q}}$ via highly virtual squark states, as shown in Figure 2.5. The proper gluino lifetime can be approximately parameterised as [61]:

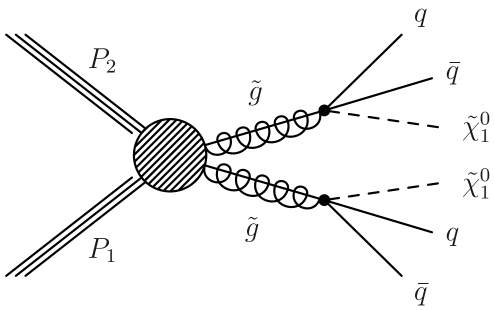


Figure 2.5: Pair production of long-lived gluinos and their subsequent decay via a squark to a pair of light-flavour quarks and the neutralino LSP.

$$\tau_0 \sim \left(\frac{m_S}{10^3\,\text{TeV}}\right)^4 \left(\frac{1\,\text{TeV}}{m_{\tilde{g}}}\right)^5 \times 10^{-4}\ \text{ns}, \tag{2.24}$$

as the squark propagator amplitude is inversely proportional to the gluino mass. If the SUSY breaking scale satisfies $m_S > 10^3$ GeV, the long-lived gluino hadronises into a colour-singlet state, known as an R-hadron [62], before decaying. The R-hadron is composed of the gluino and a light-flavour quark pair, for example, as $\tilde{g}u\bar{d}$. The interaction of the R-hadron with matter is described in terms of QCD interactions of the quark pair. In this model, the neutralino LSP remains a viable DM candidate. A summary sensitivity plot of different hadronic CMS long-lived gluino searches is shown in Figure 2.6.

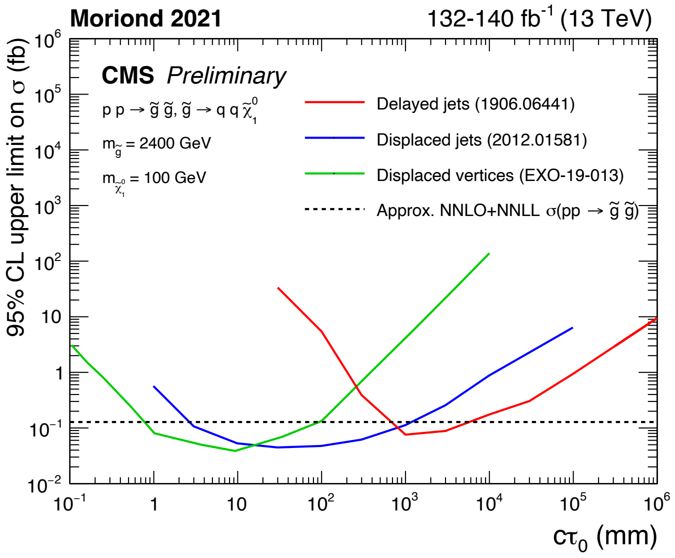


Figure 2.6: The 95% CL observed exclusion limit on the pair production cross-section of a long-lived gluino in a split SUSY model. The gluino $\tilde{g}$ and neutralino $\tilde{\chi}_1^0$ masses are assumed to be 2400 GeV and 100 GeV, respectively and the gluino is treated as a neutral, non-interacting particle [63].

Gauge-mediated symmetry breaking

In gauge-mediated symmetry breaking (GMSB) [64], the SUSY partners of SM fermions receive the dominant part of their masses via Yukawa interactions. The LSP is a very light gravitino, $\tilde{G}$, which can be produced in the decay of a long-lived gluino in association with a gluon, $\tilde{g} \longrightarrow g\tilde{G}$. As is the case for split SUSY, all other SUSY particles are assumed to be ultraheavy and decoupled from the interaction.

R-parity violating supersymmetry

Even though certain models introduce R-parity violation in SUSY, the R-parity violating (RPV) coefficients are generally expected to be small [65]–[67]. Hence, displaced decays of the superpartners are commonplace in RPV models. A long-lived top squark, for example, can form an R-hadron, before decaying to a b quark and a lepton, $\tilde{t} \longrightarrow b\ell$.

2.4.3 Long-lived heavy neutral leptons

An exciting signature occurs in the region of parameter space where HNLs are LLPs such that their decay products are sufficiently displaced with respect to the primary event vertex and distinguishable from it. For long-lived HNLs produced at the LHC, the phase space corresponding to the proper lifetime range 0.01 mm–10 m and the mass range of 1–20 GeV is the most relevant. Figure 2.7 shows the constant-$c\tau_0$ curves for Majorana HNLs in the HNL mass-coupling plane, assuming an equal ("democratic") coupling to all three active neutrino generations simultaneously.

Several direct searches for long-lived HNLs at the LHC have been performed recently. The CMS collaboration performed a search in events with three prompt leptons, with an interpretation provided for long-lived HNLs in the mass range of 1–20 GeV [48]. The ATLAS experiment did a search for prompt and long-lived HNLs, probing the HNL mixing with the second neutrino generation in the long-lived case by requiring a displaced vertex in the mass range of 4.5–10 GeV [47]. A search for HNLs in the mass range of 5–50 GeV was published by the LHCb collaboration [68]. These searches are used as a benchmark for the results obtained within this thesis.

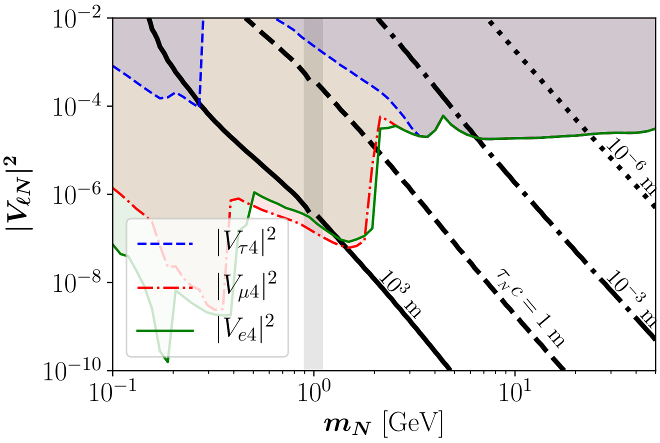


Figure 2.7: Contour lines indicate the HNL decay length $c\tau_0$ in the mass-coupling plane. Shaded areas correspond approximately to HNL scenarios excluded by various experiments. Figure taken from Ref. [69].

2.5 Summary

The SM has been a very successful theory, correctly describing most known physical phenomena. Several issues within the SM remain open, and leading candidate BSM theories predict new particles such as supersymmetric partners or HNLs. These particles can be long-lived and produce displaced signatures in the CMS detector, and much of the phase space for LLPs remains uncovered at the LHC.

Chapter 3

Experimental setup

The analysis presented in this thesis is based on pp collision data recorded with the CMS detector in 2016, 2017, and 2018 at a centre of mass energy of 13 TeV. In this Chapter, a brief overview of the European Organisation for Nuclear Research (CERN) accelerator complex and the LHC is provided. The major subdetectors of the CMS experiment are described: the solenoid magnet, the tracking system, the calorimeter systems, and the muon detector. Finally, the data acquisition system (DAQ) and trigger systems are discussed.

3.1 The Large Hadron Collider

The LHC is a superconducting hadron collider installed in a 26.7 km tunnel at CERN [70]. One of the main aims of the LHC is to search for BSM physics by colliding beams of protons or heavy ions (lead, xenon) with energies of up to 7 TeV or 2.76 TeV per nucleon, respectively. To achieve this, a multi-stage accelerator complex, depicted in Figure 3.1, is required. The following description refers specifically to the case of the acceleration of protons. The acceleration is done in several stages. Protons from hydrogen gas are first accelerated by LINAC2 to 50 MeV and injected into the Proton Synchrotron Booster (PSB). The PSB then accelerates the protons to 1.4 GeV. The following two stages are the Proton Synchrotron and the Super Proton Synchrotron (SPS), which achieve beam energies of 25 GeV and 450 GeV, respectively. Finally, the beams are injected into the LHC, with a maximum design beam energy of 7 TeV.

Each of the two LHC beams is designed to contain up to 2808 distinct bunches of protons, separated by 25 ns (around 7.5 m). The beams have up to 10^{11} protons. The beams are steered using superconducting dipole magnets, providing magnetic fields of up to

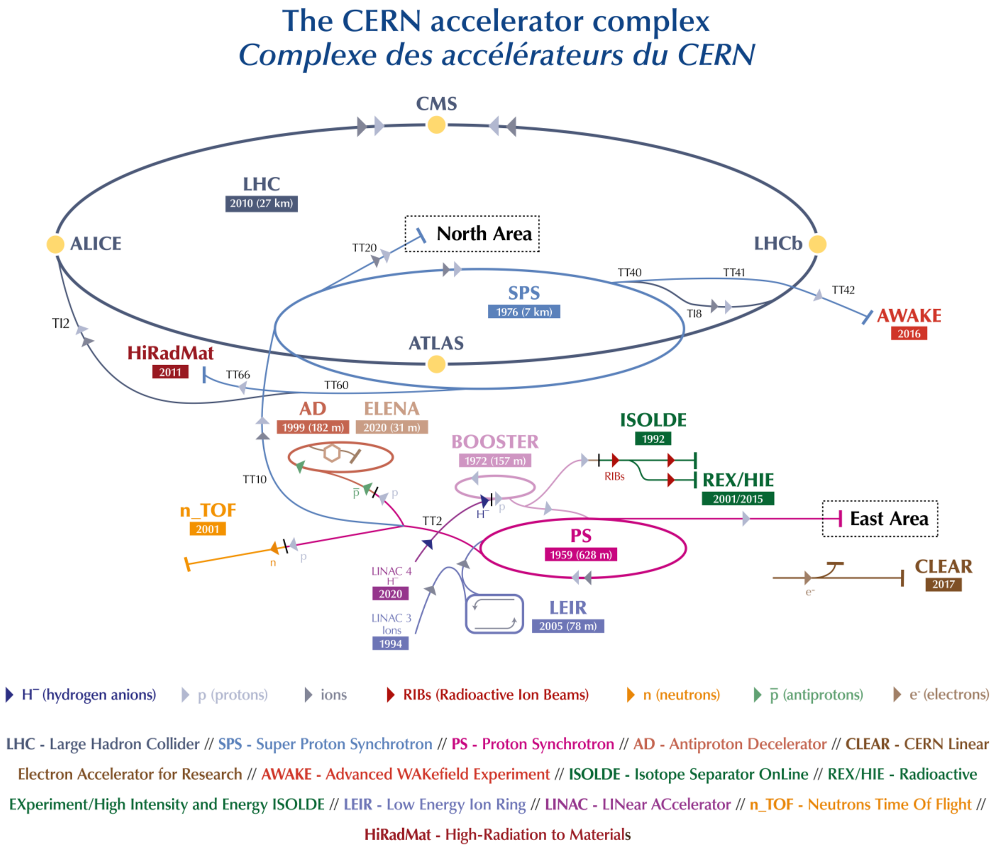

The CERN accelerator complex
Complexe des accélérateurs du CERN
CMS
LHC
2010 (27 km)
ALICE
North Area
LHCb
TT20
TT40
TT41
TT42
SPS
1976 (7 km)
AWAKE
2016
TI2
TI8
ATLAS
HiRadMat
2011
TT66
TT60
AD
1999 (182 m)
ELENA
2020 (31 m)
ISOLDE
1992
BOOSTER
1972 (157 m)
RIBs
REX/HIE
2001/2015
TT10
n_TOF
2001
p̄
p
H⁻
TT2
East Area
PS
1959 (628 m)
n
LINAC 4
2020
CLEAR
2017
e⁻
LEIR
2005 (78 m)
LINAC 3
Ions
1994
H⁻ (hydrogen anions)
p (protons)
ions
RIBs (Radioactive Ion Beams)
n (neutrons)
p̄ (antiprotons)
e⁻ (electrons)
LHC - Large Hadron Collider // SPS - Super Proton Synchrotron // PS - Proton Synchrotron // AD - Antiproton Decelerator // CLEAR - CERN Linear Electron Accelerator for Research // AWAKE - Advanced WAKefield Experiment // ISOLDE - Isotope Separator OnLine // REX/HIE - Radioactive EXperiment/High Intensity and Energy ISOLDE // LEIR - Low Energy Ion Ring // LINAC - LINear ACcelerator // n_TOF - Neutrons Time Of Flight // HiRadMat - High-Radiation to Materials

Figure 3.1: A schematic representation of the CERN accelerator complex. Image credit to CERN.

8 T. Each collision between bunches is referred to as a “bunch crossing”; the resulting interactions between protons and their products are referred to as an “event”. The number of events per second produced in the LHC collisions is:

$$\frac{dN}{dt} = L\sigma, \tag{3.1}$$

where σ is the cross-section for a particular process, while L is the instantaneous luminosity. The luminosity is determined purely by the particle beam parameters. For a Gaussian beam, it can be expressed as:

$$L = \frac{N_b^2 n_b f_{\text{rev}} \gamma_r}{4\pi \varepsilon_n \beta_*} F, \tag{3.2}$$

where N_b is the number of particles per bunch, n_b is the number of bunches per beam, f_{rev} is the revolution frequency, γ_r is the Lorentz factor, ε_n the normalised transverse beam emittance, β_* the beta function at the collision point, and F the geometric luminosity reduction factor at the interaction point (IP), arising because the beams are colliding at a non-zero crossing angle θ_c. Assuming two equal round beams, where the RMS bunch length is σ_z, and σ is the transverse beam size at the IP, F can be expressed as:

$$F = \left(1 + \left(\frac{\theta_c \sigma_z}{2\sigma*}\right)^2\right)^{-1/2}. \tag{3.3}$$

The integrated luminosity ($\mathscr{L}$) is a measure for the total amount of recorded data:

$$N = \int \frac{dN}{dt} dt = \sigma \mathscr{L}. \tag{3.4}$$

3.2 The CMS experiment

CMS [71], [72] (shown schematically in Figure 3.2) is one of the two general-purpose detectors at the LHC, designed to investigate pp collisions at high energies and instantaneous luminosities. The central feature of the CMS apparatus is a large superconducting solenoid of 6 m internal diameter, providing a magnetic field of 3.8 T in the z-direction over a radius of 3.15 m [73].

Multiple subdetectors are used to track, reconstruct and finally identify particles that traverse the detector in each bunch crossing. In increasing radial distance from the beam IP, these are the silicon pixel and silicon strip tracker, the electromagnetic calorimeter (ECAL), the hadronic calorimeter (HCAL), and the muon system. The goal of the CMS detector is to reconstruct particles as physics objects based on the detector signals they produce when traversing the active parts of the detector. By combining information from various subdetectors in an optimal way, CMS identifies the particle species as either muon, electron, charged or neutral hadron, or photon. This is called the PF algorithm and is further described in Section 4.1.

The CMS coordinate system has its origin placed at the IP and the z-axis pointing in the beam (longitudinal) direction. In the transverse (xy) plane, r is the radial distance, and ϕ is the azimuthal angle. Finally, θ is the polar angle relative to the beam axis; however, pseudorapidity is typically used instead:

$$\eta = -\ln \tan\left(\frac{\theta}{2}\right). \tag{3.5}$$

The pseudorapidity is defined differently than rapidity, y, which also depends on the particle's momentum, p. However, for relativistic particles, where $p \gg m$, the pseudorapidity converges to the rapidity, i.e. $\eta \approx y$. Differences in y are Lorentz-invariant and do not depend on longitudinal boosts of particles. The CMS detector has full coverage in ϕ, but not in pseudorapidity, where the coverage is $0 < |\eta| < 5$. The "forward" direction in a hadron collider experiment, such as CMS, is defined as being close to the beam axis, i.e. corresponding to high-values of η. The lack of complete coverage in pseudorapidity is due to the difficulty in measuring particles in the forward region, where radiation is high, and the presence of the beam pipe.

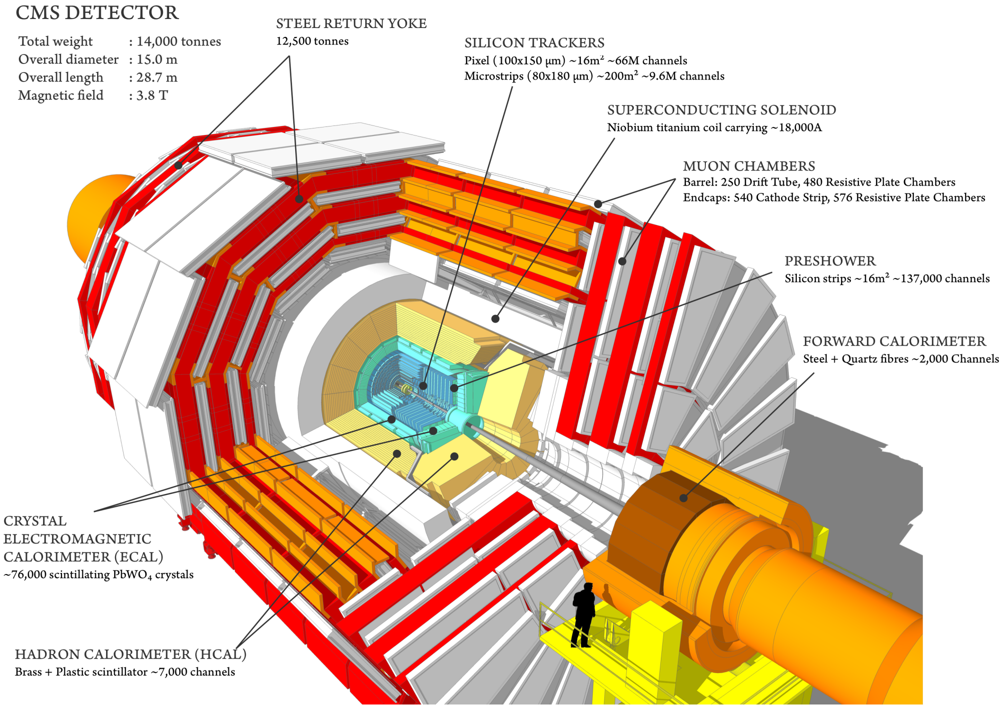

CMS DETECTOR
Total weight : 14,000 tonnes
Overall diameter : 15.0 m
Overall length : 28.7 m
Magnetic field : 3.8 T
STEEL RETURN YOKE
12,500 tonnes
SILICON TRACKERS
Pixel (100x150 μm) ~16m² ~66M channels
Microstrips (80x180 μm) ~200m² ~9.6M channels
SUPERCONDUCTING SOLENOID
Niobium titanium coil carrying ~18,000A
MUON CHAMBERS
Barrel: 250 Drift Tube, 480 Resistive Plate Chambers
Endcaps: 540 Cathode Strip, 576 Resistive Plate Chambers
PRESHOWER
Silicon strips ~16m² ~137,000 channels
FORWARD CALORIMETER
Steel + Quartz fibres ~2,000 Channels
CRYSTAL ELECTROMAGNETIC CALORIMETER (ECAL)
~76,000 scintillating PbWO₄ crystals
HADRON CALORIMETER (HCAL)
Brass + Plastic scintillator ~7,000 channels

Figure 3.2: A schematic of the CMS detector, with the physical dimensions, weight, and the number of components in various subdetectors specified. Image credit to CERN.

3.2.1 The inner tracking system

The inner tracking system [74] is located closest to the IP and serves to find charged particle trajectories ("tracks") that are bent in the magnetic field, as well as to estimate their points of origin ("vertices"). The tracker system is crucial to the scientific goals of CMS: it requires good momentum resolution for high-energy (100–1000 GeV) particles, good reconstruction efficiency for soft (few GeV) tracks, the ability to resolve nearby tracks, as well as excellent vertex resolution of ~ 10 µm.

Situated around the collision point, the cylindrical tracker, depicted in Figure 3.3, measures the momenta of charged particles in the range $|\eta| < 2.5$. The tracker system has a diameter of 2.5 m and a length of 5.8 m and uses silicon as the active material, with an effective area of 200 m^2. The basic working principle of the silicon devices is to apply a reverse bias voltage to a p-n junction to create a depletion zone free from charge carriers. A passing charged particle induces electron-hole pairs and creates a measurable signal that is digitised using readout electronics. These "hits" are used as building blocks in the reconstruction of particle tracks and vertices.

Directly around the IP and extending up to r of 20 cm, where the particle flux is maximal, finely segmented pixel detectors are used to perform three-dimensional hit position measurements, with a resolution of 10 (30) µm in the traverse (longitudinal) plane [75]. The pixel sensors are used to maintain reasonable hit occupancy and ensure good vertex resolution close to the IP. The pixel detector consisted of three barrel layers and two disks for each endcap during the 2016 data-taking period. In total, its 1440 modules cover an area of about 1 m^2 and have 66 million channels, with each module having a cell size of $100 \times 150\,\mu\mathrm{m}^2$. It was upgraded to include four barrel layers and three endcap disks for the 2017–2018 data-taking period, resulting in significantly better track reconstruction efficiency and resolution performance.

The remaining 199 m^2 is covered by larger silicon strip detectors further out, at $r > 20$ cm, where the particle flux drops. Some of the modules are placed in a "stereo" configuration by tilting modules with respect to one another, a setting that allows measuring two-dimensional hit positions. Overall, the strip modules are organised into four regions; two inner regions are called tracker inner barrel (TIB) and tracker inner disk (TID). These provide position measurements in the $r\phi$ plane with a resolution of 13–38 µm. The inner regions are surrounded by tracker outer barrel (TOB) and tracker endcaps (TECs), where the resolution drops to 18–47 µm. In total, the strip detector contains 9.3 million strips.

Notably, the tracker and most of the calorimeter system are accommodated inside the magnet, ensuring particle properties can be measured using these subsystems before any interaction with the magnet material. Nonetheless, the active tracker layers and their supporting services make up a non-negligible amount of material that particles have to traverse before reaching the calorimeters. Consequently, multiple scattering, photon conversion, electron bremsstrahlung, and hadron nuclear interactions are common and complicate the task of the track reconstruction algorithms.

For charged hadrons with $p_\mathrm{T} = 20$ GeV (momentum in the transverse plane), the momentum resolution is 1% and degrades with increasing p_T as the bending radius decreases, reaching calorimeter levels of accuracy at a few hundred GeV [75]. Charged hadrons of this energy are deviated by a few cm in the magnetic field. Due to the high strength of the magnet and the fine granularity of the calorimeter system, this can be resolved from a neutral hadron emitted in the same direction.

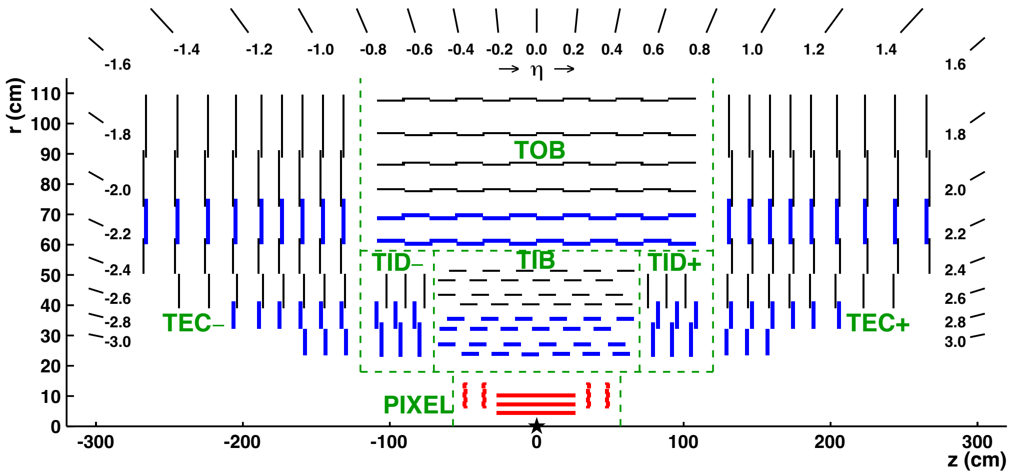


Figure 3.3: A schematic cross-section of the CMS tracker in the *rz* plane. Strip tracker modules that provide 2-D hits are shown by thin, black lines, while those permitting the reconstruction of hit positions in 3-D are shown by thick, blue lines. The pixel modules, shown by the red lines, also provide 3-D hits. Within a given layer, each module is shifted slightly in *r* or *z* with respect to its neighbouring modules, which allows them to overlap, thereby avoiding gaps in the acceptance. Figure taken from Ref. [75].

3.2.2 The electromagnetic calorimeter

The ECAL, shown in Figure 3.4, allows for high precision measurements of photon and electron positions and energies by measuring the properties of electromagnetic showers they induce in the active material. The ECAL [76] is made of high-density lead tungstate ($PbWO_4$) crystals, incorporated in a cylindrical electromagnetic barrel (EB) region ($|\eta| < 1.48$) containing 61200 crystals, which is enclosed by two electromagnetic endcaps (EEs) ($1.48 < |\eta| < 3.0$), each containing 7324 crystals. Lead tungstate is a very dense ($\rho = 8.3\,\mathrm{g cm^{-3}}$) and radiation-hard material with a short radiation length ($X_0 = 0.89\,\mathrm{cm}$). As the ECAL is around $25X_0$ deep, it almost entirely absorbs the electromagnetic showers produced by electrons and photons, containing more than 98% of energy for electrons and photons with energies of up to 1 TeV. Lead tungstate also has a small Molière radius that approximately matches the transverse crystal size. This combination of crystal properties results in a finely segmented, homogenous ECAL, with each crystal having a cross-section of 0.0174×0.0174 in the $\eta\phi$ plane, which corresponds to a rectangular wedge shape. This translates to a rectangular cross-section of $22 \times 22\,\mathrm{mm}^2$ at the front face and $26 \times 26\,\mathrm{mm}^2$ at the rear face in the barrel region. With such compact crystals, it is possible to resolve energy deposits from different particles separated by as little as 5 cm, which significantly helps to separate charged and neutral particles when considering the tracker information.

The particle signals are measured by collecting the scintillation light produced in the lead tungstate crystals. As the light yield of the crystals is relatively low, the signal is amplified by avalanche photodetectors in the barrel system and vacuum phototriodes in the endcaps. Another advantage of lead tungstate is its fast response time of 25 ns, corresponding to the typical bunch spacing, to collect 80% of the scintillation light.

To account for the reduced granularity and extra passive material at the endcaps, the ECAL is equipped with a electromagnetic pre-shower (ES) component covering $1.653 < |\eta| < 2.6$, a silicon detector with a few active layers, which helps to distinguish photons from neutral pion decays ($\pi_0 \longrightarrow \gamma\gamma$) from prompt photons.

The energy response of the EB has been determined in test beam measurements for electron energies between 20 and 250 GeV and can be expressed as [77]:

$$\left(\frac{\sigma}{E[\mathrm{GeV}]}\right)^2 = \left(\frac{2.8\%}{\sqrt{E[\mathrm{GeV}]}}\right)^2 + \left(\frac{12\%}{E[\mathrm{GeV}]}\right)^2 + (0.3\%)^2, \tag{3.6}$$

where the energy of incident electrons was determined by summing the energy deposits of a 3×3 crystal grid centred around a centrally hit crystal.

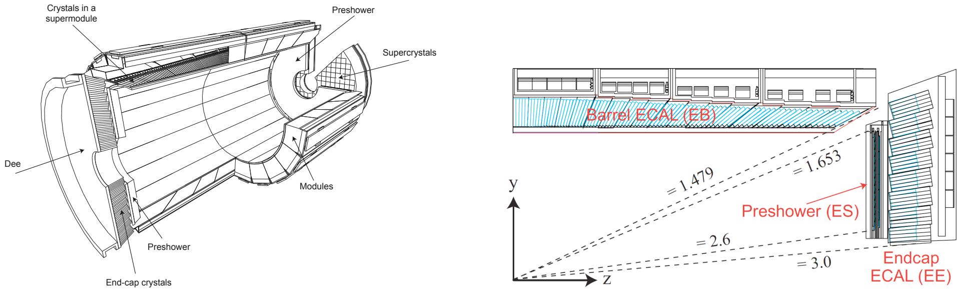


Figure 3.4: A schematic layout of the CMS ECAL, presenting the arrangement of crystal modules, supermodules, endcaps and the ES in front (left). Geometric view of one quarter of the ECAL (right).

3.2.3 The hadron calorimeter

The HCAL, shown in Figure 3.5, is designed to measure the hadronic properties of jets and provide good hermiticity to ensure that all particles, except for neutrinos, can be contained in the transverse plane of the CMS detector [78]. This allows for accurate measurement of the energy imbalance from neutrinos and potential new physics particles, such as non-interacting, long-lived DM candidates.

The HCAL is a sampling calorimeter that consists of alternating layers of non-magnetic brass absorber and plastic scintillator tiles, with a sampling fraction of about 7%. The HCAL surrounds and compliments the ECAL by absorbing showers initiated by charged hadrons in the ECAL. Traversing particles create hadronic showers in the brass layers and induce detectable light in the subsequent scintillators. The scintillation light, typically in the blue-violet region of the electromagnetic spectrum, is collected by wavelength-shifting fibres and translated and amplified by multichannel hybrid photodiodes proportionally to the magnitude of the energy deposits. The HCAL has both hadronic barrel (HB) ($|\eta| < 1.4$) and hadronic endcap (HE) ($1.3 < |\eta| < 3.0$) regions. The HCAL towers have a cross-section of 0.087×0.087 in $\eta\phi$ space for $|\eta| < 1.6$, and vary in size with $|\eta|$ after that. Because of the space constraints within the magnet cryostat, the HB thickness is limited to 5.8 hadronic interaction lengths (λ_I) at $\eta = 0$, increasing to 10 λ_I at $|\eta| = 1.2$. Hence, to catch the energy leakage from HB, layers of scintillators are placed outside the magnet: this is the hadronic outer “tail-catcher” (HO, $|\eta| < 1.26$), which increases the effective thickness of the combined HB+HE+HO system to over 10 λ_I. The HCAL is complimented by hadronic forward (HF) calorimeters, designed to endure high radiation levels and extending the coverage to $3 < |\eta| < 5$ in coarse calorimeter towers. The ECAL and the HCAL fraction of the energy deposited in each calorimeter varies non-linearly with energy, and, as a result, the raw energy measurements require substantial corrections. The combined energy resolution of the calorimeter system depends on both the ECAL and HCAL and was parameterised to be [79]:

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{85\%}{\sqrt{E[\mathrm{GeV}]}}\right)^2 + (7.0\%)^2. \tag{3.7}$$

3.2.4 The muon system

One of the main design goals of CMS is to achieve precise muon identification and energy reconstruction, as muons are key signatures of various SM and BSM processes. The muon system [81], located outside the magnet cryostat, covers $|\eta| < 2.4$ and is interleaved with iron structures containing and guiding the magnetic field. This return yoke also doubles as a filter and ensures only muons and neutrinos escape the inner part of the detector. A schematic depiction of the muon system is shown in Figure 3.6.

The basic principle of operation of the muon spectrometer is gas ionisation. Three types

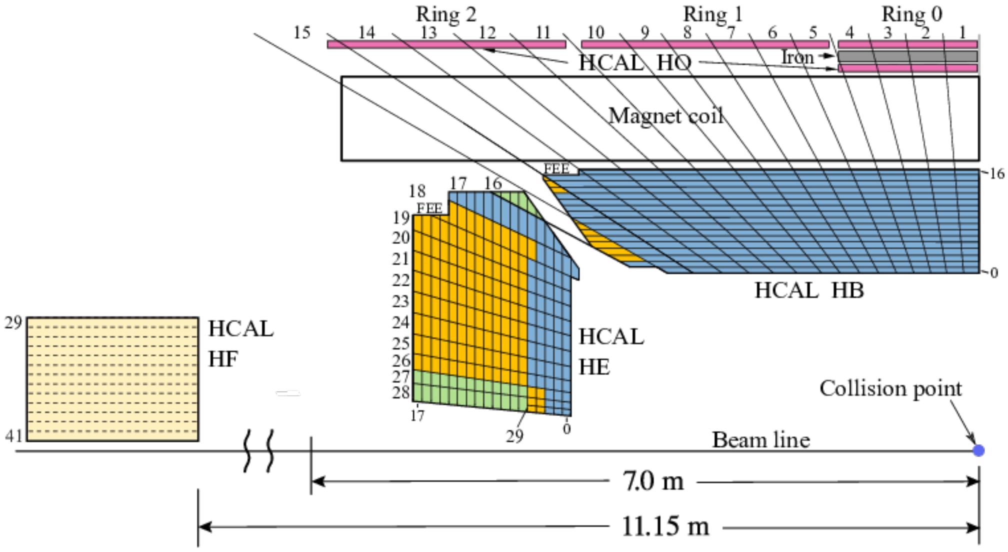


Figure 3.5: A schematic representation of the CMS HCAL. Figure taken from Ref. [80].

of separate ionisation chambers make up the muon system. As muons traverse the chambers, they ionise the gas, and the resulting freed electrons are collected using a strong electric field, which eventually causes a signal to be produced in the wires and strips in the chambers. Two system types are drift tubes (DTs), covering $|\eta| < 1.2$, and cathode strip chambers (CSCs), used for $0.9 < |\eta| < 2.4$. DTs are used in the central region where the muon rate is relatively low and the magnetic field is weak (0.4 T). With increasing η, however, the opposite conditions arise, and CSCs must be used. Hence, the muon spectrometer can be distinguished by three regions: barrel, ($|\eta| < 1.2$), two endcaps ($0.9 < |\eta| < 2.4$), and an overlap region ($0.9 < |\eta| < 1.2$). The third type of system, called resistive plate chambers (RPCs), acts as a complementary system, covering the range $|\eta| < 1.6$. RPCs offer a fast response and are crucial in associating signals to the correct bunch crossing for triggering; however, they have a much coarser position resolution than the DTs or CSCs.

The hit and track segment reconstruction efficiency for traversing muons is 94–97% [82]. Using 13 TeV measurements with cosmic ray data and combining both the inner tracker and muon systems, the momentum resolution was measured at 1% for muons of p_T = 10 GeV, decreasing to 6% at p_T = 1000 GeV.

3.2.5 The data acquisition and trigger system

At the LHC, the current bunch spacing is 25 ns, meaning collisions happen at a rate of 40 MHz. An entire event takes about a megabyte of disk space, so it is impossible to write out all events to disk due to bandwidth and storage limitations. Furthermore, the cross-sections of “interesting” processes, such as EW processes and potential BSM physics, are

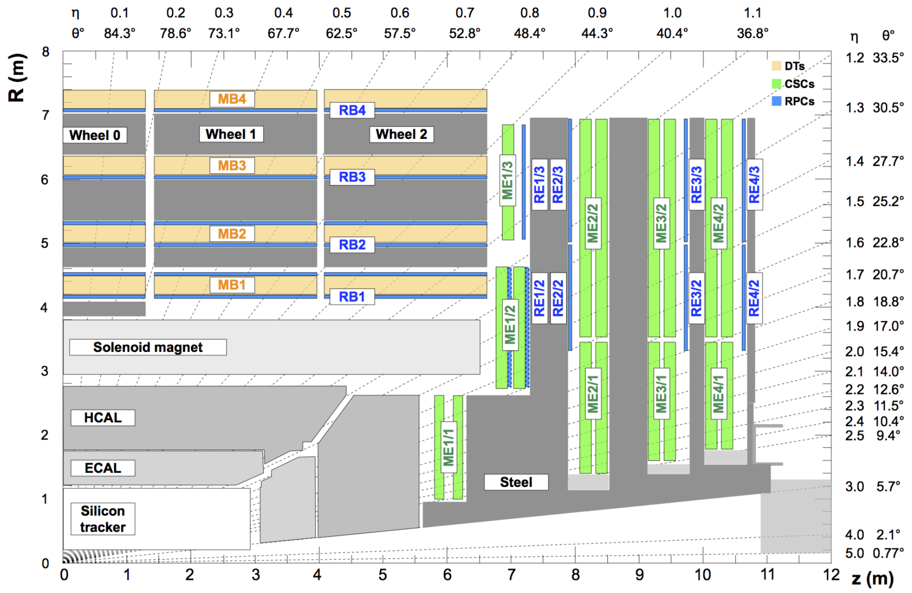

η 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0 1.1
θ° 84.3° 78.6° 73.1° 67.7° 62.5° 57.5° 52.8° 48.4° 44.3° 40.4° 36.8°
η θ°
1.2 33.5°
1.3 30.5°
1.4 27.7°
1.5 25.2°
1.6 22.8°
1.7 20.7°
1.8 18.8°
1.9 17.0°
2.0 15.4°
2.1 14.0°
2.2 12.6°
2.3 11.5°
2.4 10.4°
2.5 9.4°
3.0 5.7°
4.0 2.1°
5.0 0.77°
R (m)
z (m)
0 1 2 3 4 5 6 7 8
0 1 2 3 4 5 6 7 8 9 10 11 12
DTs
CSCs
RPCs
Wheel 0
Wheel 1
Wheel 2
MB4
MB3
MB2
MB1
RB4
RB3
RB2
RB1
ME1/3
ME1/2
ME1/1
RE1/3
RE2/3
RE1/2
RE2/2
ME2/2
ME2/1
ME3/2
ME3/1
RE3/3
RE3/2
ME4/2
ME4/1
RE4/3
RE4/2
Solenoid magnet
HCAL
ECAL
Silicon tracker
Steel

Figure 3.6: A quadrant of the CMS detector. The IP is at the lower-left corner. The locations of the various muon stations and the steel flux-return disks (dark areas) are shown. The DTs stations are labelled muon barrel (MB) and the CSCs are marked as ME. RPCs are mounted in both the barrel and endcaps of CMS, labelled RB and RE, respectively. Figure taken from Ref. [82].

very low compared to that of the inelastic scattering of partons. Inelastic pp scattering has a cross-section of 69 mb and results in an event rate of ~ 1 GHz, corresponding to about 25 simultaneous pp interactions per bunch crossing, referred to as “pileup”.

As a result of these enormous rates, no meaningful data collection can occur without a trigger in place, which performs coarse data analysis and selectively stores events via a two-tier structure [83]–[85]. The first stage is the Level-1 (L1) trigger that consists of hardware and firmware and reduces the event rate to 100 kHz. This rate is low enough for processing by a processor farm, known as the high-level trigger (HLT), further reducing the rate to about 1 kHz. The trigger system plays a core role in the DAQ system, which stores complete event data to disk for each event passing the L1 and HLT decisions.

The primary purpose of the L1 system is to discard almost all simple QCD-induced multijet events, which are produced abundantly, while at the same time allowing a low-enough threshold for precision SM measurements and potential BSM signals. The L1 trigger temporarily stores events in the detector front-end pipelines, which store information for 128 bunch crossings. The event selection algorithms must therefore operate within a latency of < 4 μs. Due to bandwidth limitations, the L1 trigger is only capable of using coarse information from trigger primitives (TPs) from calorimeter and muon systems, while the tracker is read out only if the trigger provides a positive response. Event selection is performed in several stages based on combined detector information, before deciding to keep interesting events in the global trigger (GT).

Isolated high-p_T leptons are key objects in the triggering process, as they can indicate the production of EW bosons and are thus crucial for several studies. Isolation criteria can be imposed to veto leptons originating from heavy quark (b or c) jets. In addition to triggering on isolated leptons, retaining events with energetic jets is essential as these can result from heavy new particle decays. Good efficiency for high-energy jets and energy sum quantities, such as missing transverse momentum, $p_\mathrm{T}^\mathrm{miss}$, or the scalar sum of jet energies, H_T, at the L1 trigger is a critical ingredient in searches for heavy particles, which lead to significant H_T and $p_\mathrm{T}^\mathrm{miss}$.

A crucial component of the trigger is a field-programmable gate array (FPGA) — an integrated circuit that can be configured by the user after manufacturing. This allows changing the trigger menu as needed to keep up with the evolving CMS physics programme and the detector performance. Several FPGAs, interconnected by high bandwidth optical links, make up the trigger cards. Trigger boards process and concentrate the data in several stages for final decision-making in the GT. Currently, the GT allows for up to 512 algorithms to cope with the broad CMS physics programme. A single algorithm could require, for example, to only accept events above a specific missing transverse energy threshold.

Events passing the L1 trigger are transferred to the HLT for further filtering. Its trigger decision calculation is based on more sophisticated processing that includes information from the inner tracking system. The HLT software is efficiently structured into paths that trigger events with different topologies and physics objects of interest, yet share common information to increase performance. An event is stored as soon as it is accepted by at least one trigger path or discarded once it is established that no path will accept it.

For specific certain algorithms which select standard physics processes, only a fraction of events passing the trigger criteria can be retained to restrict the output rate. This is referred to as prescaling. For a given algorithm, the prescale is adjusted during data taking as required to match the given instantaneous luminosity.

3.2.6 Data-taking performance

For the LHC data taking period of 2015–2018, referred to as Run 2 and pertinent for the data analysis discussed in this thesis, beam energy of 6.5 TeV was achieved, i.e. a centre of mass energy of $\sqrt{s} = 13\,\mathrm{TeV}$. During Run 2, the LHC achieved $L = 2 \cdot 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$, which exceeded the design value of $1 \cdot 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ through various improvements. The 2015 data set corresponds to a negligible fraction of the total Run 2 integrated luminosity and hence was not analysed in this thesis. The total integrated luminosity was measured using the pixel cluster counting method to be 137 fb^{-1} for the 2016–2018 pp collision data

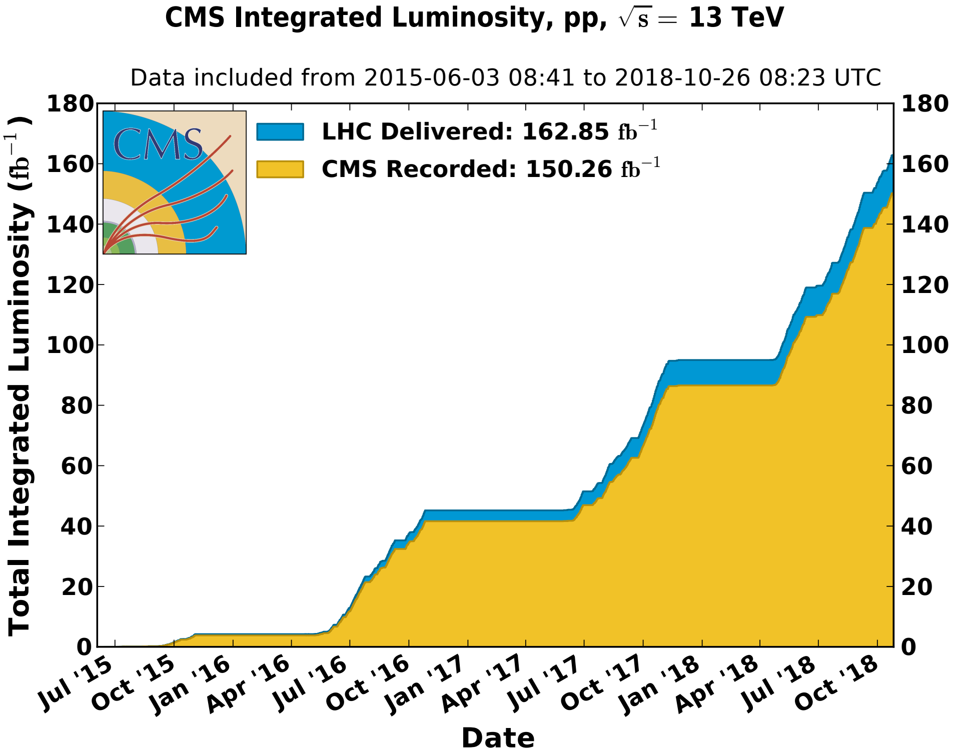


Figure 3.7: Run 2 pp collision data (2015–2018) delivered by the LHC and recorded by CMS.

set [86]–[88]. As can be seen from Figure 3.7, the overall data-taking efficiency was over 92%. The average pileup was found to be 23 (32) during the 2016 (2017–2018) data taking period, which poses a substantial challenge to particle reconstruction algorithms.

Chapter 4

Event reconstruction

This Chapter describes the identification and reconstruction of particles and their properties based on the signatures they leave in the detector during an event. In CMS, complementary information from all subdetectors is combined using the PF algorithm [89], the name of which alludes to the flow of (particle) energy throughout the detector. The PF algorithm's primary goal is to provide a global event description by identifying all stable particles produced during the collision. CMS is particularly suited for PF reconstruction due to its fine spatial granularity, which minimises the merging of signals from different particles. The primary particle type (electron, photon, muon, charged or neutral hadron) and properties (energy and direction) are identified using the PF algorithm, and physics objects are defined starting from the PF candidates, such as jets and missing energy. This Chapter aims to describe the reconstruction of the essential physics objects used within the analysis and various corresponding corrections used to describe the said objects' properties more accurately. The reconstruction of physics objects is two-fold: energy deposits from the detector are initially reconstructed as tracks, vertices, and calorimeters clusters. These are then used in PF as building blocks for physics objects in global event reconstruction.

The event selection of the search for long-lived HNLs described in this thesis requires the presence of at least two leptons, one of which can be displaced, and at least one jet. As a result, this Chapter places a particular emphasis on the reconstruction of muons, electrons, and jets. As the CMS detector has not been designed to target displaced signatures arising from LLPs explicitly, the reconstruction efficiency of various algorithms is studied to ensure they are suitable for an LLP search. Missing energy is also of particular importance as it used to train a multivariate (MVA) classifier. These objects are also crucial for defining control regions (CRs) used to train the first iteration of the DNN jet tagger, targetting long-lived split SUSY gluinos.

4.1 Particle flow event reconstruction

4.1.1 Track reconstruction

With typical instantaneous luminosities achieved at the LHC, the CMS tracker is expected to be traversed by about 1000 charged particles at each bunch crossing. In this challenging environment, maintaining a high track-finding efficiency, while ensuring that the tracking software runs fast enough in the HLT, is challenging. To this end, a different version of the tracking algorithm, which employs fewer iterations, is used at the HLT. Furthermore, it is crucial to ensure that the fraction of “fake” reconstructed tracks is small, where “fake” refers to tracks resulting from a combination of unrelated hits or a genuine particle trajectory that is poorly reconstructed.

The first step of the reconstruction process, called local hit reconstruction, involves clustering zero-suppressed signals left by traversing particles in pixel and strip modules into hits, and estimating the cluster positions and their corresponding uncertainties. The hit efficiency, defined as the probability to find a cluster in a given silicon sensor that a charged particle has traversed, is typically >99% [75]. The hits can then be used to estimate the charged particle position and momentum parameters.

In the magnetic field inside the solenoid, charged particles follow helical trajectories. Twelve iterations of a combinatorial track finder algorithm are used to reconstruct the trajectory properties. In each iteration, hits assigned to tracks in the previous iteration are removed from the hit collection to reduce the combinatorial complexity. Such an approach is advantageous because it removes easy-to-reconstruct tracks in the first few iterations. This leaves more challenging classes, such as soft or significantly displaced tracks, for the later stages.

The first four iterations aim to find prompt tracks, i.e. tracks originating close to the pp IP, while iterations 5–9 recover tracks outside the beam spot. Iterations 10–12 are unique in that they target tracks of high-p_{T} jets and muons. Each of the twelve iterations proceed in four stages. First, seed generation provides initial track candidates found using a few hits. A seed is the initial estimate of the helical particle trajectory parameters and their corresponding uncertainties. The second step is based on a Kalman filter [90], where the seed trajectories are extrapolated along the expected flight path of the particle to search for hits that could be assigned to the track candidate. The third step involves fitting the assigned hits to provide the best possible estimate of the track parameters. Once the final tracks are obtained, fake tracks are reduced by imposing quality requirements. Tracks are selected based on the number of layers that have hits, the goodness-of-fit, and how compatible they are with originating from a primary interaction vertex. The main difference between the generations is in the seed generation configuration and final track selection steps. At the end of the track selection, the tracks found by each of the twelve iterations are merged into a single collection [75], [91].

The tracking performance as a function of track p_T and simulated vertex radius is shown for Run 2 data-taking conditions in Figure 4.1 for simulated tt̄ events. It can be seen that later iterations of the tracking algorithm are critical to achieving good efficiency for displaced tracks which is crucial for LLP searches. In addition, the tracking efficiency is >99% for isolated muons with $1 < p_T < 100$ GeV. However, the efficiency is around 90% at $p_T = 10$ GeV for charged pions, with tracking inefficiencies arising due to nuclear interactions experienced by charged hadrons. For electrons, a large fraction of energy is lost via bremsstrahlung before reaching the end of the active tracker material, leading to substantial inefficiencies when using the default track reconstruction algorithm. Therefore, an additional dedicated electron reconstruction algorithm is used to improve the tracking efficiency, as described in Ref. [75].

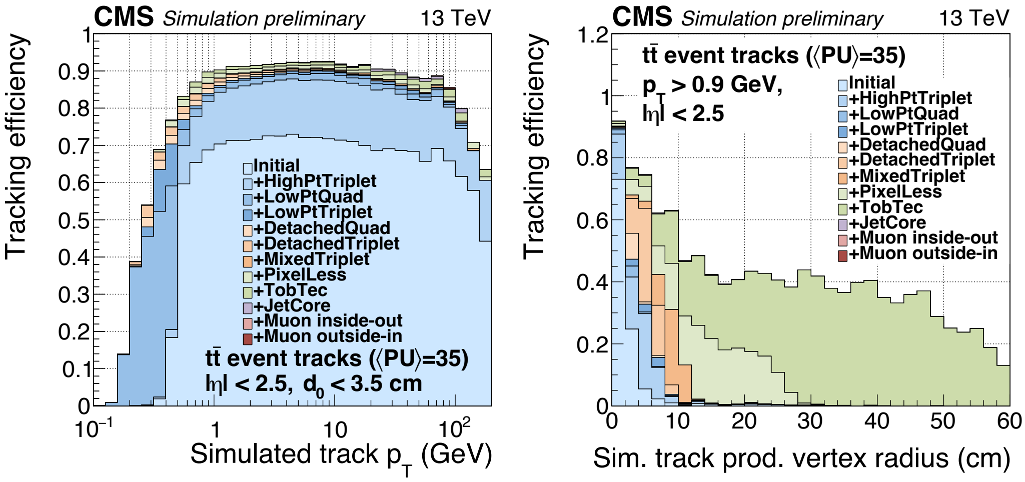


Figure 4.1: Track reconstruction efficiency for each of the twelve tracking iterations (denoted in the legend) as a function of simulated track p_T (left) and production vertex radius (right). Figures taken from Ref. [91].

4.1.2 Vertex reconstruction

In a high pileup environment, the main goal of the vertexing algorithm is to identify the primary vertex (PV) and all tracks associated with it, mask particles originating from additional collision (pileup) vertices, and identify secondary vertexs (SVs), a crucial feature in searches for LLPs.

The PV reconstruction proceeds in three steps: first, tracks consistent with being produced at the IP are identified by imposing various quality requirements. The selected tracks are then clustered based on their z-coordinates at their point of closest approach to the centre of the beam spot to identify all pp interactions produced in the same bunch crossing. The clustering is performed using a deterministic annealing algorithm [92], which performs a stochastic search for minima via a Markov chain MC method. The algorithm is analogous to the scenario of a physical system undergoing a series of temperature reductions and gradually approaching a final state of minimal energy. Finally, candidate vertices identified by the deterministic annealing procedure that contain at least two tracks are fitted to determine the position of each vertex, the corresponding uncertainty, as well as various fit quality indicators [93].

The resulting vertex with the highest p_{T}^2 sum of the associated tracks is taken to be the leading PV and is used for physics object identification and reconstruction, while all other PVs are treated as pileup interactions. For vertices with many tracks, characteristic of interesting events, the PV reconstruction efficiency is close to 100%, while the typical PV resolution is about 10–12 µm in each of the three spatial dimensions [75].

4.1.3 Cluster reconstruction

The calorimeters are used to reconstruct the energies of incident particles from shower deposits made in the active material. Energy clusters left in the calorimeter system are the only way in CMS to reconstruct the energies of photons and neutral hadrons. These are identified in the PF algorithm by finding energy clusters not associated with the extrapolated trajectory of a charged particle's track.

In the ECAL, for example, clustering algorithms are used to sum together energy deposits in adjacent crystals belonging to the same electromagnetic shower. The clustering algorithm starts by finding seed crystals, corresponding to a local energy maximum above a predefined threshold, which is generally 2 or 3 times bigger than the electronic noise expected for these crystals. Adjacent crystals with recorded energy deposits are iteratively added to the bare cluster as long as they are not assigned to a different cluster already. The resulting bare clusters are then merged to form a supercluster extended in ϕ to recover the radiated energy due to electron bremsstrahlung or photon conversions. The energy measurement is provided by summing all the energy deposits recorded within the supercluster's crystals. The energy measured by the superclustering algorithm is subject to losses for several reasons: electromagnetic shower energy in the ECAL can be lost through shower leakages, in intermodular gaps or dead crystals. An MVA regression technique, described in Ref. [94], is used to correct the supercluster energy for these energy deposits.

4.1.4 Muon tracking

The muon system allows the identification of muons with high efficiency. It provides a very high purity measurement, as most other particles (except neutrinos) are fully absorbed within the calorimeters, whereas muons only leave minor energy deposits there.

Hits within each DT and CSC detector are clustered to form seed track segments, used for pattern recognition in the muon spectrometer. The algorithm then finds all DT, CSC, and RPC hits along the muon trajectory. By fitting all hits, a so-called standalone-muon track is obtained. The muon spectrometer information is later combined with the inner tracker in the PF algorithm to obtain a more precise energy measurement.

4.2 Physics object reconstruction

One particle can, in general, produce several PF elements in different subdetectors. Hence, the first step of PF is a link algorithm, matching particle signals produced in different subdetectors:

1. A link between tracks in the central tracker and a calorimeter cluster is made by extrapolating the track direction from the last hit in the tracker to the ECAL and the HCAL within a given geometric acceptance.
2. Tangents to the electron trajectory are extrapolated to the ECAL from the track and tracker material's intersection points to collect photons resulting from electron bremsstrahlung.
3. Links are formed between ECAL and HCAL clusters.
4. Finally, links between tracks in the central tracker and the muon subsystem are established.

The linking algorithm produces PF blocks of linked elements. The reconstruction and identification sequence then proceeds for each block as follows:

1. Muon candidates are identified by linked tracker and muon chamber hits. The tracks, as well as any associated calorimeter clusters, are removed from the blocks.

2. The PF electron seed candidates are formed by tracks linked to ECAL clusters, and isolated photon seed candidates are likewise considered if there is no matching electron track. In both cases, a requirement of no more than 10% of the supercluster energy in a surrounding HCAL region, defined as all clusters within 0.15 in the $\eta\phi$ plane, is imposed. All selected tracks and clusters in the PF block are masked from further processing. Bremsstrahlung photons associated with the electron track are also removed and added to the electron energy.

3. At this stage, the remaining particles are products of jet fragmentation and hadronisation. Calorimeter clusters not matched to any tracks are reconstructed as PF neutral hadrons and (non-isolated) photons. The precedence in the ECAL is given to photons over neutral hadrons. Finally, tracks that are linked to compatible ECAL and HCAL clusters form PF charged hadrons.

The particles identified and reconstructed by the PF algorithm are used in the analysis to build physics objects. In the presence of no pileup interactions, these objects would correspond to the collision's stable final state products (except for neutrinos). Reconstruction and identification of isolated electrons and muons is a crucial part of the CMS detector design, as this is the key signature of the production of W/Z/H bosons. For a given particle, relative isolation is a dimensionless quantity calculated as the transverse momentum sum of all other particles present in a cone of a given radius ΔR_{max} in the $\eta\phi$ plane around that particle, relative to the p_{T} of the particle itself:

$$I_{\text{rel}} = \frac{\sum_i p_{\text{T}}^i}{p_{\text{T}}}, \quad \Delta R(\mathbf{p}, \mathbf{p}^i) < \Delta R_{\text{max}} \forall i. \tag{4.1}$$

In the following Sections, the reconstruction of each type of object is explored in more detail.

4.2.1 Muon reconstruction

Muon identification in the PF algorithm exploits the information from the inner tracking system tracks and the muon spectrometer, resulting in "global" and "tracker" muon classes. Each standalone-muon track is matched to an inner track if the parameters of the two tracks propagated onto a common surface are compatible to obtain global muons. The inner tracks and standalone-muon tracks are combined and fit to form a global-muon track. The combined fit improves the momentum resolution with respect to the tracker-only fit, especially for $p_{\text{T}} > 200\,\text{GeV}$. Tracker muons, on the other hand, are found by extrapolating inner tracks to the muon system. If the extrapolated track matches at least one segment, the muon is considered a tracker muon. About 99% of the muons produced within the muon system's geometrical acceptance are reconstructed either as a global muon or a tracker muon. All muons considered in this analysis are required to be either tracker or global muons.

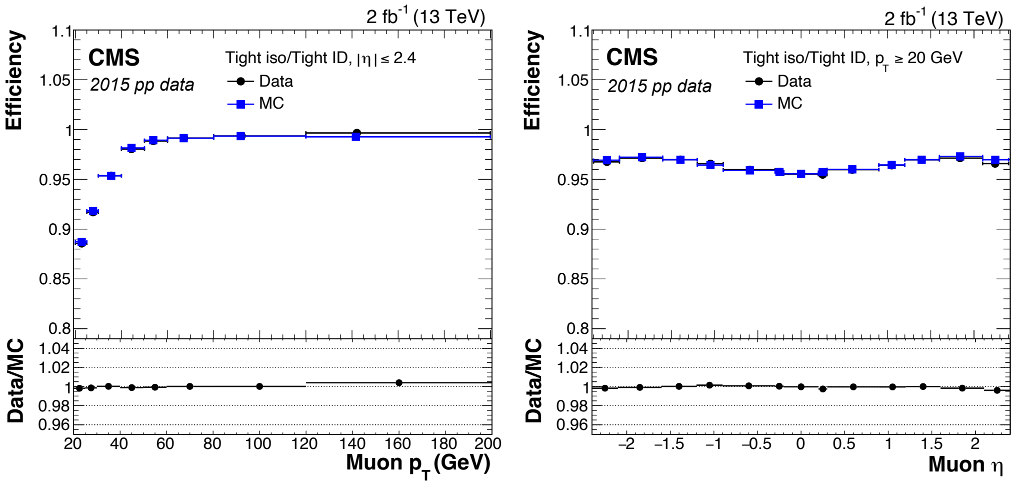


Figure 4.2: Efficiency for the tight PF muon isolation and identification working point versus p_T (left) and versus η for muons with $p_T > 20$ GeV (right) for the Run 2 data-taking conditions. Figure taken from Ref. [82].

Charged hadrons can be misreconstructed as muons if some of the hadron shower penetrates through to the muon system (punch-through). Different identification criteria can be applied to the muon tracks to sufficiently reduce the rate of fake muon tracks.

In the analysis, several types of muon candidates are considered. To preferentially target genuine muons produced in prompt decays of W and Z bosons, tight muon identification and isolation requirements are imposed. For tight muon candidates, a global muon track with $\chi^2/\text{ndof} < 10$ is required using at least one hit from the muon system and at least six hits from the inner tracker, of which at least one is in the pixel detector. Additionally, muon track segments are required in at least two muon stations, which suppresses accidental association of tracks to uncontained hadronic showers spilling into the muon system. In addition, the muon track needs to originate from the vicinity of the PV within $|d_{xy}| < 0.01$ cm and $|d_z| < 0.05$ cm. The requirement is used to suppress muons from decays of b and c hadrons with minimal loss of efficiency for prompt muons.

Tight muon candidates are required to be isolated with a PF-based combined relative isolation of $I^{\mu}_{\text{rel.}} < 15\%$. The relative "delta-$\beta$" isolation for muons is defined as:

$$I^{\mu}_{\text{rel.}} = \frac{I_{\text{ch.had.}} + \max\left(0,\ I_{\text{neut.had.}} + I_{\gamma} - \beta \cdot I_{\text{PU}}\right)}{p^{\mu}_{T}}, \qquad \beta = \frac{1}{2}, \tag{4.2}$$

where $I_{\text{ch.had.}}$, $I_{\text{neut.had.}}$, I_{γ} denote the summed energy deposits of charged hadrons, neutral hadrons, and photons within a cone of $\Delta R < 0.4$ around the muon direction, respectively. Half of the contribution by tracks associated with pileup vertices, I_{PU}, is taken to estimate the neutral energy deposit from pileup, and used as a correction. The efficiency of tight criteria, shown in Figure 4.2, was measured using tag-and-probe methods [82] and is generally at least 90%.

On the other hand, only loose muon identification requirements are employed to reconstruct displaced muon candidates; no requirement on the muon isolation nor the loose muon track's compatibility with the PV is employed. Loose muons are required to have a transverse momentum of at least 3 GeV within $|\eta| < 2.4$. The efficiency for a displaced muon from an HNL decay to be reconstructed a PF muon and subsequently identified using the loose criteria, is shown in Figure 4.3. The reconstruction efficiency is close to 95% for prompt muons with $p_{\text{T}} > 5\,\text{GeV}$, decreasing to 80% at a displacement of 10 cm as the outer layers of the inner tracking system are being reached. On the other hand, the identification efficiency is close to 100% due to the muon spectrometer.

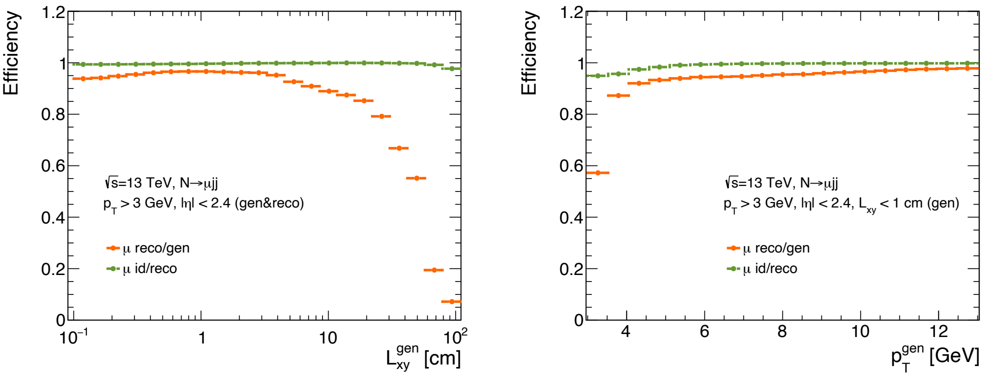


Figure 4.3: Muon from HNL decay reconstruction and identification efficiency as a function of generator-level lab frame displacement (left) and generator-level p_{T} (right).

4.2.2 Electron reconstruction

Electron candidates obtained from tracks linked to calorimeter superclusters must satisfy specific identification criteria. These are two-fold: one requires the electron to satisfy a sequential set of requirements ("cut-based"); the other is based on an MVA approach.

As is the case for muons, one of the essential quantities which allows identifying electrons originating from W and Z bosons is the relative isolation based on effective areas, which is defined as:

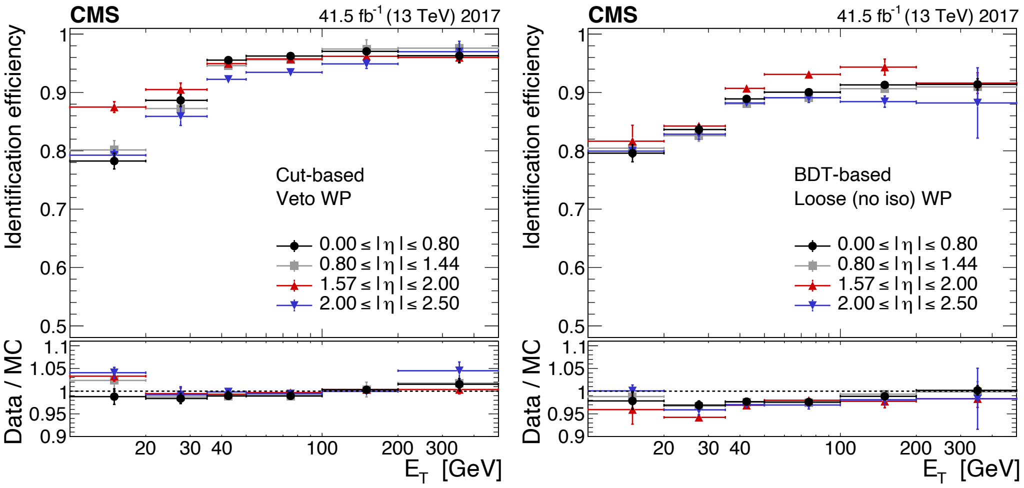


Figure 4.4: Electron identification efficiency measured in data (upper panels) and data-to-simulation efficiency ratios (lower panels), as a function of the electron energy, for the cut-based identification veto working point (left) and the BDT-based (without isolation) loosest working point (right). Figure taken from Ref. [94].

$$I^{\mathrm{e}}_{\mathrm{rel.}} = \frac{I_{\mathrm{ch.had.}} + \max\left(0,\ I_{\mathrm{neut.had.}} + I_{\gamma} - \rho \cdot A_{\mathrm{eff.}}\right)}{p_{\mathrm{T}}^{\mathrm{e}}}, \tag{4.3}$$

where the charged and neutral energy deposits within a cone of $\Delta R < 0.3$ around the electron candidate are used. A pileup correction is applied whose contribution is estimated as $I_{\mathrm{PU}} = A_{\mathrm{eff}} \cdot \rho$, where A_{eff} denote η-dependent areas and ρ the median of the transverse energy density in $\delta\eta \times \delta\phi$, calculated from all PF particles within an event. Non-prompt electrons are rejected based on exploiting shower-shape variables. For instance, even though it may not be possible to fully resolve a pair of photons decaying from a neutral hadron, a wider shower profile is still expected, on average, compared with a single incident electron. Some typically used handcrafted variables include $E_{\mathrm{had.}}/E_{\mathrm{EM}}$ i.e. the hadronic over electromagnetic energy ratio as well as $\sigma_{i\eta i\eta}$; defined as the second moment of the log-weighted distribution of crystal energies in η. Several variables related to the electron track properties are also employed. One such variable is $E_{\mathrm{SC}}^{-1} - p^{-1}$, where E_{SC} is the supercluster energy, and p is the track momentum. Another pair of angular variables that are used is $\Delta\eta^{\mathrm{in}}_{\mathrm{seed}}$, defined as the difference in η of the seed cluster and the extrapolated track position as well as $\Delta\phi_{\mathrm{in}}$, which uses the supercluster energy-weighted position in ϕ instead of the seed cluster. An important source of backgrounds to prompt electrons arises from secondary electrons produced in photon conversions in the tracker material. The presence of missing hits in the first tracker layer is exploited to reject this background. The exact combination of these requirements depends on the electron η and is described in detail in Ref. [94].

The MVA approach combines the variables mentioned above and is expanded by several others in a BDT. Two versions of the BDT are trained, with and without including the isolation variable, to allow maximum flexibility.

The efficiencies of the MVA-based and cut-based approaches were measured using tag-and-probe methods and are shown in Figure 4.4. Considering also the false positive rate of misidentified electrons, the MVA-based identification substantially outperforms the cut-based approach, and, for that reason, is used in this analysis to target prompt electrons from vector boson decays. However, the cut-based ID can be more easily customised by adjusting certain requirements; for this reason, it is used to ensure sufficiently high efficiency for displaced electron candidates resulting from long-lived HNL decays. Loose electrons are required to have a transverse momentum of at least 5 GeV within $|\eta| < 2.4$ and fulfil a set of cut-based criteria listed in Table 4.1. The efficiency for a displaced electron from an HNL decay to be reconstructed as a PF electron and subsequently identified using these criteria is shown in Figure 4.5. The efficiency to reconstruct a prompt electron of $p_{\mathrm{T}} > 10$ GeV is around 80%, dropping to 50% for electrons of $p_{\mathrm{T}} = 5$ GeV or $L_{xy} = 10$ cm. Electrons, especially low-p_{T} ones, exhibit significantly worse reconstruction and identification performance compared with muons. This is not surprising as muon reconstruction benefits heavily from the identification capabilities of the muon chambers.

Table 4.1: Cut-based loose electron identification criteria.

Quantity	EB	EE
$\sigma_{i\eta i\eta}$	< 0.0112	< 0.0425
$\Delta\eta^{\mathrm{in}}_{\mathrm{seed}}$	< 0.00377	< 0.00674
$\Delta\phi_{\mathrm{in}}$	< 0.0884	< 0.169
$E_{\mathrm{had.}}/E_{\mathrm{EM}}$	$< 0.05 + 1.16/E_{\mathrm{SC}} + 0.0324\rho/E_{\mathrm{SC}}$	$< 0.0441 + 2.54/E_{\mathrm{SC}} + 0.183\rho/E_{\mathrm{SC}}$
$E_{\mathrm{SC}}^{-1} - p^{-1}$	< 0.193	< 0.111
conversion veto	yes	yes

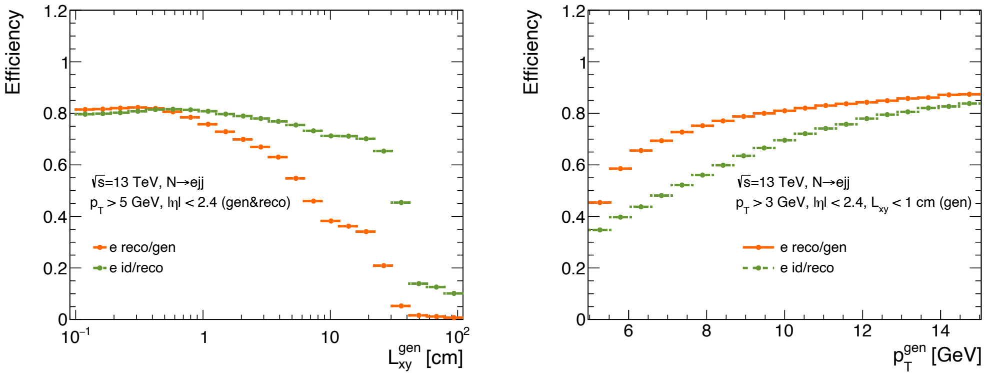


Figure 4.5: Electron from HNL decay reconstruction and identification efficiency as a function of generator-level lab frame displacement (left) and generator-level p_{T} (right).

4.2.3 Jet reconstruction

Collimated jets of particles are produced by the hadronisation of partons (quarks and gluons) resulting from deep inelastic pp collisions. By combining all particles produced in this process, the four-momentum of the original parton can be obtained. For the jet properties to be theoretically tractable, the jet clustering algorithm needs to satisfy the requirements of infrared safety, i.e. not be sensitive to the addition of soft particles, and that of collinear safety, i.e. insensitivity to the collinear splitting of particles. In CMS, PF candidates are clustered into jets using the anti-k_T [95] algorithm, satisfying the properties mentioned above. The determination of the flavour of the parton and its charge is challenging and relies on ML algorithms.

The input clusters for the jet clustering algorithm are the momenta of PF candidates or previously clustered 4-vectors of several PF candidates. The algorithm starts by introducing a distance measure, d_{ij}, between pairs of clusters (i, j), as well as a beam-cluster distance, denoted as d_{iB}:

$$d_{ij} = \min\left(\frac{1}{p_{\mathrm{T},i}^2}, \frac{1}{p_{\mathrm{T},j}^2}\right)\left(\frac{\Delta R_{ij}}{R}\right)^2, \quad d_{iB} = \frac{1}{p_{\mathrm{T},i}^2}, \tag{4.4}$$

where $\Delta R_{ij} = \sqrt{(\phi_i - \phi_j)^2 - (\eta_i - \eta_j)^2}$, and R determines the maximum distance for clustering pairs, as well as fixes the maximum jet area to πR^2. The algorithm then calculates and ranks all distances d_{ij} and d_{iB} for all clusters in the event. If the smallest distance happens to be an inter-jet distance, i and j are combined into a new cluster. If the smallest distance is a beam-jet cluster, the cluster is considered a final state jet and removed from further clustering. The procedure is then repeated until no more clusters remain. The anti-k_T algorithm preferentially clusters hard and closeby particles before continuing with soft and distant ones. It results in desirable experimental properties, such as robust reconstruction of the jet axis, and a circular shape in the $\eta\phi$ plane.

In this analysis, PF candidates are clustered with a jet size of $R = 0.4$, optimal to capture particles associated with a jet at $\sqrt{s} = 13$ TeV, while being resilient against pileup. Charged jet constituents not belonging to the PV are removed from the jet to reduce pileup in a method known as charged hadron subtraction [96]. The fine-grained tracker allows for a pure reconstruction of charged particles (65% of jet energy), while the ECAL allows for charged hadrons, neutral hadrons, and photons to be clearly separated, adding an excellent measurement of another 25% of PF jet energy. The HCAL provides the remaining 10% of jet energy, albeit with a modest resolution.

To match the energy of the reconstructed jet to that of the parton that produced it, the jet four-momentum must be calibrated. This is done in CMS in a factorised approach named jet energy corrections (JECs) to obtain the corrected jet four-momentum [97]. The JECs are derived using a detailed MC simulation of the detector and adjusted with data:

- The first stage of JECs is to remove the “offset” energy originating from pileup interactions and detector noise. The pileup offset corrections are determined from the simulation of a sample of QCD dijet events, processed with and without pileup overlaid.
- After correcting for pileup, the relative “response” corrections aim to calibrate the detector response of truth-level jets, which are obtained by clustering all generator-level particles excluding neutrinos. The response function is derived as a function of particle jet p_T and η by matching nearby truth-level and reconstructed jets in simulated dijet events. The obtained response function is then inverted to find the required correction to the reconstructed jet four-momentum.
- After the first two stages, residual corrections are applied to data to correct for remaining minor differences (of the order of a few per cent) within jet response between data and simulation. These corrections are determined with a well-measured reference object, e.g. jet recoil from Z bosons.
- An optional step, not applied in this analysis, is to correct for the jet flavour composition. This is done to account for the variations in jet fragmentation energy and the particle composition between jets of different flavours.

All steps in the chain have uncertainties associated with the corresponding JECs. As shown in Figure 4.6, the JEC uncertainties for Run 2 data taking conditions can be as large as 6% for jets with $p_T = 30\,\mathrm{GeV}$ [98], which is typical for displaced jets resulting from HNL decays. Finally, to account for the different jet energy resolutions (JERs) between data and simulation, the jet energy for simulated jets is smeared to match the jet energy resolution to the one observed in data.

Additional requirements, referred to as jet identification criteria, are applied to reject spurious jets from detector noise. The criteria are motivated by the fact that a genuine jet consists of several PF particles, both charged and neutral. The exact requirements are adjusted slightly for each year of data taking. Typically, jets are required to consist of at least two constituents, and for the neutral hadron fraction and neutral electromagnetic energy fraction to be both less than 90%. In addition, for jets within the tracker acceptance ($|\eta| < 2.4$), at least one constituent must be a charged hadron, and the charged electromagnetic energy fraction is usually required to be less than 99%. The efficiency for a jet that results from an HNL decay to be reconstructed and subsequently identified

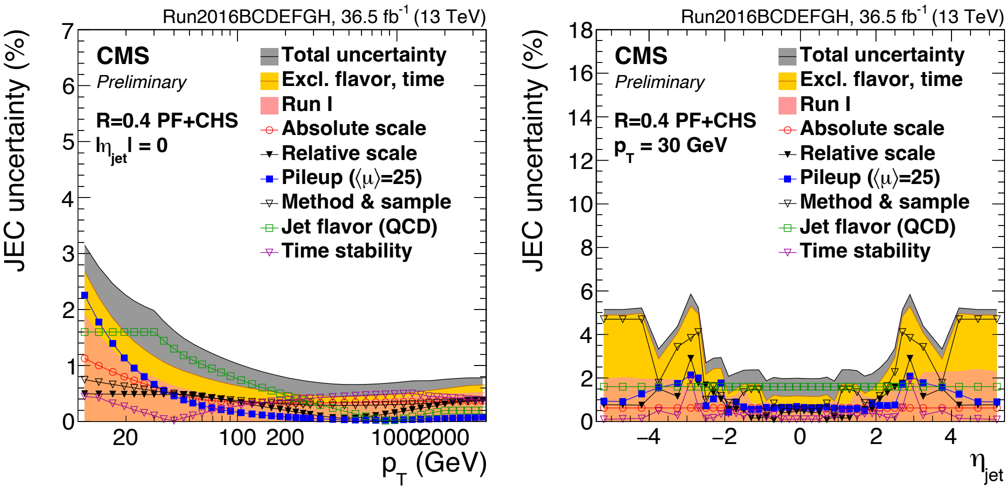


Figure 4.6: JEC uncertainty breakdown for anti-k_t, $R = 0.4$ charged hadron-subtracted central jets as a function of p_T (left) and for jets with $p_T = 30\,\text{GeV}$ as a function of η (right). Figure taken from Ref. [98].

is shown in Figure 4.7. Due to a finite jet energy resolution, a turn-on effect is observed; otherwise, the efficiency to reconstruct and identify a jet from a prompt HNL decay is close to unity, dropping to 80% for $L_{xy} = 1$ m. Overall, the PF jet algorithm has excellent performance for displaced jets, even at the outer edge of the CMS detector, as a significant fraction of neutral PF jet constituents is still reconstructed, as shown in Figure 4.8.

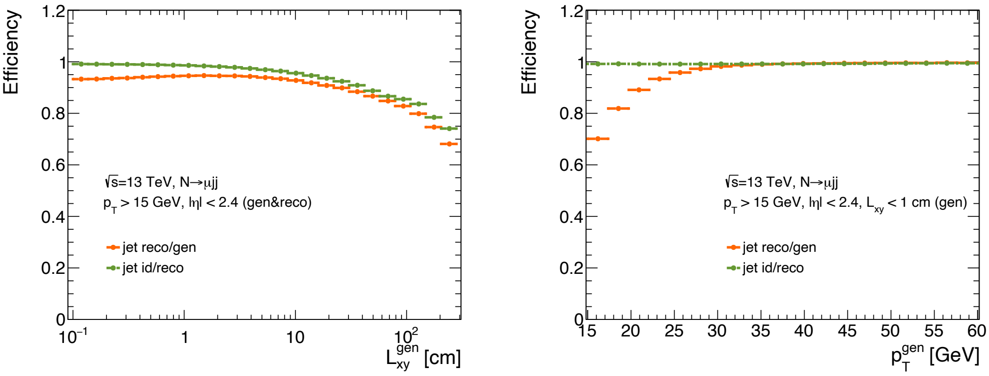


Figure 4.7: Jet reconstruction and identification efficiency as a function of generator-level lab frame displacement (left) and generator-level p_T (right).

4.2.4 Missing energy reconstruction

Because the system of colliding partons has little initial transverse momentum, the transverse momentum of all final state particles must add up to zero within the detector resolution. Any deviation in this is referred to as missing transverse energy, $\vec{p}_T^{\text{miss}}$, and it signals the presence of neutrinos or other weakly interacting particles escaping the

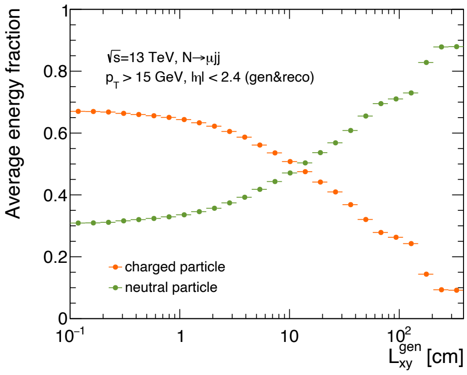


Figure 4.8: Average charged and neutral PF candidate energy fraction as a function of jet generator-level lab frame displacement.

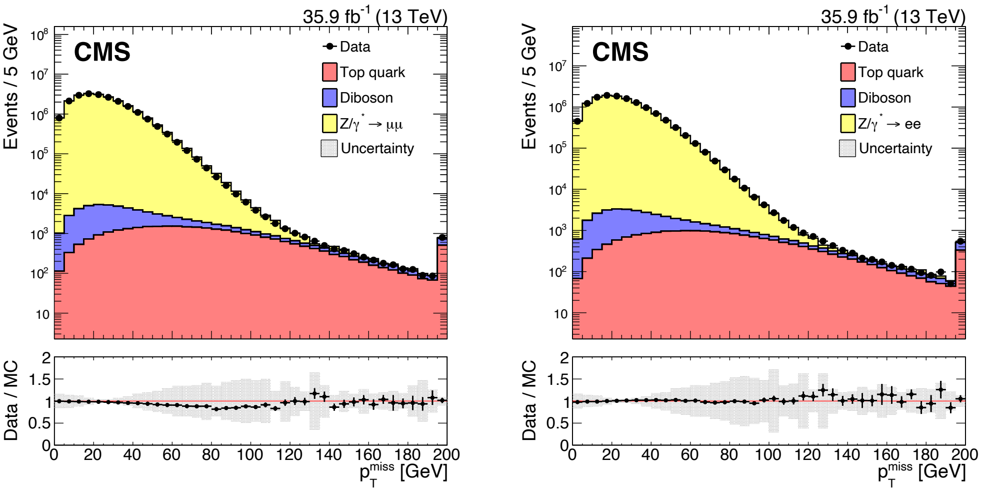


Figure 4.9: Distributions of $p_{\mathrm{T}}^{\mathrm{miss}}$ in $\mathrm{Z} \longrightarrow \mu^+\mu^-$ (left) and $\mathrm{Z} \longrightarrow e^+e^-$ (right) samples. Figure taken from Ref. [99].

detector and is, therefore, an essential quantity in BSM physics searches. The $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ is defined as the negative vectorial sum of all reconstructed PF candidates, and $p_{\mathrm{T}}^{\mathrm{miss}}$ is its magnitude:

$$\vec{p}_{\mathrm{T}}^{\mathrm{miss}} = -\sum \vec{p}_{\mathrm{T}}^{\mathrm{PF}}, p_{\mathrm{T}}^{\mathrm{miss}} \equiv |\vec{p}_{\mathrm{T}}^{\mathrm{miss}}|, \tag{4.5}$$

The calculation of $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ is biased by effects like minimum calorimeter energy thresholds, inefficiencies in tracking, non-linear calorimeter responses, and electronics noise. These effects are accounted for by propagating JECs to the calculation of $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ as:

$$\vec{p}_{\mathrm{T}}^{\mathrm{miss,corr.}} = \vec{p}_{\mathrm{T}}^{\mathrm{miss}} - \sum_{\mathrm{jet}} (\vec{p}_{\mathrm{T\ jet}}^{\mathrm{corr}} - \vec{p}_{\mathrm{T\ jet}}), \tag{4.6}$$

which uses all corrected jets with $p_{\mathrm{T}} > 15\,\mathrm{GeV}$ that have less than 90% of their energy deposited in the ECAL. This requirement ensures photons and electrons reconstructed as jets are not considered in the missing energy correction. The four-momenta of any muons found in jets are subtracted when performing the correction and added back to the corrected object. The scale and resolution of $p_{\mathrm{T}}^{\mathrm{miss}}$ are verified in $\mathrm{Z} \longrightarrow \mu^{+}\mu^{-}$ and $\mathrm{Z} \longrightarrow e^{+}e^{-}$ candidate events, as shown in Figure 4.9 [99]. The propagated JEC uncertainties are also displayed.

Anomalies causing high-$p_{\mathrm{T}}^{\mathrm{miss}}$ signatures can arise from spurious deposits due to particles directly striking ECAL photodetectors or from real particle showers caused by hadronic activity not associated with pp collisions, so-called beam halo effects. These are usually caused by the production of particles in upstream collisions, which travel parallel to the beam axis at a constant ϕ. Another source of fake $p_{\mathrm{T}}^{\mathrm{miss}}$ is poor reconstruction in the muon chambers, leading to muons being considered as charged hadrons in the PF algorithm. Dedicated cleaning algorithms are performed and are crucial for not mistaking a spurious high-$p_{\mathrm{T}}^{\mathrm{miss}}$ signature as a signal of new physics [100].

Chapter 5

Analysis techniques

In this Chapter, several techniques used in this analysis are described. A high-level overview of MC event generators, used to simulate the final state particles resulting from pp collisions, is provided. Subsequently, two families of ML classification algorithms, namely NNs and BDTs, are described. Lastly, the frequentist maximum likelihood framework for statistical inference is introduced.

5.1 Monte Carlo event generation

It is practically unfeasible to perform an analytical calculation of the expected multi-particle collision signature in the detector for a given physical process. Hence, the experimental signatures have to be simulated via MC integration. The analysis makes use of simulated samples for studying the event selection, training MVA discriminants, and performing statistical inference. The simulation is factorised into several main stages [101]: the hard parton scattering, parton shower (PS), secondary interactions, hadronisation, and hadron decays. These steps are shown schematically in Figure 5.1, and a typical general-purpose MC event generator, such as PYTHIA [102], can perform most of them.

5.1.1 Parton-level event generation

The event simulation is based on and starts with a specific hard scattering process of interest. The process can be both an SM process or a BSM process. The factorisation theorem allows expressing the cross-section for a given process with large momentum transfer by treating the soft initial state physics independently from the hard parton scattering process [103]. In this approach, the colliding partons are treated as free particles in an essentially "frozen" proton, as the interactions between the partons within the proton occur over significantly longer space-time scales:

$$\sigma(\mathrm{pp} \longrightarrow X) = \int_0^1 \int_0^1 \sum_{i,j} \mathrm{PDF}\left(x_1, f_i, \mu_\mathrm{F}\right) \mathrm{PDF}\left(x_2, f_j, \mu_\mathrm{F}\right) \sigma\left(f_i f_j \longrightarrow X\right) \mathrm{d}x_1\, \mathrm{d}x_2, \tag{5.1}$$

where $\sigma\left(f_i f_j \longrightarrow X\right)$ is the cross-section for the production of a final state, X, from two partons of flavours i, j. The cross-section can be calculated by considering the relativistic formulation of Fermi's golden rule (Equation 2.17) and a Lorentz-invariant formulation of the particle flux. For instance, the cross-section of the two-body process $i + j \longrightarrow a + b$ can be expressed as:

$$\sigma = \int \frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} \mathrm{d}\Omega = \int \frac{1}{64\pi^2 \hat{s}} \frac{p_j^*}{p_i^*} \left| \mathscr{M}_{ij \longrightarrow ab}(\mu_\mathrm{F}, \mu_\mathrm{R}) \right|^2 \mathrm{d}\Omega, \tag{5.2}$$

where $\mathrm{d}\Omega = \mathrm{d}(\cos\theta)\mathrm{d}\phi$ is a differential solid angle element, $\sqrt{\hat{s}} = \sqrt{x_1 x_2}\sqrt{s}$, where $\sqrt{s}$ is the centre-of-mass energy of the pp system, and $p_{i,j}^*$ are momenta of the two partons in the centre-of-mass frame. $\mathscr{M}_{ij \longrightarrow ab}$ is the ME, introduced in Equation 2.17, and computed to some order in perturbation theory. MC event generators perform the integral of Equation 5.1 by employing MC sampling techniques according to Equation 5.2. These methods are advantageous for simulating high-dimensional final-state parameter space integrals, as the uncertainty does not depend on the number of dimensions but only on the number of events as $\sim \sqrt{N}$. The events are generated in such a way that they all contribute with the same probability to the integral. For a given physical process with a cross-section, σ, for which N events have been generated, an event weight can be calculated as:

$$\omega_\mathrm{gen} = \frac{\sigma \mathscr{L}}{N}, \tag{5.3}$$

and applied to simulated events to match the expected number of events in a data set of integrated luminosity $\mathscr{L}$.

In a pp collision, secondary "remnant" interactions not associated with the hard parton scattering can occur. This is referred to as the underlying event (UE) and collectively comprises initial state radiation (ISR) and final state radiation (FSR) with small momentum transfer, beam-beam remnants, and multiple parton interactions.

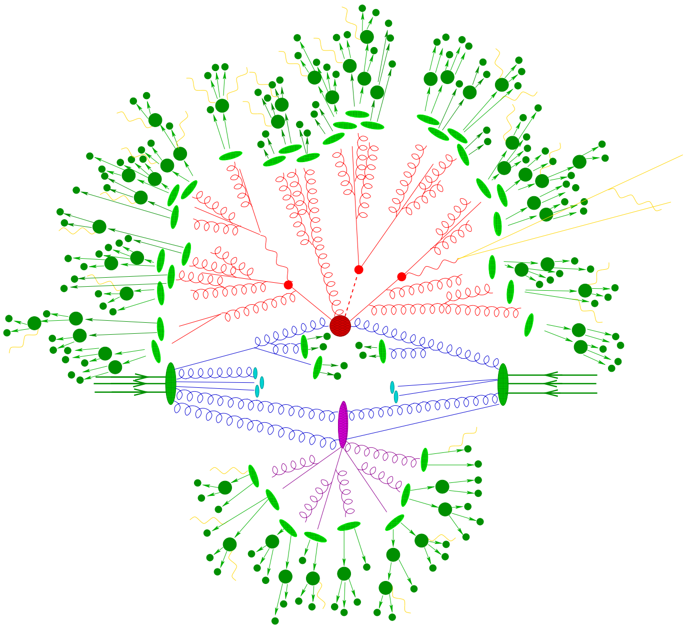


Figure 5.1: Schematic representation of event simulation using MC generators. The incoming partons are shown as blue lines. Stages of the simulation are displayed: the hard process (red circle), PS (red gluon lines), hadronisation (light green circles), the UE interactions (purple circle), and hadron decays (in dark green). Figure taken from Ref. [104].

5.1.2 Parton showering and hadronisation

The final state partons resulting from the hard interaction process possess a colour charge and can radiate gluons. The resulting FSR is predominantly very soft or almost collinear with the outgoing partons. This radiation gives rise to further gluons or quark-antiquark pairs, and the process of showering repeats until the energy of the final state particles are low enough to form bound states. Event generators treat PS as a Markov chain and recursively simulate it using MC techniques until the parton energy evolves to a scale of $\mathcal{O}(1\,\text{GeV})$. PS models have several free parameters which are tuned to match observed UE data [105].

As perturbative QCD can no longer be used at low energies, PS is replaced by various QCD-inspired hadronisation models, such as the string model [106]. In these models, colour-connected systems of partons are hadronised collectively, which ultimately results in colour-neutral hadronic final states. Finally, unstable hadrons are decayed until a set of detector-stable particles is obtained.

A complication arises because the FSR of partons can produce additional partons that are indistinguishable from partons produced in the equivalent hard scattering process. To avoid this double-counting, which would result in an overestimation of the hadronic activity in an event, a jet matching procedure is employed [107]. The ME calculation of the hard process cannot reliably describe soft gluon emissions. On the other hand, shower algorithms are inaccurate for hard, wide-angle emissions, i.e. additional resolved jets. Hence, such matching schemes typically introduce a specific cut-off scale, below which only partons from the showering are retained. Above this scale, partons originating from the ME calculation are used instead. An alternative approach to deal with this issue is to employ a jet merging scheme instead, where jets from ME and PS calculation are merged.

For the generated samples used in this analysis, minimum bias events are overlaid with the hard scattering event to simulate pileup interactions, with the multiplicity distribution matched to that observed in pp collision data. The PYTHIA program is used to describe the PS and hadronisation for all simulated samples. The NNPDF3.0 LO and NLO parton distribution functions [12] are used with the event generators.

In summary, a general-purpose MC event generator can simulate a range of interesting physical processes occurring during pp collisions at the LHC, and can likewise be used to simulate the signatures of new particles.

5.1.3 Detector simulation

The final step of event generation is simulating the interactions of all final state particles with the active detector material. In CMS, the detector simulation is done based on the GEANT4 toolkit [108]. It consider interactions with matter, the production of secondary particles, and propagates particles through electromagnetic fields. Finally, the readout of various detector components is simulated in the digitisation step.

5.1.4 Background samples

For the analysis described in this thesis, the simulated background samples are separated into two categories: those used in the statistical analysis and those used to train various MVA discriminants. Independent samples must be used for the two purposes to prevent a bias arising from an ML algorithm performing better on the training data set.

Long-lived SUSY and HNL used for training the displaced jet tagger are further described in Section 6.4.5 and Section 7.2, respectively. In all cases, the signal samples are partitioned into independent samples for the training and evaluation of MVA algorithms.

For the samples used in the statistical analysis, major SM background processes resulting in at least one and potentially several leptons in the final state are simulated. The MADGRAPH5_aMC@NLO event generator is used to produce samples of W+jets at NLO precision in QCD with up to two additional partons at the ME level and are merged with jets from the subsequent pythia PS simulation using the FxFx scheme [109]. Samples of Z/γ^*+jets are produced in two “slices” depending on the dilepton invariant mass, $m_{\ell\ell}$, at NLO (LO) precision, with MadGraph5_aMC@nlo (MadGraph), and merged (matched) using the FxFx (MLM) [107] scheme. The powheg generator is used to simulate $\mathrm{t\bar{t}}$ production [110] as well as t-channel [111] and tW [112] production of single top quarks at NLO accuracy. Additionally, leptons can be produced during the hadronisation of jets; hence, the QCD multijet process is essential to consider. Multijet events are simulated at LO accuracy using pythia. As the production of leptons in the final state in QCD multijet events is rare, during the generation of QCD multijet samples, a filter is optionally applied, requiring at least one lepton at the generator level. The samples are summarised in Table 5.1.

For the training of various MVA classifiers described within this thesis, LO samples of W+jets, Z/γ^*+jets, and $\mathrm{t\bar{t}}$ generated with MadGraph were used. Independent p_{T}-binned samples of QCD multijet production, generated with pythia, were also used. The samples are summarised in Table 5.2.

Table 5.1: Background samples for statistical inference.

Process	Samples (ME)	Generator	Colour
W+jets	W+0 jets W+1 jet W+2 jets	MadGraph5_aMC@nlo	🟩
Z/γ^*+jets	$m_{\ell\ell} > 50\,\mathrm{GeV}$ $10 < m_{\ell\ell} < 50\,\mathrm{GeV}$	MadGraph5_aMC@nlo MadGraph	🟦
top-quark	$\mathrm{t\bar{t}} \longrightarrow \ell\bar{\ell}\nu\bar{\nu}b\bar{b}$ $\mathrm{t\bar{t}} \longrightarrow \ell\nu b + \mathrm{q\bar{q}}'\bar{b}$ single-top, t-channel single-top, tW-channel	powheg	🟨
QCD multijet	p_{T}-binned, $\mu(e)$-enriched H_{T}-binned	pythia	⬜
V$\gamma^\star$+jets	Z($\longrightarrow \ell\ell)\gamma^\star$+0,1 jets W($\longrightarrow \ell\nu)\gamma^\star$+0,1 jets	MadGraph MadGraph5_aMC@nlo	🟪

Table 5.2: Background samples for MVA discriminant training.

Process	Samples (ME)	Generator
W+jets	inclusive	MadGraph
Z/γ^*+jets	$m_{\ell\ell} > 50$ GeV	MadGraph
$t\bar{t}$	inclusive	POWHEG
QCD multijet	p_T-binned	PYTHIA

5.2 Machine learning

ML is a subfield of artificial intelligence concerned with the development of a computer algorithm (model) that can automatically improve through experience. Conventional ML algorithms can be broken down into three main parts:

Decision process ML algorithms involve making either classification or regression tasks. Classifications tasks aim to accurately predict a discrete class label given an example or the corresponding probability, while regression is concerned with predicting a continuous value, i.e. a number. Algorithms are also subdivided depending on whether the training data is labelled or not, called supervised and unsupervised learning, respectively. The final possibility is that of semi-supervised learning, where the training data is partially unlabelled. Typically, supervised ML algorithms are employed in HEP to perform object identification or event classification and rely on labelled simulated samples. A canonical example of unsupervised learning is clustering algorithms.

Loss function A loss (cost) function, which the algorithm aims to minimise, and which assigns a single number ("cost value") associated with a given sample (event).

Model optimisation process An optimisation algorithm aims to determine the internal model parameters that minimise the chosen loss function for a given training data set.

The focus of this Section is on classification algorithms, which are employed extensively within this thesis.

5.2.1 Neural networks

Neuron

A single artificial neuron is a mathematical function described by a set of weights ($\boldsymbol{\omega}$) and a bias term (b) which act on a set of inputs $\mathbf{x}$ to produce a single output:

$$y = \sigma\left(\sum_i \omega_i x_i + b\right) \equiv \sigma(z), \tag{5.4}$$

where σ is a (typically non-linear) activation function, and we have defined the weighted input as z.

Neural networks

The simplest NN, called a perceptron, corresponds to a single neuron. A feed-forward NN consists of a series of layers of neurons in which the connections between the neurons do not form a cycle. In such a network, where the signals flow in the forward direction only, the outputs of a given layer play the role of inputs to the following one. As all the neurons in each layer "see" the same inputs, their output can only be different if the weights themselves are different. For a vector of neurons in a layer l, the output of the layer can be expressed in matrix form as:

$$\mathbf{y}_l = \sigma_l\left(\Omega_l \mathbf{y}_{l-1} + \mathbf{b}_l\right) \equiv \sigma_l\left(\mathbf{z}_l\right), \tag{5.5}$$

where $\Omega_l = \mathbb{R}^{N\times M}$, and $\mathbf{y}_l, \mathbf{b}_l, \mathbf{z}_l = \mathbb{R}^M$. Assuming there are L layers in total, $\mathbf{x} \equiv \mathbf{y}_0$ denote the NN inputs, and $\hat{\mathbf{y}} \equiv \mathbf{y}_L$ its output. The intermediate (hidden) layer ($1 < l < L$) outputs are usually called latent features. The set of network weights can be collectively expressed as:

$$\Omega = \Omega_1 \circ \ldots \circ \Omega_l \circ \ldots \circ \Omega_L, \tag{5.6}$$

with the bias term absorbed into the definition. The same notation is adopted for the activation functions:

$$\sigma = \sigma_1 \circ \ldots \circ \sigma_l \circ \ldots \circ \sigma_{L-1}. \tag{5.7}$$

For the final network layer, an activation function is not applied on each score separately; instead, a softmax function transforms the output vector to a vector of class probabilities:

$$\hat{y}_k = \frac{e^{z_L^k}}{\sum_n e^{z_L^k}}, \tag{5.8}$$

where n is the number of classes considered. The network output (class probability) is then used as a discriminant in a physics analysis. The set of weights and activation functions is called a model and completely describes the response of the network to input data:

$$\hat{\mathbf{y}} = f(\mathbf{x}|\Omega, \sigma). \tag{5.9}$$

A simple example of a NN used for binary classification is shown in Figure 5.2. To define a procedure for optimising the model, a loss function, describing its accuracy, has to be determined. A powerful result, called the universal approximation theorem [113], states that a sufficiently wide NN with only a single hidden layer can approximate any well-behaved one-dimensional function over a specific interval.

Loss function

The typical loss function used for classification tasks is categorical cross-entropy, defined for two discrete probability distributions, p and q, as:

$$H(p,q) = -\sum_{i=1}^{n} p_i \log q_i = H(p) - D_{\mathrm{KL}}(p,q), \tag{5.10}$$

where p_i, q_i are probabilities for occurrence (class) $i \in \mathbb{R}^n$, and $H(p)$ is the entropy of p:

$$H(p) = -\sum_{i=1}^{n} p_i \log p_i. \tag{5.11}$$

Finally, D_{KL} is the Kullback-Leibler (KL) divergence, a measure of divergence of p from q:

$$D_{\mathrm{KL}}(p,q) = \sum_{i=1}^{n} p_i \log\left(\frac{p_i}{q_i}\right). \tag{5.12}$$

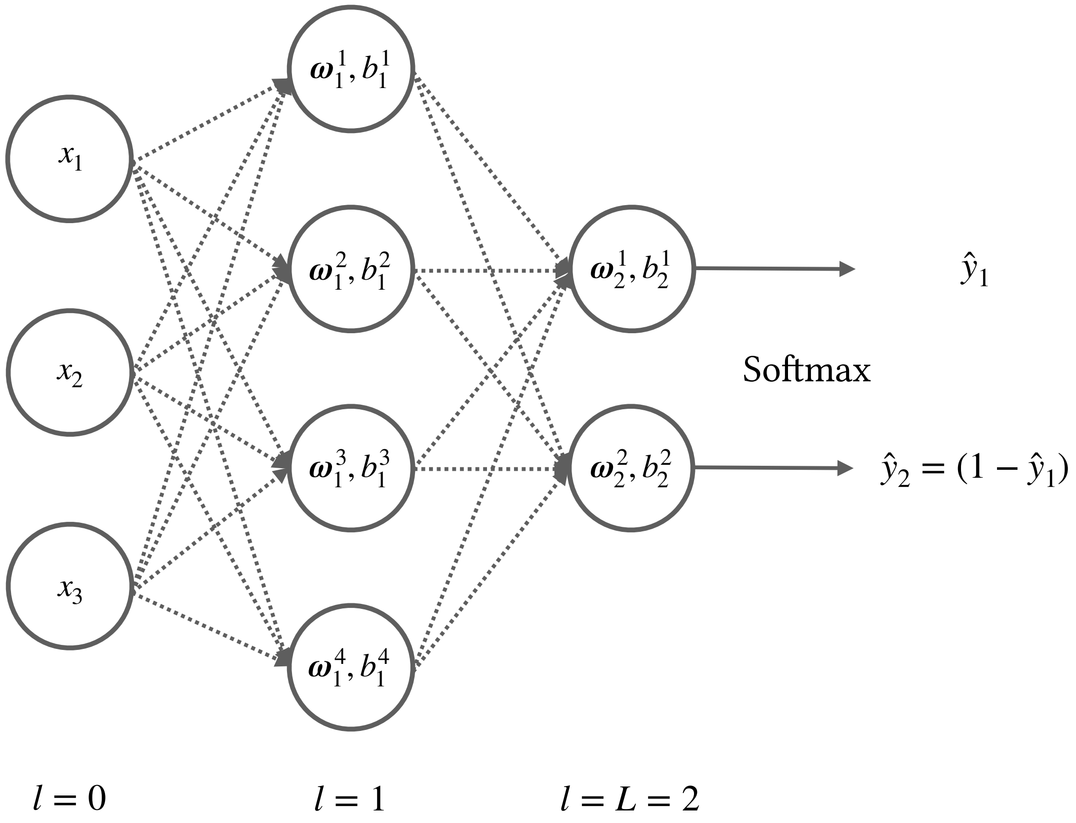


Figure 5.2: An example of a simple NN model that could solve binary classification problems. The network has three input features, one hidden layer, and outputs probabilities for the two classes. The activation function of the hidden layer is not shown explicitly.

In the particular case of solving a binary classification problem by using binary cross-entropy, a single neuron with a sigmoid activation function, i.e. $\sigma(z) = 1/(1 + e^{-z})$, corresponds precisely to the logistic regression model, one of the simplest and easiest to interpret classification models in ML.

The Shannon mutual information (MI) entropy is used [114] in the analysis described within this thesis to quantify the independence between variables. In the discrete case, this is defined in terms of the individual and joint entropy distributions:

$$M(p,q) = H(p) + H(q) - H(p,q). \tag{5.13}$$

Entropy is an additive quantity for statistically independent variables, and hence the MI is zero in this case. This is a stronger statement than the linear (Pearson) correlation coefficient (ρ) being zero. An exact relationship between the two exists, assuming both marginal distributions are normally distributed:

$$M = -\frac{1}{2}\log\left(1 - \rho^2\right). \tag{5.14}$$

Finally, to make similarity comparisons between two probability distributions, it is helpful to consider the Jensen-Shannon divergence (JSD), a symmetrised version of the KL divergence:

$$\mathrm{JSD}(p, q) = \frac{1}{2}D_{\mathrm{KL}}(p, m) + \frac{1}{2}D_{\mathrm{KL}}(q, m), \tag{5.15}$$

where $m = (p + q)/2$ is the mean of the probability distributions. Hence, the JSD is often called the total distance to the average. The JSD is bound to $[0, 1]$, and takes a zero value for identical distributions.

In supervised learning, a single labelled training example y_i is usually encoded as a one-hot (dummy) variable:

$$y_i = \delta_{ij}, \tag{5.16}$$

where δ_{ij} is the Kronecker delta and j is the class label. For a sample of size N, the total loss is defined in terms of the cross-entropy as:

$$L(\mathbf{y}, \hat{\mathbf{y}}(\mathbf{x})) = -\frac{1}{N}\sum_{m=1}^{N}\sum_{i=1}^{n} y_i^m \log \hat{y}_i^m(\mathbf{x}), \tag{5.17}$$

where $\hat{y}_i^m(\mathbf{x})$ is the NN probability estimate for class i for a single training example, m, as defined in Equation 5.9. The goal of training the NN is to find a set of weights, Ω, which minimises the loss function.

Gradient descent

The loss function is minimised through an iterative procedure known as gradient descent. A small change in the loss function due to a change in the weights is calculated as:

$$\Delta L = \sum_l \left(\frac{\partial L}{\partial \Omega_l} \cdot \Delta\Omega_l\right) \equiv \nabla L \cdot \Delta\Omega. \tag{5.18}$$

In each iteration, the network weights are updated in the direction of the gradient, i.e. steepest descent in the gradient space:

$$\Delta\Omega_l = -\eta \cdot \nabla L_l, \tag{5.19}$$

where η is the learning rate of the algorithm, restricted to $0 < \eta < 1$ to ensure convergence. There are two main practical implementations of gradient descent, called batch and stochastic gradient descent (SGD). Batch gradient descent computes the gradient by using the entire available dataset, which is computationally inefficient for large data sets, in which case using SGD becomes critical. In SGD, the gradient is repeatedly computed for subsets (minibatches) of data, approximating the entire batch's gradient.

Various extensions to the SGD algorithm are used, and one such widespread algorithm is called Adaptive Moment Estimation (Adam) [115], where the weights are updated by first calculating the bias-corrected mean and variance estimates m^t, v^t of gradients at iteration t:

$$\begin{aligned} \hat{m}_l^t &= \left(\beta_1^t \cdot m_l^{t-1} + \left(1 - \beta_1^t\right) \cdot \nabla L_l\right) / (1 - \beta_1^t), \\ \hat{v}_l^t &= \left(\beta_2^t \cdot v_l^{t-1} + (1 - \beta t_2) \cdot \nabla {L_l}^2\right) / (1 - \beta_2^t), \end{aligned} \tag{5.20}$$

where $\beta_{1,2}^t$ are exponentially decaying coefficients with decay rates $(\beta_{1,2})$ for each of the moment estimates, respectively. The weights are updated as:

$$\Delta\Omega^t = -\eta \cdot \frac{\hat{m}_l^t}{\sqrt{\hat{v}_l^t} + \varepsilon}, \tag{5.21}$$

where ε is a small number to prevent division by zero. While SGD can be interpreted as a massless ball rolling down the slope of a multi-dimensional surface, Adam can be compared with a heavy ball with friction, which tends to travel in the same direction, preventing perturbations from local minima. Another advantage of the Adam algorithm is that the learning rate is effectively adapted for each weight. The authors of the algorithm have shown that the algorithm compares favourably to other adaptive learning-method algorithms.

Backpropagation

The backpropagation algorithm [116] allows to compute the gradient in the Ω-space of the NN efficiently. It proceeds by computing the gradient recursively backwards, layer-by-layer. To derive this recursive principle, let us consider the derivative of the loss with respect to the weights of the layer l:

$$\nabla L_l = \frac{\partial L}{\partial \Omega_l} = \frac{\partial L}{\partial \mathbf{z}_l}\frac{\partial \mathbf{z}_l}{\partial \Omega_l} = \frac{\partial L}{\partial \mathbf{z}_l}\mathbf{y}_{l-1}, \tag{5.22}$$

where we have made use of the chain rule and Equation 5.5. We now derive a recursive relation between the gradients for layers $l < L$ and $l + 1$:

$$\frac{\partial L}{\partial \mathbf{z}_l} = \frac{\partial L}{\partial \mathbf{z}_{l+1}}\frac{\partial \mathbf{z}_{l+1}}{\partial \mathbf{z}_l} = \frac{\partial L}{\partial \mathbf{z}_{l+1}}\sigma_l'\left(\mathbf{z}_l\right)\Omega_{l+1}, \qquad l < L. \tag{5.23}$$

As a final step, we need to explicitly consider the output layer of the network, $l = L$, with the softmax function defined in Equation 5.8, which can be shown to give:

$$\frac{\partial L}{\partial \mathbf{z}_L} = \frac{\partial L}{\partial \hat{\mathbf{y}}}\frac{\mathrm{d}\hat{\mathbf{y}}}{\mathrm{d}\mathbf{z}_L} = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{\mathbf{y}}_i - \mathbf{y}_i\right). \tag{5.24}$$

Thus, the total gradient can be calculated recursively, starting from the last layer and propagating back. This is summarised in Algorithm 1.

Algorithm 1 Backpropagation and SGD

```
Require: Initial weights $\Omega^0$, learning rate $\eta$, minibatch size $N$.
Shuffle the training data set and divide it into $n$ minibatches of size $N$.
while $\Omega$ has not converged do
    for minibatch = $1 \ldots n$ do
        Propagate input through the network, storing values of $\mathbf{z}^l, \Omega^l$ for each layer.
        Calculate the loss $L(\mathbf{y}, \hat{\mathbf{y}}(\mathbf{x}))$ according to Equation 5.17.
        Determine the error and gradients of the last layer via Equation 5.24 and Equation 5.22.
        Recursively update the gradients according to Equation 5.23, Equation 5.22, and Equation 5.19.
    end for
end while
```

The choice of initial weights Ω^0 can have a significant impact on the training procedure. In general, the initialisation strategy should be adapted according to the shape of the weight matrix Ω as well as the chosen activation function σ by sampling from an appropriate uniform or normal distribution. Commonly used initialisation procedures are Lecun, Glorot, and He [117]–[119].

Input features are often transformed to have a similar range to make backpropagation more efficient. A common normalisation is to have zero mean and unit variance for each input feature distribution [117].

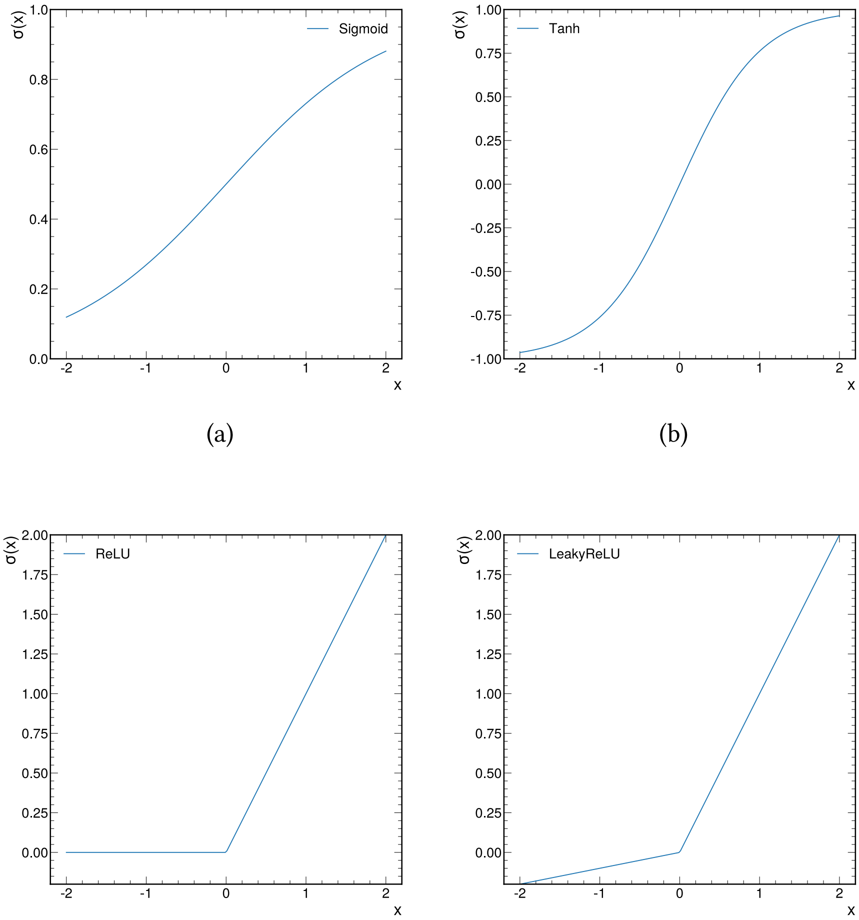
Activation functions

Non-linear activation functions are essential for NNs to approximate non-linear functions, as the final output of a NN with only linear activation functions will be based on a linear combination of the input features. Additionally, a continuously differentiable function is necessary to enable gradient descent-based optimisation algorithms to work. Typical activation functions are shown graphically in Figure 5.3. While the rectified linear unit (ReLU) and LeakyReLU functions are not differentiable at the singular computationally rare point of $x = 0$, the gradient is usually manually set to zero in such cases.

The choice of suitable activation functions is motivated by their desirable mathematical properties and empirical performance. The original development was based on mimicking the behaviour of biological neurons, which resulted in using the sigmoid, and later the hyperbolic tangent activation functions. The rectifier function and its variants are the most widely used activation function nowadays due to their practicality. When using the sigmoid or tanh activations, the output, and hence the maximum gradient in a layer, saturates and decreases exponentially in the subsequent layers of backpropagation, as can be inferred from the recursive relation of Equation 5.23. This phenomenon is known as the vanishing gradient problem and can be mitigated using the ReLU activation, as the gradient is constant. This is of particular importance when training DNNs.

Bias-variance trade-off

The bias-variance trade-off is a central problem in supervised learning, which means that it is impossible to have a model that perfectly captures the features in the training data while generalising well to previously unseen data. Having an algorithm that does not adequately capture the underlying structure of the data is called underfitting. On the other hand, having a model too sensitive to noise in the training data set is referred to as overfitting. This is highly undesirable as using an overtrained classifier in a physics analysis would introduce bias and result in an avoidable loss of sensitivity. To monitor for under and overfitting, the available data is usually split into training and validation samples, and only the training set is used in the algorithm's training. The loss function of the validation data set is then monitored as a function of the training cycle, as shown in Figure 5.4. In addition, the similarity between classifier score distributions, obtained by evaluating the NN on training and validation data sets, is compared using, for example, a Kolmogorov-Smirnov (KS) test [121].



(a) (b)

(c) (d)

Figure 5.3: Commonly used activation functions (a) sigmoid $\sigma(x) = 1/(1+e^{-x})$; (b) hyperbolic tangent $\sigma(x) = (e^{x} - e^{-x})/(e^{x} + e^{-x})$; (c) ReLU $\sigma(x) = \max(0, x)$ [120]; (d) LeakyReLU $\sigma(x) = \max(\alpha x, x)$ where $\alpha = 0.1$ is the leakage factor.

No free lunch theorem

The “no free lunch theorem” is an important result stating that it is impossible to train a single MVA discriminant with optimal performance for all possible problems [122]. In other words, it is futile to try to find a general-purpose, universal optimisation strategy.

Regularisation

To avoid overfitting, regularisation techniques are often employed. For NNs, a commonly used method is dropout [123], where a fraction of neurons and their connections are randomly severed during the network training.

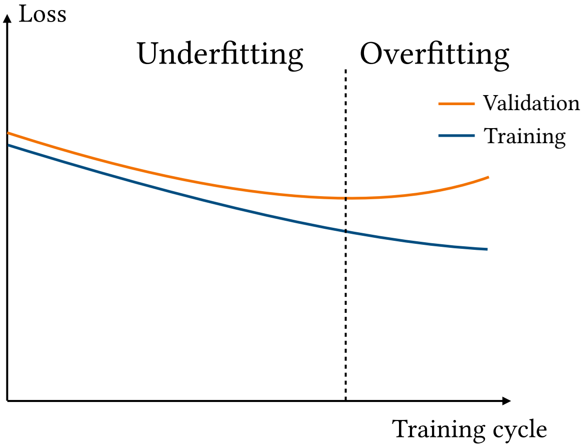


Figure 5.4: The progression of the training and validation data set loss as the NN is trained using backpropagation and SGD. At some point, the validation loss starts to increase again, which signals that the network is overfitting the training data set.

Hyperparameters

The free parameters (weights) of the network are optimised during the training procedure. Conversely, the overall network architecture, the optimisation algorithm parameters, such as the learning rate, and regularisation parameters, such as the dropout rate, need to be set in advance. These parameters are called hyperparameters and are subject to manual optimisation outside of the training process.

Deep learning

DL refers to the training of a class of NNs that involve multiple hidden layers. Deep NNs have achieved state-of-the-art performance in various tasks such as speech recognition [124], computer vision [125], and natural language processing [126]. It has been shown in Ref. [127] that networks of width $n + 4$ with ReLU activation functions can approximate any continuous function in n-dimensional input space, assuming the network depth is allowed to grow.

5.2.2 Decision trees

A decision tree is a very popular algorithm due to its simplicity and interpretability [128]. One advantage of decision trees is that, unlike NNs, they do not require input feature distribution normalisation, and work well with both categorical and numerical features. The process of building a decision tree is called decision tree learning. Decision tree learning algorithms typically work in a top-down approach, choosing the best feature to split a data set into two new samples. The information gain is a commonly used metric that measures the relative decrease in entropy due to a given split:

$$\mathrm{IG}(b, a) = H(b) - H(b \mid a), \tag{5.25}$$

where $H(b)$ and $H(b \mid a)$ are the sample entropies before and after the separation, respectively. Supposing we have a training sample of size N, the entropy before the split can be obtained by substituting the fraction of samples of each class i:

$$p_i = \frac{N_i}{N}, \tag{5.26}$$

into Equation 5.11. At a given node, the sample is split into two subsamples, $a_{1,2}$, with $N^{(1)} + N^{(2)} = N$. The entropy after the separation is found as a weighted sum of the individual resulting entropies:

$$H(b \mid a) = \frac{N^{(1)}}{N} H(b \mid a_1) + \frac{N^{(2)}}{N} H(b \mid a_2). \tag{5.27}$$

The split is chosen as the one which maximises the information gain, and tree learning algorithms proceed recursively on the partitioned subsets until some specified stopping condition is met. This happens, for example, when a given node contains less than some predefined number of events. At this point, the algorithms make class assignments by creating leaf nodes.

Although decision trees are interpretable and straightforward, they are relatively inaccurate and prone to overtraining. To overcome these shortcomings, decision trees are often combined in ensemble methods such as random forests [129] or gradient BDTs [130]. These algorithms work by combining several weak classifiers into a single strong learner.

Boosted decision trees

A gradient-boosted decision tree or simply a BDT is an ensemble of decision trees, where the individual learners (trees) are trained via gradient boosting, which is somewhat similar to the gradient descent algorithm used to train NNs, described in Equation 5.18. For the case of NNs, the network parameters (weights) are simply numbers. Decision trees, on the other hand, need to be parameterised by a function. New trees are added iteratively during the training, with each new tree following the gradient while keeping

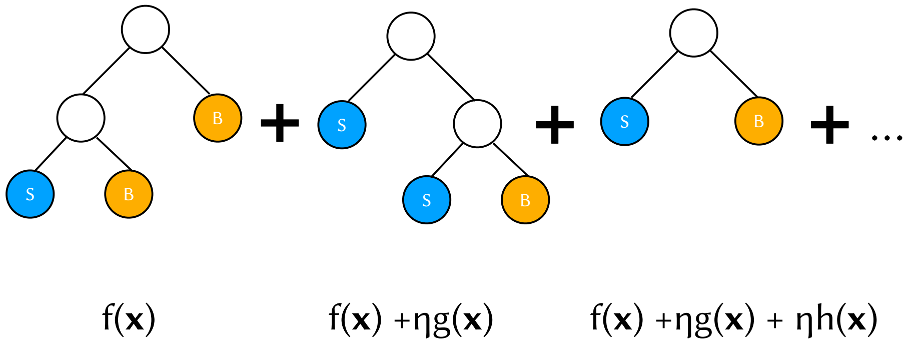


Figure 5.5: A sketch of several decision trees combined into a BDT.

the previous ensemble of trees fixed. To make a prediction, the ensemble trees perform a weighted majority vote, with later trees contributing less to slow down the learning by the algorithm. This weighting, as before, is called the learning rate. Practically, it has been observed that such algorithms have good performance even if the individual decision trees are stumps, i.e. they only have a single node. A rudimentary sketch of a BDT is shown in Figure 5.5. One fast and scalable gradient-boosting algorithm, that achieves state-of-the-art performance on a wide range of problems and is widely used in HEP, is called `XGBoost` [131].

5.2.3 Domain adaptation

Domain adaptation (DA) is a general area of ML, relevant whenever there is a substantial difference between the features of an algorithm's training data set, and the data set it encounters once deployed. Specifically for our case, this means the situation of training an algorithm on simulated samples (jets) yet applying it to real pp collision data. The features of simulated jets might be mismodelled due to, for instance, limitations in the simulation software, an overly naive detector description, or the impossibility to model specific background processes such as nuclear interactions of particles with the detector material. Hence, the discriminant distribution might not be very representative of the actual data. In the presence of many input features, manual selection of well-modelled input features is impractical.

A common approach in HEP adopted to a posteriori account for the mismodelling of a discriminator output, which ultimately arises due to mismodelling of the algorithm's input features, is to derive correction scale factors that quantify the disagreement in the discriminator output between the data and simulation. This is typically done in a particular CR, enriched by a specific process of interest. However, it might be more beneficial to a priori account for any such mismodelling when training the algorithm, which DA ultimately allows.

In this analysis, unsupervised DA by backpropagation [132] is used, enabling a straightforward extension to a standard feed-forward NN architecture to accommodate the differences between the domains of interest, which are data and simulation. Here, "unsupervised" refers to the fact that the target domain (pp collision data) does not possess any class labels; however, this is not necessary for the algorithm to work. The search described in this thesis is the first data analysis using DA at the LHC. DA was also succesfully used at the MINERvA neutrino experiment to reduce NN model bias [133].

The desired behaviour can be achieved by requiring that the latent network features discriminate between classes of the source domain and are invariant with respect to the differences between the source and target domains. As was first shown in Ref. [132], this can be easily achieved by introducing a single gradient reversal layer (GRL) and training with the standard backpropagation algorithm. The class (source) data set contains simulated samples from the source domain with class labels, whereas the domain (target) data set includes samples that only possess a source or target domain label, corresponding to simulation and data. Ideally, the two data sets would correspond to the same region of phase space. However, the two data sets can also be drawn from different regions, i.e. a particular set of fiducial requirements can be required on the class data set but not on the domain data set. The classification problem for the source domain is described by the usual loss term of Equation 5.17, denoted as L_c, and an additional loss term for the target domain is introduced:

$$L_\mathrm{d} = -\frac{1}{M}\sum_{m=1}^{M}(y_m \log \hat{y}_m + (1 - y_m)\log(1 - \hat{y}_m)), \tag{5.28}$$

where $y_m = 1$ if the m-th sample belongs to the target domain (data) and zero if it belongs to the source domain (simulation).

The network input is mapped to a space of latent features $\mathbf{f}$ using a standard feed-forward architecture. We call this part the feature extractor and denote its weights as Ω_f. The NN architecture is extended to have two independent branches originating from the latent feature layer $\mathbf{f}$: one branch to classify samples from the source domain, and another to determine the sample domain, as shown in Figure 5.6. The two branches are called the class predictor and the domain predictor, with their weights denoted as Ω_c and Ω_d, respectively.

As before, the training goal is to obtain a set of weights Ω_f and Ω_c, which minimises the class loss L_c, as this would achieve the goals of class discrimination. Crucially, we would also like to obtain a set of weights Ω_f, which maximises the domain loss L_d, while

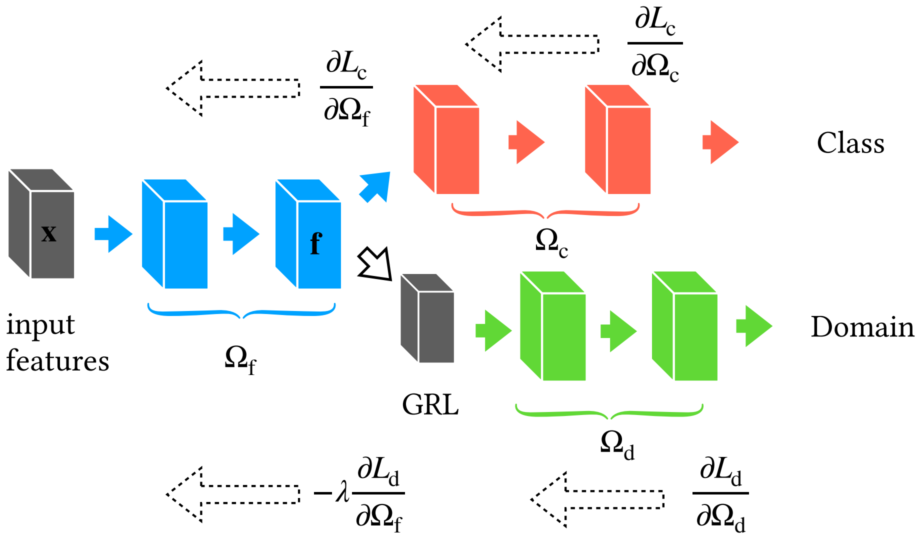

$\frac{\partial L_c}{\partial \Omega_c}$
$\frac{\partial L_c}{\partial \Omega_f}$
Class
x
f
Ω_c
Domain
input
features
Ω_f
GRL
Ω_d
$-\lambda \frac{\partial L_d}{\partial \Omega_f}$
$\frac{\partial L_d}{\partial \Omega_d}$

Figure 5.6: Sketch of a NN architecture to perform DA by backpropagation. The feature extractor (blue) and a label predictor (red) together form a standard feed-forward architecture and predict class labels. A domain classifier (green) is connected to the feature extractor via a GRL that multiplies the gradient by a certain negative constant during the training via backpropagation.

simultaneously getting a set of weights Ω_d, which minimises L_d. This would achieve the goal of domain invariance, as the domain predictor would not be able to tell the domain label based on the latent space vector f. The total loss function of the NN is introduced as the sum of the class and domain contributions:

$$L = L_\mathrm{c}(\Omega_\mathrm{f}, \Omega_\mathrm{c}) + \lambda L_\mathrm{d}(\Omega_\mathrm{f}, \Omega_\mathrm{d}), \tag{5.29}$$

where λ is a hyperparameter controlling the strength of DA and has to be optimised. The GRL is the final ingredient, introduced as an interface between the feature extractor and the domain predictor. The GRL acts as the following:

$$f(x) = \begin{cases} x & \text{forward pass,} \\ -x & \text{backpropagation.} \end{cases} \tag{5.30}$$

The GRL ensures that the following function is effectively minimised for the feature extractor when trained in a standard way using backpropagation:

$$L_\mathrm{f} = L_\mathrm{c}(\Omega_\mathrm{f}) - \lambda L_\mathrm{d}(\Omega_\mathrm{f}), \tag{5.31}$$

which forces the network to penalise mismodelled input features while retaining discrimination between classes.

5.3 Statistical inference

This Section describes the framework for frequentist statistical inference, specifically for setting ULs on model parameters.

5.3.1 Likelihood model

The first step of statistical inference is constructing a likelihood model that describes the compatibility of a given parameter of interest (POI) with a given set of observations. For a multi-channel cut-and-count experiment, given an expected background yield b_i, and an expected signal yield s_i in channel $1 \leq i \leq N$, the expected number of events in channel i is:

$$E[n_i] = b_i(\boldsymbol{\theta}) + \mu s_i(\boldsymbol{\theta}), \tag{5.32}$$

where μ is a signal strength parameter. Both the signal and background yields depend on a set of nuisance parameters (NPs), $\boldsymbol{\theta}$, which are used to model the effects of systematic uncertainties. A Poisson likelihood model for the POI μ, given an observation n_i, is constructed as:

$$\mathscr{L}_i(\mu, \boldsymbol{\theta} \mid n_i) = \text{Poisson}\left(n_i \mid b_i(\boldsymbol{\theta}) + \mu \cdot s_i(\boldsymbol{\theta})\right) \cdot p(\boldsymbol{\theta}), \tag{5.33}$$

where

$$\text{Poisson}(n \mid \lambda) \equiv e^{-\lambda} \frac{\lambda^n}{n!}. \tag{5.34}$$

The notation $\mathscr{L}_i(\mu, \boldsymbol{\theta} \mid n_i)$ is used to stress the fact that the likelihood is not a probability density function of the POIs. The prior distribution for NPs is assumed to be Gaussian:

$$p(\theta_\kappa) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\theta_\kappa^2}. \tag{5.35}$$

5.3.2 Nuisance parameters

NPs introduce a Bayesian component to an otherwise frequentist treatment. The NPs modify the nominal expected signal or background yield as:

$$s_i(\boldsymbol{\theta}) = s_i \cdot \prod_{\kappa}^{\dim \theta} \pi(\theta_\kappa). \tag{5.36}$$

The NPs are described either as rate-changing or as shape-changing.

Rate-changing nuisance parameters

Rate-changing NPs are conventionally chosen to be represented by a log-normal distribution, which gives:

$$\pi(\theta_\kappa) = (1 + \sigma_\kappa)^{\theta_\kappa}, \tag{5.37}$$

where σ_κ is the uncertainty on the NP, obtained from some auxiliary measurement or theory. Using a log-normal distribution ensures that the predictions, calculated by varying the NP, cannot go negative.

Shape-changing nuisance parameters

Shape-changing NPs are produced by recreating a template within the analysis where the systematic uncertainty is varied according to the boundaries of its central 68% confidence interval. For the likelihood function to be applicable, a “morphing” procedure is performed to obtain a continuous relation between θ_κ and $\pi(\theta_\kappa)$. The morphing algorithm interpolates between the central template $\pi(0)$ and two templates varied at a single deviation $\pi(\pm\sigma_\kappa)$ using a polynomial of the third degree, whereas a linear extrapolation is applied outside that region.

5.3.3 Profile likelihood ratio

The likelihood is a function of the POI, i.e. the signal strength μ, and the NPs. In this context, the primary purpose of frequentist statistical inference is the determination (fitting) of the signal strength given the input data, obtained by maximising the likelihood function. On the other hand, the NPs are specific to the measurement and are not of great general interest. To deal with the prior distributions of NPs, the profile likelihood, obtained by finding a conditional set of NPs that maximise the likelihood for a given value of μ, is used:

$$L(\mu) = \sup_{\boldsymbol{\theta}} \mathscr{L}(\mu, \boldsymbol{\theta} \mid \boldsymbol{n}). \tag{5.38}$$

The profile likelihood is a function of the POIs only, and can be used in frequentist inference. The profile likelihood ratio is a normalised version of the profile likelihood:

$$\lambda(\mu) = \frac{\sup_{\boldsymbol{\theta}} \mathscr{L}(\mu, \boldsymbol{\theta} \mid n)}{\sup_{\boldsymbol{\theta},\boldsymbol{\mu}} \mathscr{L}(\mu, \boldsymbol{\theta} \mid \boldsymbol{n})} \equiv \frac{\mathscr{L}(\mu, \hat{\hat{\boldsymbol{\theta}}}(\mu) \mid \boldsymbol{n})}{\mathscr{L}(\hat{\mu}, \hat{\boldsymbol{\theta}}, \mid \boldsymbol{n})}, \tag{5.39}$$

where the denominator is the overall maximum unconstrained likelihood. We have introduced $\hat{\hat{\boldsymbol{\theta}}}(\mu)$ and $\hat{\boldsymbol{\theta}}$ as the conditional (for a fixed μ) and unconstrained (arbitrary μ) maximum likelihood estimators of $\boldsymbol{\theta}$, respectively, and $\hat{\mu}$ is the maximum likelihood estimator of μ.

In a search of a new process that can only increase the mean event rate beyond what is expected from background alone, the signal strength is strictly non-negative, i.e. $\mu \geq 0$, and so a one-sided profile likelihood ratio is used instead:

$$\tilde{\lambda}(\mu) = \begin{cases} \frac{\mathscr{L}(\mu, \hat{\hat{\boldsymbol{\theta}}}(\mu) | \boldsymbol{n})}{\mathscr{L}(\hat{\mu}, \hat{\boldsymbol{\theta}} | \boldsymbol{n})} & \hat{\mu} \geq 0, \\ \frac{\mathscr{L}(\mu, \hat{\hat{\boldsymbol{\theta}}}(\mu) | \boldsymbol{n})}{\mathscr{L}(0, \hat{\hat{\boldsymbol{\theta}}}(0) | \boldsymbol{n})} & \hat{\mu} < 0. \end{cases} \tag{5.40}$$

The choice of the test statistic is motivated by the Neyman-Pearson lemma, which states that a likelihood-ratio test has the highest statistical power or the lowest type-II (false positive) error among all variants [134].

5.3.4 Hypothesis testing

A search for new physics aims to perform a statistical hypothesis test to distinguish between a null (background-only) hypothesis and an alternative (signal-plus-background) hypothesis, for which an appropriate test statistic needs to be defined. A single measurement corresponds to a single value of the test statistic, i.e. the test statistic summarises the characteristics of the data relevant to a specific statistical inquiry. The test statistic is chosen depending on whether a discovery is being tested or an UL is set.

Given a test statistic, x, distributed according to a probability density function $f(x)$, the p-value is calculated as follows for a one-sided right-tail test:

$$p = \int_{\hat{x}}^{\infty} f(x)\mathrm{d}x, \tag{5.41}$$

where $\hat{x}$ is the measured test statistic value. The null hypothesis is rejected in favour of the alternative hypothesis if $p < p_0$, where p_0 is the significance level set in advance. The p-value is commonly translated to the number of standard deviations of the normal distribution, such that the right-sided area of the normal distribution above this number equals p_0 [135]:

$$p_0 = 1 - \Phi(Z), \tag{5.42}$$

where $\Phi(x)$ is the cumulative distribution function of the normal distribution and Z is the required significance, conventionally taken to be $Z = 5$ for discovery. For setting ULs, $p_0 = 0.05$ (95% confidence level) is usually required, which corresponds to $Z \approx 1.64$.

Discovery

In discovery testing, the null hypothesis corresponds to the SM being true with $\mu = 0$. The alternative hypothesis is some BSM scenario being true, where $\mu > 0$. The discovery test statistic is [136]:

$$q_0 = \begin{cases} -2\ln\lambda(0) & \hat{\mu} \geq 0, \\ 0 & \hat{\mu} < 0, \end{cases} \tag{5.43}$$

where $\lambda(\mu)$ is defined in Equation 5.39. The value of test statistic is larger the more the data is inconsistent with the background-only hypothesis.

Upper limit setting

For deriving ULs, the roles of the null and alternate hypotheses are reversed. The null hypothesis corresponds to the case where a BSM scenario is true, whereas the alternate hypothesis is $\mu \leq \hat{\mu}$. The following test statistic is widely used [136]:

$$\tilde{q}_\mu = \begin{cases} -2\ln\tilde{\lambda}(\mu) & \hat{\mu} \leq \mu, \\ 0 & \hat{\mu} > \mu. \end{cases} \tag{5.44}$$

From the definition of the test statistic, it can be seen that higher values of $\tilde{q}_\mu$ represent greater incompatibility between the data and the hypothesised value of μ. In the case of an upward fluctuation of data, corresponding to $\hat{\mu} > \mu$, the test statistic is assigned a zero value.

Goodness-of-fit

A goodness-of-fit (GOF) test describes the compatibility of observation with a given prediction hypothesis in the absence of an alternative. This scenario arises when comparing a background-only hypothesis with the observation in the absence of a signal-plus-backgound hypothesis. In this case, there is no single optimal test, but a likelihood-ratio-like test statistic can be constructed as follows [137]:

$$q_{\mathrm{sat}} = -2\ln\lambda_{\mathrm{sat}}(\mu,\boldsymbol{\theta}) = -2\ln\frac{\mathcal{L}(\mu,\boldsymbol{\theta})}{\mathcal{L}_{\mathrm{sat}}}. \tag{5.45}$$

The saturated likelihood $\mathcal{L}_{\mathrm{sat}}$ is constructed by solving $\frac{\partial\mathcal{L}}{\partial\mu} = 0$ and hence provides the best possible fit to data. It effectively acts as a normalisation term. For the simple Poisson likelihood model, the test statistic would correspond to:

$$q_{\mathrm{sat}} = -2\ln\left(\frac{\mathrm{Poisson}(n \mid \lambda)}{\mathrm{Poisson}(n \mid n)}\right) = -2(n - \lambda + n\ln\lambda - n\ln n). \tag{5.46}$$

The goodness-of-fit test statistic asymptotically follows a χ^2 distribution.

The CL_s criterion

For a single measurement, a single value of $\tilde{q}_{\mathrm{obs}}$ will be obtained. The p-value given this measurement can be calculated as:

$$p_{\mathrm{s+b}} = \int_{\tilde{q}_{\mathrm{obs}}}^{\infty} f\left(\tilde{q}_\mu \mid \mu\right) d\tilde{q}_\mu, \tag{5.47}$$

where $f\left(\tilde{q}_\mu \mid \mu\right)$ is the probability distribution of $\tilde{q}$. The equation would have to be solved to find a value of μ which gives the required significance level corresponding to a p-value $p_{\mathrm{s+b}} = 0.05$. However, it turns out that it is preferable to use the so-called CL_{s} criterion instead [138], which is defined as:

$$p_{\mathrm{s}} = \frac{p_{\mathrm{s+b}}}{p_{\mathrm{b}}}, \tag{5.48}$$

where we have normalised by the background-only p-value:

$$p_{\mathrm{b}} = \int_{\tilde{q}_{\mathrm{obs}}}^{\infty} f\left(\tilde{q}_{\mu} \mid 0\right) d\tilde{q}_{\mu}. \tag{5.49}$$

The p-value is effectively penalised by dividing by p_{b}. As a result, p_{s} is found to be a more conservative approach than $p_{\mathrm{s+b}}$, and the difference between the two is especially pronounced in case of strong overlap between $f\left(\tilde{q}_{\mu} \mid \mu\right)$ and $f\left(\tilde{q}_{\mu} \mid 0\right)$. Figure 5.7 shows scenarios of moderate (left) and strong (right) overlap between the two test statistics. In both cases, $p_{\mathrm{s+b}}$ is small, suggesting that the null hypothesis should be rejected. However, excluding such cases based on $p_{\mathrm{s+b}}$ is undesirable in the latter case where the null and the alternative hypotheses are not well separated. On the other hand, using p_{s} would reject the alternative hypothesis in the former, but not in the latter case. The use of p_{s} remains controversial among statisticians; a more detailed discussion can be found in Refs. [139], [140].

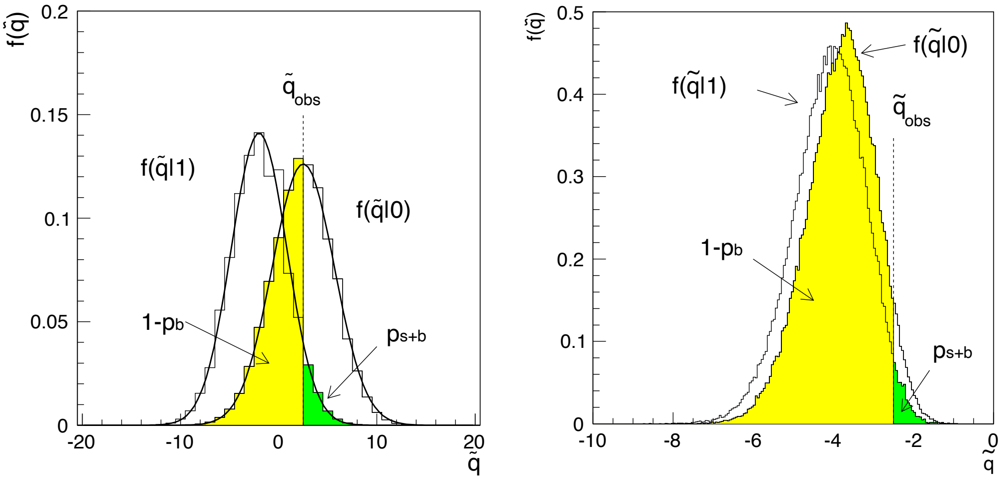


Figure 5.7: The CL_{s} criterion construction in case the signal+background and background hypotheses, denoted as $f\left(\tilde{q}_{\mu} \mid 1\right)$ and $f\left(\tilde{q}_{\mu} \mid 0\right)$, respectively, for well separated (left) and strongly overlapping (right) scenarios.

The asymptotic approximation

The probability distribution $f\left(\tilde{q}_{\mu} \mid \mu\right)$ cannot be obtained from the first principles, and approximate approaches are used to derive it. One computationally expensive solution is to rely on MC simulation by generating toy experiments that approximate the probability distribution. The second approach, used in this analysis, makes use of the asymptotic approximation [136]. The approximation is based on Wilk's theorem [141], which states that the probability distribution of a likelihood-ratio test statistic $f\left(\tilde{q}_\mu \mid 0\right)$ can be approximated by a χ^2 distribution if the null hypothesis is true and the sample size is sufficiently large. An extension by Wald [142], which applies to the alternative hypothesis, states that the probability distribution $f\left(\tilde{q}_\mu \mid \mu\right)$ asymptotically follows a non-central χ^2 distribution.

The Asimov data set

The Asimov data set is used to obtain expected limits from simulation. The Asimov data set is a fictitious data set in which all free parameters are set to their expected values from simulation [136]. The Asimov data set allows approximating the median outcome under a given hypothesis straightforwardly.

Discovery significance

In the asymptotic limit, the median expected discovery significance can be approximated from the Asimov data set as [136]:

$$Z = \sqrt{2\left((s+b)\ln\left(1+\frac{s}{b}\right)-s\right)}, \tag{5.50}$$

where s (b) are the expected signal (background) yields. The approximation is a useful quantity for identifying optimal event selection criteria in the design phase of a physics search.

Nuisance parameter impacts

The impact of a NP determines the shift in the best-fit value of a POI when the NP is fixed to $\pm 1\sigma$ of its nominal value:

$$\text{impact} = \Delta\mu^{\pm} = \hat{\mu}_{\theta\pm\sigma_\theta} - \hat{\mu}. \tag{5.51}$$

The impact reveals the correlation between the POI and the NP in question. NPs with low impacts values do not contribute significantly to the best fit value of the POI.

Chapter 6

Deep neural network-based displaced jet tagger

6.1 Introduction

Experiments at the CERN LHC routinely employ ML algorithms to perform event reconstruction, particle identification, event classification, and other tasks when analysing data samples [143]. Specifically, ML techniques have been widely adopted to classify jets. For example, jets originating from the hadronisation of b quarks exhibit characteristic experimental signatures that dedicated algorithms can exploit to identify b jets in a procedure known as b-tagging. B mesons, with proper lifetimes of $\mathcal{O}(10^{-12}\,\mathrm{s})$, travel distances of approximately 1–10 mm before decaying. As a result, charged particle tracks in b jets can originate from displaced vertices and have significant impact parameter values.

The latest b-tagging algorithm developed by the CMS Collaboration is the DeepJet tagger [144], a multiclass classifier that discriminates between jets originating from the hadronisation of heavy- or light-flavour quarks or gluons with substantial performance gains over its predecessors. The algorithm is based on a DNN suited to exploit low-level particle-level features. This information can be combined with high-level engineered features, such as, for example, the pseudoangular distance between the jet axis and the four-momentum sum of all the tracks, which have been used in preceding b-tagging algorithms [145]. Likewise, the ATLAS collaboration adapts DL by combining jet constituent inputs and the outputs of low-level algorithms into a single BDT or a DNN [146].

Inspired by the advances in b-tagging, we have developed a novel displaced jet tagger based on a DNN to achieve high sensitivity to models predicting displaced jets [1]. The main difference with respect to flavour tagging is that the LLP lifetime is a free model parameter, unlike the fixed heavy meson lifetime. As a result, the properties of the resulting jets will range from detector-prompt to highly displaced decays at the edge of the detector acceptance. To consider multiple possibilities at once, the network was parameterised based on the LLP lifetime. The network is trained on jets obtained from simulated and real pp collision events with DA to ensure that only well-modelled jet input features are exploited. In this Chapter, the architecture of the algorithm is motivated, and its technical implementation is described. The modelling of the DNN output in data and the performance achieved for different LLP models is presented.

6.2 Signal samples

The simplified split SUSY models, used for both training and evaluation of the displaced jet tagging algorithm, are defined by three parameters: the $c\tau_0$ of the gluino, the mass of the gluino $m_{\tilde{g}}$, and the neutralino mass, $m_{\tilde{\chi}_1^0}$. In training the displaced jet algorithm, direct detection of the gluino decay is targeted. The following mass parameter space, motivated by prompt SUSY searches, is considered when generating the samples: $1000 < m_{\tilde{g}} < 2500\,\mathrm{GeV}$, and $m_{\tilde{g}} - m_{\tilde{\chi}_1^0} > 100\,\mathrm{GeV}$. The production of gluino pairs is generated at LO precision in QCD with the MadGraph5_amc@nlo program [147]. The decay of the gluino is performed with the pythia program. The rhadrons package within the pythia program describes the formation of R hadrons through the hadronisation of gluinos [148], [149]. The gluino is assumed to decay democratically to any light quark flavour pair. The samples are generated separately for fourteen fixed, geometrically-spaced values of gluino $c\tau_0$, ranging from 30 µm to 30 m, resulting in approximately 3.4M events in total. The lower and upper bounds of the $c\tau_0$ range are motivated by the $\mathcal{O}(10\,\mu\mathrm{m})$ position resolution of the tracker and the physical dimensions of the CMS detector, respectively. Figure 6.1 shows the number of generated events in bins of the gluino and $\tilde{\chi}_1^0$ mass for two selected gluino lifetimes. Overall, the samples are preferentially generated to have more events for challenging compressed mass spectra, i.e. $m_{\tilde{g}} - m_{\tilde{\chi}_1^0} \ll m_{\tilde{g}}$, which result in softer jets with a lower selection efficiency.

A similar procedure is performed for the GMSB- and RPV-inspired benchmark models used as auxiliary samples to assess the tagger performance. These samples are generated only for two benchmark lifetimes of 1 mm and 1 m. The two lifetimes are motivated by the physical extent of the tracker and calorimeter systems, respectively. Altogether, we define the following four long-lived SUSY benchmark scenarios, which result in different flavours of displaced jets in the final state:

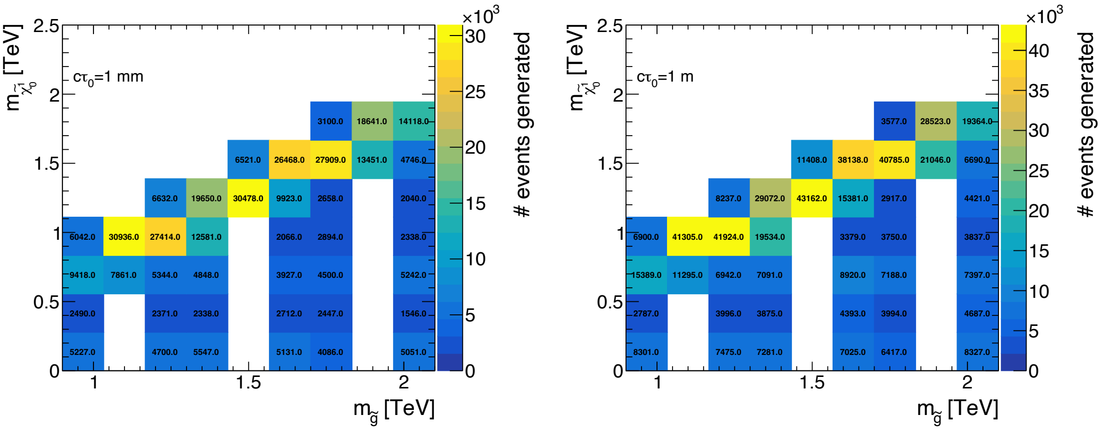


Figure 6.1: Split SUSY samples used for training the DNN jet tagger algorithm. The number of generated events as a function of the gluino and $\tilde{\chi}_1^0$ mass is shown for lifetimes of $c\tau_0$ = 1 mm (left) and 1 m (right).

Uncompressed A split SUSY scenario with a large mass difference between the gluino and the $\tilde{\chi}_1^0$, namely $(m_{\tilde{g}}, m_{\tilde{\chi}_1^0}) = (2000, 0)$ GeV. The scenario results in energetic final state jets resulting from the hadronisation of displaced uds quarks.

Compressed A nearly degenerate split SUSY spectrum, $(m_{\tilde{g}}, m_{\tilde{\chi}_1^0}) = (1600, 1400)$ GeV, with relatively soft final state jets compared to the uncompressed scenario.

GMSB A GMSB scenario with a long-lived gluino of 2500 GeV and a light gravitino of mass 1 keV. In this case, the jets originate from the hadronisation of displaced gluons.

RPV An RPV scenario with a long-lived top squark of 1200 GeV. In this model, the stop decays result in displaced b quarks, which hadronise and traverse the detector further before decaying, resulting in a tertiary decay vertex. The production of a tertiary vertex occurs because of the finite proper lifetime of B mesons of $c\tau_0 \approx 0.5$ mm. The other top squark decay product, a displaced lepton, can also be incidentally clustered into a jet by the PF algorithm, but such cases are explicitly rejected.

6.3 Definition of jet labels

The displaced jet tagger is trained as a multiclass classifier using jets from simulated QCD multijet, $t\bar{t}$, and long-lived gluino events. A sound definition of both SM jet and displaced jet labels (truth categories) is crucial to achieving optimal performance. As the hadronisation of partons is a highly stochastic process, resulting in intermediate partons in FSR or heavy flavour quark decay chains, jet labelling is generally somewhat ambiguous. We start the discussion by referring to a widely used SM jet labelling scheme called “ghost labelling” [150] and show why it is not well suited to label displaced jets from LLP decays. To this end, we propose an alternative labelling scheme.

6.3.1 Ghost labelling

The ghost labelling procedure assigns each reconstructed jet a parton flavour (u, d, s, g, c, b). The algorithm starts by considering as input all heavy-flavour hadrons at the truth (generator) level. The four-momenta of these hadrons are downscaled to minuscule values, and they become “ghost” hadrons, retaining only the original directional information. The jet clustering algorithm is then rerun by adding these truth-level particles to the reconstructed PF candidate collection. Since the momenta of ghost hadrons are negligible, their inclusion in jets does not affect the overall jet properties.

Consequently, a jet is labelled b or c jet if it contains at least one ghost hadron of the relevant kind. B hadrons can decay into D hadrons, and a jet containing both hadron flavours is labelled b jet. After this step, the remaining unlabelled jets are considered candidates for light flavour (u, d, s, g) labels. In this case, partons produced in the hard interaction before hadronisation are used as ghost particles. The flavour of a jet is then determined from the clustered ghost partons by choosing the flavour of the hardest parton. The remaining jets that do not have any label assigned are called pileup jets. The procedure is summarised as Algorithm 2.

The validity of the ghost labelling stems from the fact that the ghost particles follow the jet direction. The algorithm could be applied to label jets by choosing the LLPs or their (partonic) decay products as ghost particles. However, such partons produced at a displaced vertex can still interact strongly with each other and lead to additional jets, as shown in Figure 6.2. The resulting jets are not necessarily produced in the direction of the initial partons (shown in blue) nor the LL gluino (shown in red). Due to this loss of directionality, neither the LLP nor its decay products can be used as ghost particles. One possibility, however not explored in this thesis, would be to use the final copies of partons (shown in green). A significant shortcoming of using such an approach is that these partons might overlap with prompt ISR jets or UE activity, and the fraction of energy stemming from the LLP decay would not be dominant.

Two examples of decays of long-lived gluinos ($\tilde{g} \longrightarrow q\bar{q}\tilde{\chi}_1^0$), displaying the issue of using ghost tagging for LLP labelling, are shown in Figure 6.3. In each case, the final-state particles resulting from one of the quarks’ hadronisation are clustered into multiple jets. At the same time, ghost tagging can only account for a ghost particle producing a single

Algorithm 2 Ghost labelling

```
Require: Gen-level B, D hadrons and partons, PF candidates, jet clustering algorithm
Scale down the momenta of generator-level heavy-flavour hadrons and partons to
infinitesimally small values.
while not all jets clustered do
    Run jet clustering algorithm
    if any ghost particle clustered inside the jet then
        if B hadron among ghost particles then
            label as b jet
        else
            if D hadron among ghost particles then
                label as c jet
            else
                label by hardest parton flavour (u, d, s, g)
            end if
        end if
    else
        label as pileup jet
    end if
end while
```

jet. The algorithm may even fail to associate the ghost particle with any jets if the jets are sufficiently distanced in $\eta\phi$ space.

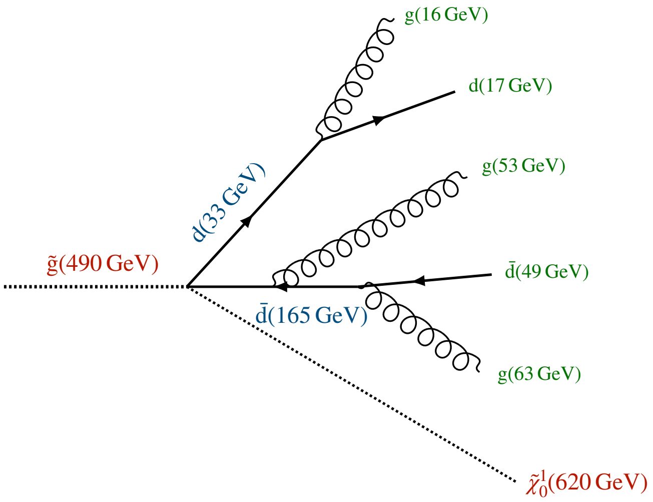
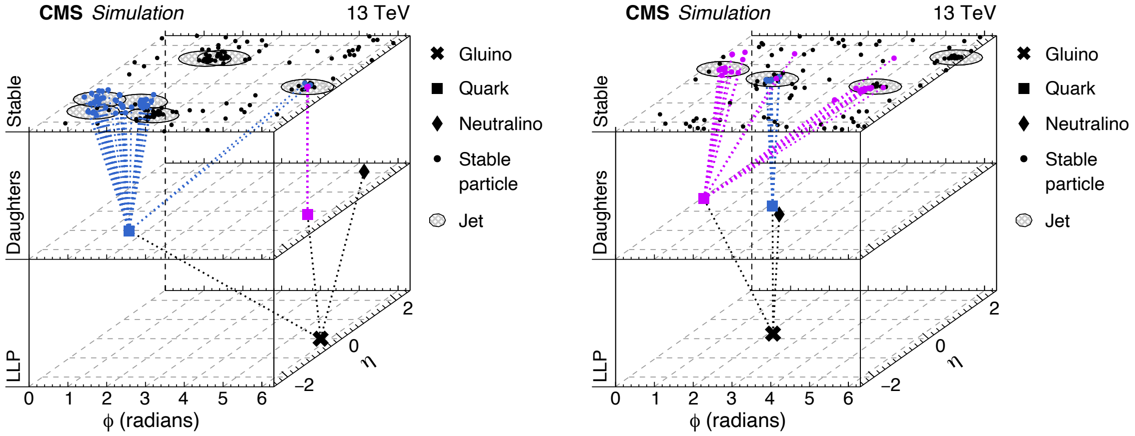


Figure 6.2: Example of a decay of a long-lived gluino R-hadron (in red) to a pair of d quarks (in green) and the $\tilde{\chi}_1^0$ (also in red) is shown. Several additional partons (in green) result from hard gluon radiation before hadronisation. The R-hadron or its daughters might not follow the direction of the final particles and thus cannot be reliably used for ghost labelling. The particle p_T is shown in brackets.



(a) (b)

Figure 6.3: Two examples of long-lived gluino decay chains, constructed from information provided by MC event generators. The positions of various particles in the $\eta\phi$ plane are shown: the LLP gluino and its daughter particles are shown in the lower and middle planes, respectively; the upper plane depicts the location of the stable particles after hadronisation, with shaded ellipses overlaid to indicate the reconstructed jets. Each quark and its decay is assigned a unique colour. The dotted lines indicate the links between parent and daughter particles. Figure taken from Ref. [1].

6.3.2 Displaced jet labelling

A novel generator-level procedure to label displaced jets was devised using information about the displacement of particles and their decay chain at the generator level. The algorithm identifies truth-level jets that have most of their energy originating from a displaced LLP decay vertex. The procedure works for labelling any LLP type; it is not specific to long-lived gluino decays.

First, all stable particles are grouped according to their vertex position. The resulting particle groups are linked via LLPs by considering LLP production and decay vertices. Next, jets at the generator level, clustered from stable generator particles that exclude neutrinos and the LSP, are associated with the vertices by splitting the jet momentum according to the vertex positions of the clustered particles. The jet-vertex shared-momentum fraction is defined per vertex v and jet with constituents j as:

$$f_v(\text{jet}) = \frac{\left(\sum_j \mathbf{p}_j \middle| \, j \in \text{vertex } v\right) \cdot \mathbf{p}_\text{jet}}{p^2_\text{jet}}, \quad f_v(\text{jet}) \in [0;\ 1], \quad \sum_v^{\text{vertices}} f_v(\text{jet}) = 1\,. \tag{6.1}$$

Jets at the generator level are associated with the vertex from which most of its momentum originates. This means choosing the vertex v, which maximises $f_v(\text{jet})$. This criterion prevents the coincidental association of jets containing very few or very soft displaced particles with a displaced vertex, while most of the clustered particles stem from the hard interaction or the UE instead. Finally, if the vertex v associated with the jet is an LLP decay vertex, the jet is labelled as an LLP jet. The labelling procedure for generator-level jets is summarised as Algorithm 3.

Algorithm 3 Displaced jet labelling

```
Require: Stable particles, jets, LLP decay vertex position $\mathbf{L}_{\text{LLP}}$
Cluster stable particles into vertices
for jet in jets do
    for vertex $v$ in vertices do
        $\mathbf{p}_v^{\text{vtx}} = 0$
        for constituent $i$ in jet do
            if $i \in v$ then
                $\mathbf{p}_v^{\text{vtx}} = \mathbf{p}_v^{\text{vtx}} + \mathbf{p}_i$
            end if
        end for
        $f_v = \mathbf{p}_v^{\text{vtx}} \cdot \mathbf{p}_{\text{jet}} / p_{\text{jet}}^2$
    end for
    $\hat{v} = \text{argmax}\, f_v, v \in \text{vertices}$
    if $\mathbf{L}_{\hat{v}} = \mathbf{L}_{\text{LLP}}$ then
        label jet as displaced
    end if
end for
```

An important distinction is that this new procedure applies to generator-level jets, unlike the ghost labelling algorithm, used with reconstructed jets instead. Hence, as a final step, the displaced jet label is transferred to any reconstructed jet found within $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2} < 0.4$ of the generator-level jet. The LLP label has priority and is assigned to signal jets over any other label obtained via ghost labelling.

It should be noted that a perfect split between the defined SM labels (b, c, uds, g) and the LLP jet label is not achieved. During initial tests in NN training, the displaced jet probability never reached 100% confidence for true displaced jets. This effect was due to genuine displaced jets contaminating the background classes in the signal samples due to the imperfect labelling scheme. As a result, jets with SM labels from signal samples are discarded during training, and only LLP jets are retained. QCD multijet and $t\bar{t}$ simulated samples are used to populate the background jet classes instead.

6.4 Training setup

In this Section, key ingredients of the DNN tagger setup are detailed: the input features used to train the displaced jet tagger, the chosen DNN architecture, the network parameterisation based on LLP lifetime, the kinematic resampling based on the signal jet kinematics, hyperparameter optimisation, and the technical implementation of the training.

6.4.1 Input features

The network features (input variables) are split into four broad categories: charged constituent properties, neutral constituent properties, SV features, and global jet variables. Zero-padding, i.e. replacing missing constituent values by zero, is used in the case of a less-than-required number of constituents for all categories. The network has 662 input variables in total.

Figure 6.4 shows several representative input features for signal and background jets. No single variable provides strong discrimination between the jet classes, motivating using a DNN that can exploit correlations between the variables. In addition, signal jet features differ significantly depending on the LLP $c\tau_0$, highlighting the need for a parametric network.

Charged PF constituents

The list of charged PF candidates properties retained for training is described in Table 6.1. Up to 25 candidate tracks are retained for training for each jet, ranked hierarchically by decreasing transverse impact parameter significance. The impact parameter, d_{xy}, is defined as the spatial distance between the originating pp collision and the track at its point of closest approach, while its significance is obtained by dividing the impact parameter by the uncertainty on its measurement: $d_{xy}/\sigma(d_{xy})$.

Neutral PF constituents

The list of neutral PF candidates per jet retained for training is described in Table 6.2. For each jet, up to 25 neutral candidates are retained for training. The neutral candidates are sorted by their minimum distance to an SV. However, if no SV is reconstructed in the jet, neutral candidates are sorted by their momentum instead.

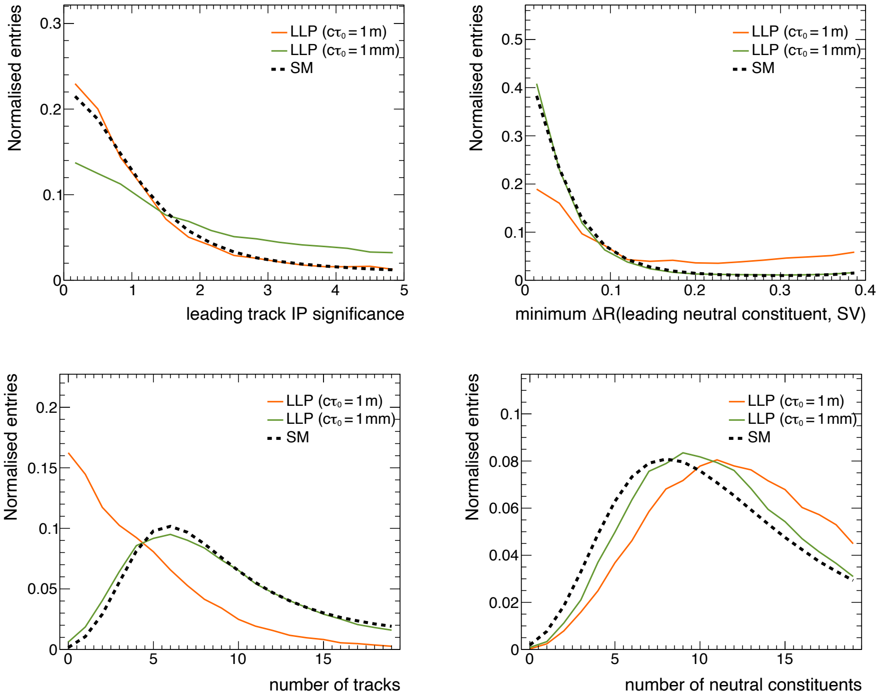


Figure 6.4: Various tagger input features for SM background jets and signal split SUSY jets for $c\tau_0$ values of 1 mm and 1 m.

Table 6.1: Charged PF candidate features.

Internal name	Description
`cpf_trackPtRel`	p_T relative to jet axis
`cpf_trackEtaRel`	η relative to jet axis
`cpf_trackPPar`	momentum parallel to jet axis
`cpf_trackDeltaR`	ΔR(track,jet)
`cpf_trackSip2dVal`	transverse impact parameter
`cpf_trackSip2dSig`	significance of transverse impact parameter
`cpf_trackSip3dVal`	impact parameter
`cpf_trackSip3dSig`	impact parameter significance
`cpf_trackJetDistVal`	distance from the track to the jet
`cpf_ptrel`	track p_T/jet p_T
`cpf_drminsv`	ΔR between track and closest SV within the jet
`cpf_vertex_association`	indicates whether the track is used in the PV fit
`cpf_from_PV`	indicates whether the track is associated to the PV
`cpf_puppi_weight`	pileup likelihood assigned by the PUPPI algorithm [151]
`cpf_track_chi2`	track χ^2
`cpf_track_ndof`	track fit number of degrees of freedom
`cpf_track_quality`	track reconstruction quality

Table 6.2: Neutral PF candidate features.

Internal name	Description
`npf_ptrel`	particle p_T/jet p_T
`npf_deltaR`	ΔR(jet,candidate)
`npf_isGamma`	flag for passing loose photon identification requirements
`npf_hcal_fraction`	fraction of energy deposited in the HCAL
`npf_drminsv`	ΔR between track and closest SV within the jet
`puppi_weight`	pileup likelihood assigned by the PUPPI algorithm

Secondary vertices

For each jet, up to four reconstructed SVs are retained for training, ranked by the same procedure as the charged PF candidates. There is often no reconstructed SV; in this case, the network can utilise other discriminating features such as the charged PF candidate impact parameters instead. The properties associated with SVs used in training are listed in Table 6.3.

Table 6.3: SV features.

Internal name	Description
`sv_pt`	SV p_T
`sv_deltaR`	ΔR between jet axis and SV flight direction
`sv_mass`	invariant mass of reconstructed SV
`sv_ntracks`	number of tracks associated to the SV
`sv_chi2`	SV χ^2
`sv_ndof`	SV fit number of degrees of freedom
`sv_dxy`	SV transverse impact parameter
`sv_dxysig`	SV transverse impact parameter significance
`sv_d3d`	SV impact parameter
`sv_d3dsig`	SV impact parameter significance
`sv_costhetavpv`	$\cos\theta$ of the SV with respect to PV
`sv_enratio`	ratio of SV energy with respect to the jet

Global jet features

Global jet features comprise kinematic variables (p_T, η), the number of charged and neutral PF candidates, and the number of SVs, among others. The list also includes various features used in the training of a previous CMS b-tagging algorithm. The complete list of features is presented in Table 6.4.

Table 6.4: Global jet features.

Internal name	Description
`global_pt`	jet p_T
`global_eta`	jet η
`global_ncpf`	number of charged PF candidates
`global_nnpf`	number of neutral PF candidates
`global_nsv`	number of SVs
`global_trackSumJetEtRatio`	E_T of four-momentum of all tracks, divided by the jet E_T
`global_trackSumJetDeltaR`	ΔR between four-momentum of all tracks and jet axis
`global_vertexCategory`	RecoVertex, PseudoVertex or NoVertex
`global_trackSip2dValAboveCharm`	transverse impact parameter value of the first track that raises the invariant mass of the system of all tracks above 1.5 GeV
`global_trackSip2dSigAboveCharm`	transverse impact parameter significance of the first track that raises the invariant mass of the system of all tracks above 1.5 GeV
`global_trackSip3dValAboveCharm`	impact parameter value of the first track that raises the invariant mass of the system of all tracks above 1.5 GeV
`global_trackSip3dSigAboveCharm`	impact parameter significance of the first track that raises the invariant mass of the system of all tracks above 1.5 GeV
`global_jetNSelectedTracks`	number of selected tracks
`global_jetNTracksEtaRel`	number of tracks for which trackEtaRel is calculated

6.4.2 Network architecture

A complete schematic of the network architecture is shown in Figure 6.5. Overall, the choice of the network's input features and the architecture of the feature extractor is heavily influenced by the DeepJet model used for tagging b jets in CMS [144].

The network's first step is to simultaneously reduce the vast number of input features to a more tractable number and engineer highly discriminating latent (hidden) features. The vector of features of each charged and neutral PF candidate and each SV is passed through several 1×1 convolutional layers with a kernel size of 1. The goal of these layers is to compress the input features in several stages and finally obtain a reduced vector of only the most discriminating features. Convolutional layers of sizes 64, 32, 16, 8 and 4 are used, depending on the input features. The 1-dimensional convolutional layers change the shape of the feature vector for each particle based on a linear combination of the preceding layer features, followed by a non-linear activation function. After each convolutional layer, a leaky ReLU activation function is used. Dropout layers are inserted throughout the network after activation layers with a 10% dropout rate.

After the final convolutional layer, the compressed latent feature vectors of the charged and neutral PF candidates and the SVs are flattened, i.e. the information from all PF candidates is combined into a single vector. This resulting vector is also combined with the global jet features. At this stage, the lifetime is introduced to parameterise the network. The performance of this approach is further studied in Section 6.4.3. The resulting feature vector is fed into a series of fully connected (dense) layers to perform jet class and domain prediction. The dense layers have 100 neurons in the label predictor and 50 in the domain classifier, respectively. The final outputs are fed through a softmax activation function in both branches, and the loss function used is categorical cross-entropy. The resulting set of five jet class probabilities comprises gluons (g), light-flavour quark jets (uds), and two heavy-flavour jet classes: b and c jets. Finally, displaced jets from LLP decay, here labelled as LLP, are included. The domain class probability output is binary, i.e. either data or simulation.

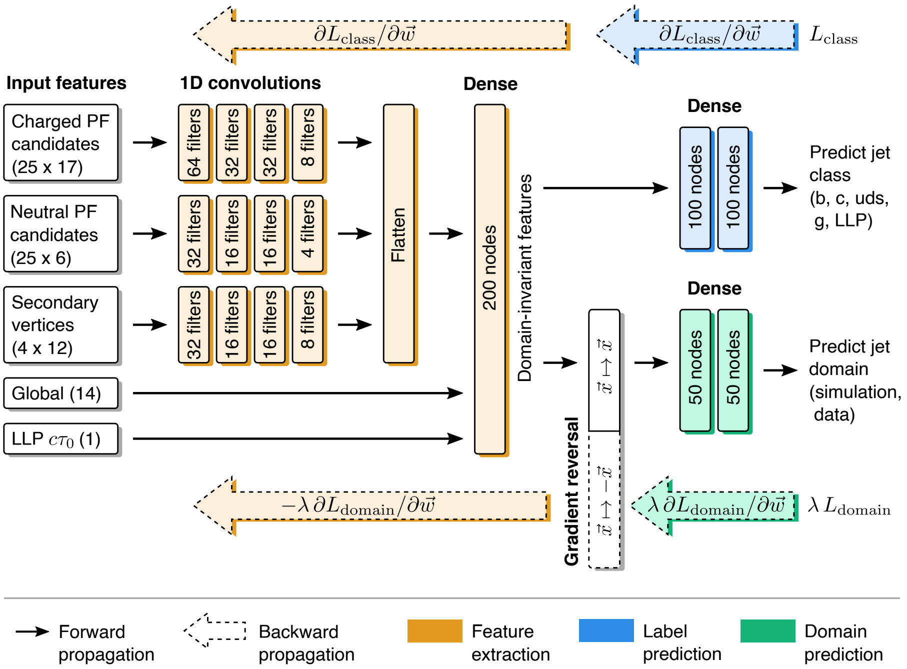

$\partial L_{\mathrm{class}}/\partial\vec{w}$
$\partial L_{\mathrm{class}}/\partial\vec{w}$
L_{class}
Input features
1D convolutions
Dense
Charged PF candidates (25 x 17)
Neutral PF candidates (25 x 6)
Secondary vertices (4 x 12)
Global (14)
LLP $c\tau_0$ (1)
64 filters
32 filters
32 filters
8 filters
32 filters
16 filters
16 filters
4 filters
32 filters
16 filters
16 filters
8 filters
Flatten
200 nodes
Domain-invariant features
Dense
100 nodes
100 nodes
Predict jet class (b, c, uds, g, LLP)
Dense
$\vec{x} \to \vec{x}$
50 nodes
50 nodes
Predict jet domain (simulation, data)
Gradient reversal
$\vec{x} \to -\vec{x}$
$-\lambda\,\partial L_{\mathrm{domain}}/\partial\vec{w}$
$\lambda\,\partial L_{\mathrm{domain}}/\partial\vec{w}$
$\lambda\, L_{\mathrm{domain}}$
Forward propagation
Backward propagation
Feature extraction
Label prediction
Domain prediction

Figure 6.5: An overview of the DNN architecture, which comprises convolutional and dense layers; the numbers of filters and nodes are indicated. Dropout layers and activation functions are not shown. The input features are grouped by object type; $(m \times n)$ indicates the maximum number of objects (m) and the number of features per object (n). The gradients of the class and domain losses with respect to the weights ω used during backpropagation are shown. Figure taken from Ref. [1].

6.4.3 Parametric network

In contrast to b jets, displaced signal jets result from LLPs with arbitrary lifetimes. A simple solution for this could be to train a different model depending on the LLP lifetime;

however, this is not very practical computationally. Nonetheless, such a model is used as a benchmark for parameterising the network. In the first iteration, the NN was parameterised as a function of the LLP lifetime. A parameterised or parametric NN means that the network output, i.e. class probabilities, depends on a specific input feature, treated as a free parameter [152]. For our case, we denote the DNN output as a function of hypothesised LLP lifetime, $P(\text{class}|c\tau_0)$, class $\in \{\text{b}, \text{c}, \text{uds}, \text{g}, \text{LLP}\}$. Furthermore, we only consider the displaced jet class probability $P(\text{LLP}|c\tau_0)$, which we use as a discriminating variable (likelihood score). The user must specify a $c\tau_0$ value to evaluate the probability. The parameterisation allows obtaining a single network operating over a wide range of lifetimes with extrapolation capabilities. It allows all displaced jets, regardless of their displacement, to be included for training.

The parameterisation is achieved by adding the LLP lifetime to the vector of global jet features listed in Table 6.4. For signal jets, the LLP lifetime is the lifetime of the mother gluino. As the LLP lifetime is undefined for background jets, a “fake” lifetime is sampled from the $c\tau_0$ distribution of LLP jets used in the training. The main consequence is that the network cannot directly discriminate based on $c\tau_0$ but learns the behaviour of displaced jets for different lifetimes through correlations with other jet features. The relative importance of these correlations is expected to differ based on the LLP lifetime. The performance of this approach compared with training a non-parameterised network is further studied in Section 6.5.3.

Similarly, background jets are selectively sampled to match the kinematic distributions of LLP jets, namely p_T and η. This downsampling prevents direct discrimination based on jet kinematic quantities dictated in simulated MC samples by a specific physical process. For instance, it is undesirable to learn to discriminate entirely based on the fact that LLP jets from a heavy gluino decay have a significantly harder p_T spectrum than, say, heavy-flavour jets from $\text{t}\bar{\text{t}}$. However, the correlation of the p_T and η with other jet input features can still be exploited to obtain optimal discrimination in bins of jet p_T and η. As a final step, the jets are further downsampled to obtain a balance of all five jet classes. The resulting p_T, η, and $c\tau_0$ distributions of the resampled jets used to train the class predictor are shown in Figure 6.6.

For the first iteration of the algorithm, pileup jets were ignored in training, as they pose a completely negligible source of background for $p_\text{T} > 50$ GeV, i.e. in the bulk of the LLP jet p_T distribution. However, they were later included when extending the tagger to target significantly softer jets resulting from HNL decays, as such LLP jets have an average p_T of 40 GeV.

6.4.4 Control regions

To train the NN with DA and validate its output, we define two independent μ+jets and $\mu\mu$+jets CRs, corresponding to events with one or two muon candidates, respectively, and at least several jets. Both CRs correspond to a dataset of $36\,\mathrm{fb}^{-1}$ collected in 2016 using a trigger condition that requires the presence of a single isolated muon with $p_\mathrm{T} > 24\,\mathrm{GeV}$. The μ+jets CR comprises, in approximately equal measure, events from the W+jets and $\mathrm{t\bar{t}}$ processes. The $\mu\mu$+jets CR contains Z/γ^*+jets events with subdominant contributions from $\mathrm{t\bar{t}}$. The requirements for both CRs are summarised in Table 6.5. Requirements on $p_\mathrm{T}^\mathrm{miss}$ or a $p_\mathrm{T}^\mathrm{miss}$-like quantity are imposed in the CRs to mimic the scenario of long-lived gluino decays, which result in significant genuine $p_\mathrm{T}^\mathrm{miss}$ from the decay of the LSP or jets decaying outside the detector acceptance. The specific requirements for the physics objects are further described in Chapter 4.

Table 6.5: Overview of selection requirements in data CRs for training and validating DA.

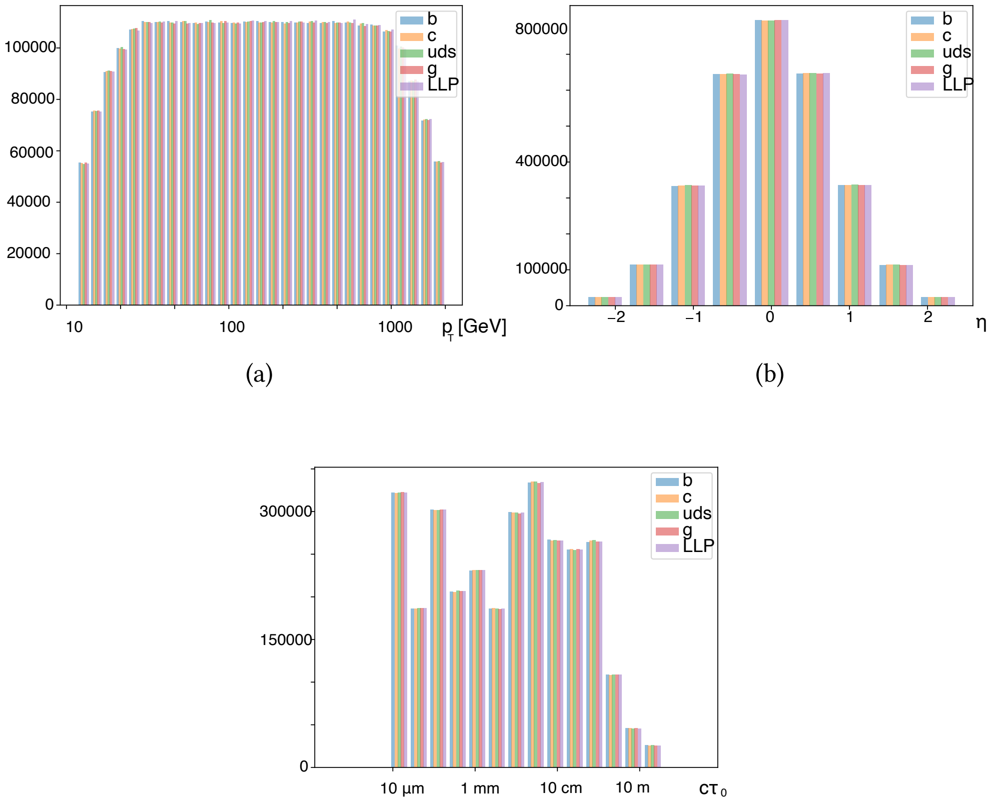
	μ+jets (training)	$\mu\mu$+jets (validation)
Lepton selection		
Tight muons ($I^\mu_\mathrm{rel.} < 15\%$)	$p_\mathrm{T} > 26\,\mathrm{GeV}$	$p_\mathrm{T}(\mu_{1,2}) > 26, 15\,\mathrm{GeV}$
Veto muons ($p_\mathrm{T} > 10\,\mathrm{GeV}, I^\mu_\mathrm{rel.} < 25\%$)	✓	✓
Veto electrons ($p_\mathrm{T} > 15\,\mathrm{GeV}$)	✓	✓
Jet selection		
Number of central jets ($p_\mathrm{T} > 30\,\mathrm{GeV}$)	≥ 2	≥ 2
Veto forward jets ($p_\mathrm{T} > 50\,\mathrm{GeV}$)	✓	✓
Event observables		
$p_\mathrm{T}^\mathrm{miss} > 150\,\mathrm{GeV}$	✓	
$p_\mathrm{T}(\mu\mu) > 100\,\mathrm{GeV}$		✓

6.4.5 Training samples

Only jets in the tracker acceptance ($|\eta| < 2.4$) and with uncorrected $p_\mathrm{T} > 20\,\mathrm{GeV}$ are retained for training. Uncorrected p_T is used in order not to make the model dependent on a specific set of JECs. As ROOT files used in this analysis store events as entries rather than jets, suitable jets are unpacked from different events and randomly mixed and split into multiple new files. The resulting files used for training contain one jet per entry stored randomly, with only relevant input features saved.

In total, the training sample contains about 4M displaced jets from split SUSY samples. The background samples of QCD and $\mathrm{t\bar{t}}$ contain tens of millions of jets; however, after the kinematic resampling described in Section 6.4.3, we are left with about 20M jets of all classes in total. Minibatches of 10 000 MC jets, containing about 2000 jets of each class, taken from simulated long-lived gluino, QCD multijet and $t\bar{t}$ samples, are used for training with SGD. In total, one epoch consists of processing 2000 batches.

The training of the domain classifier uses minibatches of 10 000 jets, obtained from events in the μ+jets CR, drawn from pp collision data, W+jets, and $t\bar{t}$ samples in relative, approximate quantities of 55%, 35% and 10%, respectively. The selection leads to about 1.2M jets in total for the domain classifier, and jets must be reused multiple times per epoch. For the simulated jets in each domain minibatch, the class label is not used.



(a)

(b)

(c)

Figure 6.6: Training jet (a) transverse momentum, (b) pseudorapidity, and (c) proper lifetime distributions for signal and background classes after kinematic resampling.

6.4.6 Technical implementation

The NN is trained using the TensorFlow [153] system with an interface to Keras [154], which facilitates the construction of NN architectures. To optimise the network weights, TensorFlow uses automatic differentiation to implement backpropagation. Backpropagation can be accelerated using a graphical processing unit (GPU), resulting in substantially shorter training times than training on a central processing unit (CPU). Hence, an NVIDIA Tesla P100 GPU is used to speed up training. The class and domain minibatches are merged into a single minibatch of 20 000 jets to train all network parts. The choice of the minibatch size is dictated by the memory constraints of the GPU (16 GB of RAM). One epoch of training takes approximately one hour. 80% of the total sample is used for training and the remaining 20% serves as a validation data set to monitor for overtraining.

A custom ROOT-to-TensorFlow pipeline was developed to read ROOT files and interface them with the TensorFlow queue system. The same pipeline additionally performs on-the-fly jet resampling based on jet p_{T} and η and fake $c\tau_0$ generation for background jets. A more detailed description with a minimal working example can be found in Ref. [155].

The training is performed for up to fifty epochs using the Adam optimiser with default settings implemented in Keras. The initial learning rate is chosen to be 0.01; its decay schedule is optimised as discussed in Section 6.4.7. Additional auxiliary NN models are trained by using ten-fold cross-validation to assess the uncertainty of a given model. The cross-validation is achieved by splitting the training dataset into ten partitions and obtaining ten permutations, each leaving out a single partition. The models obtained this way are used to provide the uncertainty contour on the nominal model, trained using the entire available data set.

6.4.7 Hyperparameter optimisation

This Section describes the optimisation procedure for some of the hyperparameters used in the tagger training. In all cases, an independent validation data set is used for the optimisation.

Learning rate

Arguably, the learning rate is the most critical hyperparameter in NN training, and hence it is optimised first. To find the optimal learning rate decay schedule, the network is trained exclusively on the label predictor for fifty epochs, with the learning rate decay parameterised as:

$$\eta = \frac{0.01}{1 + \eta_0 \cdot n}, \tag{6.2}$$

where η_0 is the decay rate parameter and n is the epoch number. An appropriate metric has to be chosen to evaluate the network performance. In a realistic physics analysis,

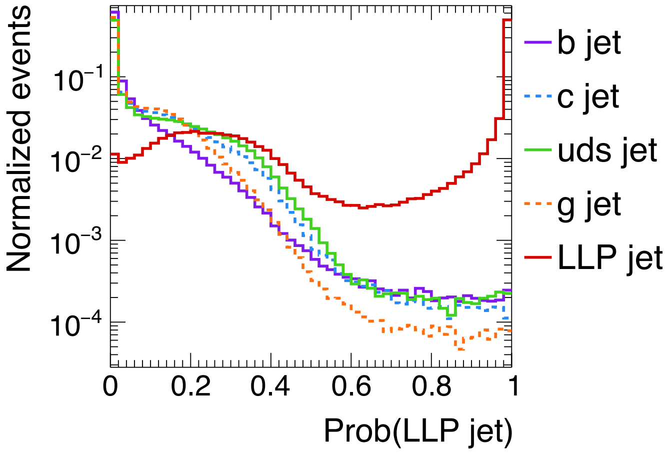


Figure 6.7: An example of probability distributions in a validation data set for the five jet classes obtained after training the NN. The $c\tau_0$ values used to evaluate the probability were generated in the same way as for the training data set.

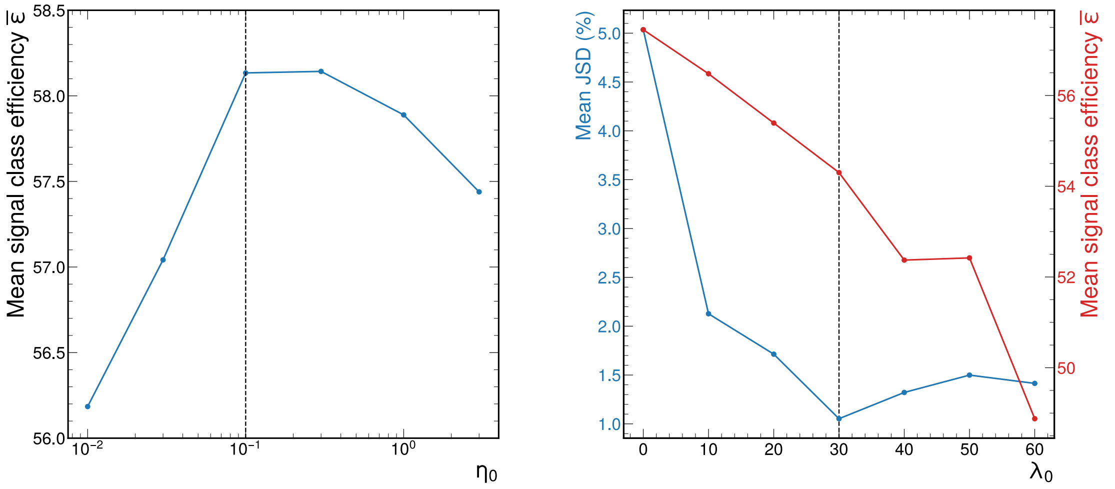
searching for an LLP decaying to displaced jet(s), the SM background rate is typically several orders of magnitude larger than the signal yield. To reduce the SM background to a manageable rate, obtaining a low background jet false positive or “mistag” rate is imperative. The efficiency of the tagger to correctly identify the LLP jet class and to mistag background jets as LLP depends on a chosen working point, defined by a threshold (“cut”) requirement on the jet class probability $P(\mathrm{LLP}|c\tau_0)$. To assess the tagger performance, we apply four different thresholds on the LLP class probability (an example of which is shown in Figure 6.7). Each threshold is tailored to reject a specific background jet class, corresponding to a fake identification (mistag) rate of 0.01%. The LLP jet efficiency resulting from applying such a threshold is denoted as $\varepsilon(\mathrm{LLP} \mid \mathrm{background@}10^{-4})$. Finally, the mean signal efficiency is calculated as follows:

$$\bar{\varepsilon} = \frac{\varepsilon(\mathrm{LLP} \mid \mathrm{uds@}10^{-4}) + \varepsilon(\mathrm{LLP} \mid \mathrm{g@}10^{-4}) + \varepsilon(\mathrm{LLP} \mid \mathrm{b@}10^{-4}) + \varepsilon(\mathrm{LLP} \mid \mathrm{c@}10^{-4})}{4}. \quad (6.3)$$

As shown in Figure 6.8a, the optimal value of $\eta_0 = 0.1$ corresponds to the training configuration with the largest average displaced jet tagging efficiency. No strong dependence on η_0 is observed, suggesting a well-chosen initial learning rate.

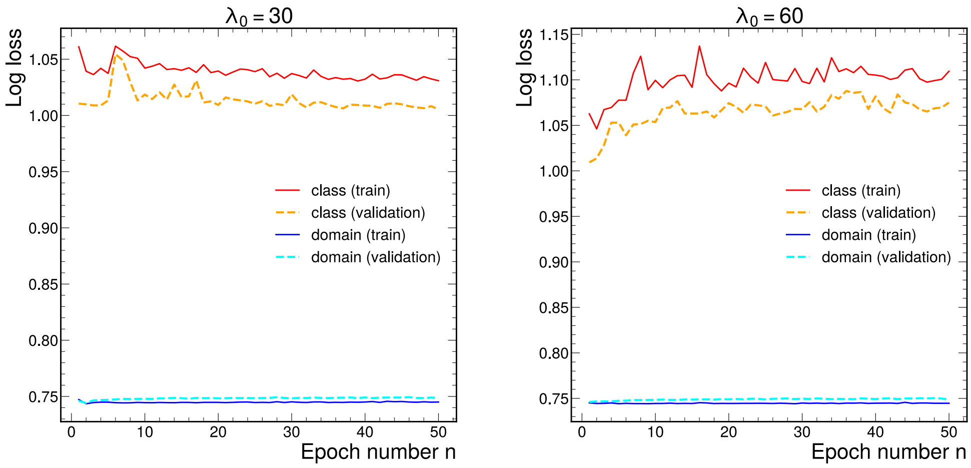
Domain loss weight

The parameter controlling the relative importance of the domain loss term is gradually increased according to the following schedule:



(a) (b)

Figure 6.8: Hyperparameter optimisation. In (a), the optimisation of the learning rate decay parameter η_0 with no DA is shown. A value of $\eta_0 = 0.1$ is chosen as the best performing scenario. In (b), DA hyperparameter optimisation is shown. The mean signal efficiency and the mean JSD score is investigated as a function of the hyperparameter λ_0. The value of $\lambda_0 = 30$ is chosen as optimal.



(a) (b)

Figure 6.9: The evolution of class and domain validation loss during training is shown with (a) $\lambda_0 = 30$ and (b) $\lambda_0 = 60$.

$$\lambda = \lambda_0 \left(\frac{2}{1 + \exp(-0.2 \cdot n)} - 1 \right), \tag{6.4}$$

which converges after around 30 epochs to the final domain loss weight, λ_0. This schedule ensures the network achieves an optimal set of weights for the class prediction first, i.e.

it is pre-trained on the label predictor. Here, λ_0 is set as a hyperparameter and optimised. To assess the effectiveness of different values of λ_0, we use the JSD score. The test is applied to distributions of the domain data set, evaluated at different $c\tau_0$ values ranging from 10 µm to 1 m, to ensure good data/MC agreement for different lifetimes. In addition, the mean signal efficiency $\bar{\varepsilon}$ is monitored to maintain adequate network performance, as shown in Figure 6.8. For values $\lambda_0 > 30$, the performance of the tagger drops significantly, while the JSD score plateaues, which suggested that increasing λ_0 further does not bring additional improvement. This effect is further shown in Figure 6.9, where the evolution of the loss function is shown for two models with $\lambda_0 = 30$ and $\lambda_0 = 60$. For the latter, the loss eventually starts to increase, suggesting an unbalanced value of λ_0. A value of $\lambda_0 = 30$ is ultimately chosen, corresponding to acceptable performance degradation and simultaneously a significantly improved JSD score over a model with no DA.

For the following Sections, we have chosen two benchmark models: a nominal “DA” model with $\lambda_0 = 30$ and a reference “no DA” model with $\lambda_0 = 0$. The data/MC agreement and the performance of the two models are explored in the following Sections.

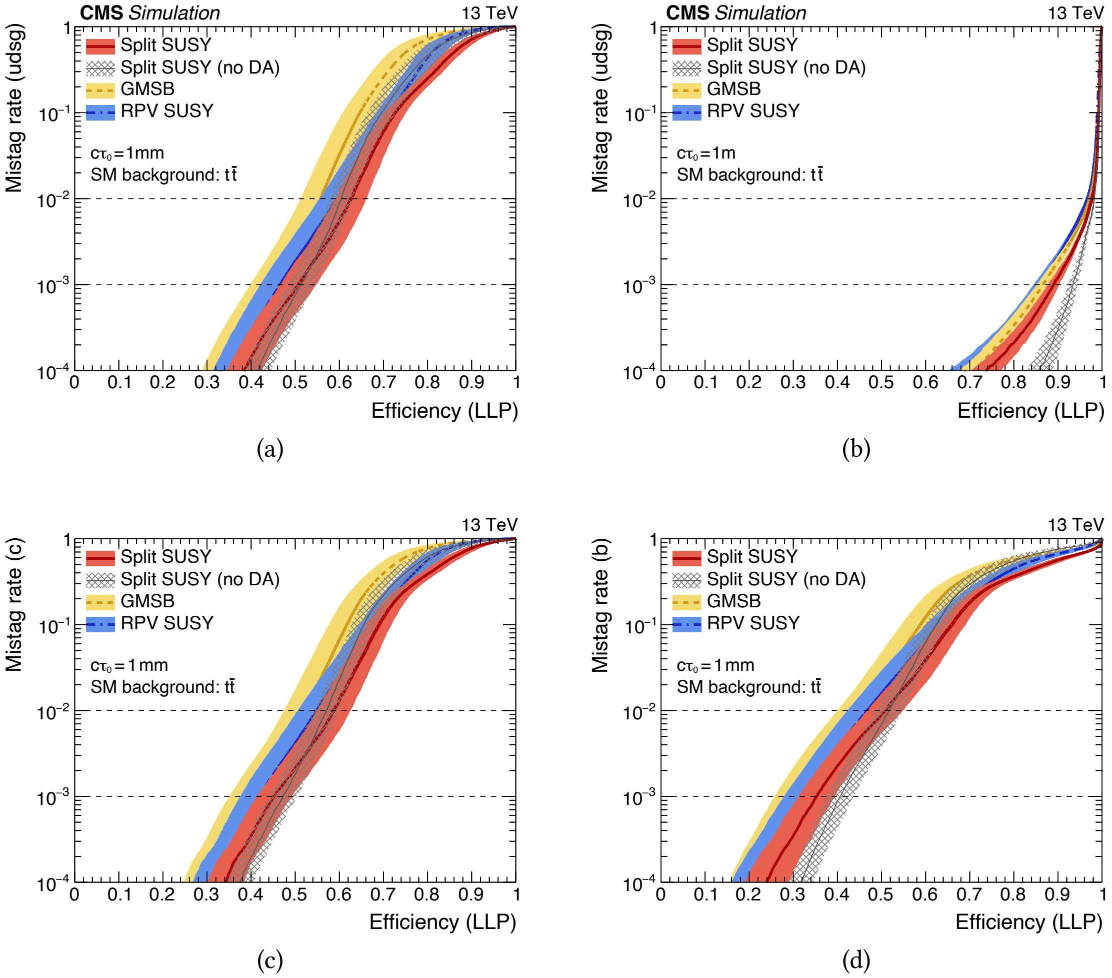
6.5 Tagger performance

The tagger performance is studied using an inclusive sample of simulated $\mathrm{t\bar{t}}$ events. For these studies, the tagger performance is assessed under realistic conditions, i.e. under the typical background jet p_T and η distributions expected in pp collision data. The $\mathrm{t\bar{t}}$ samples provide both light-flavour (udsg) jets, through ISR and hadronic decays of the W boson, and heavy-flavour (b, c) jets, with $p_\mathrm{T} > 30\,\mathrm{GeV}$ and $|\eta| < 2.4$. The performance is benchmarked against a range of long-lived SUSY scenarios listed in Section 6.2, using LLP jets obtained from those samples.

6.5.1 Inclusive performance

The generalisation of the tagger to LLP jets of different flavours is investigated for the benchmark lifetimes of 1 mm and 1 m. The two values of $c\tau_0$ give greater weight to the roles of the tracker and calorimeter systems, respectively. The receiver operating characteristic (ROC) curves that provide the LLP jet tagging efficiency as a function of the background mistag rate for different scenarios are shown in Figure 6.10. It can be seen that all flavours of displaced jets have a similar performance within the uncertainties. Excellent signal jet efficiencies $\varepsilon(\mathrm{LLP}) \equiv \varepsilon(\mathrm{LLP} \mid \mathrm{t\bar{t}}@0.01)$ of 40 and 70% are obtained for split SUSY models with $c\tau_0$ values of 1 mm and 1 m, respectively.

Furthermore, the performance is better when discriminating against background light-flavour than heavy-flavour jets, which is expected, as heavy-flavour jets are significantly displaced with respect to the IP, corresponding to ε(LLP) of 40, 30, and 25% discriminating against udsg, c, and b jets, respectively. Lastly, it can be seen that for the 1 mm scenario, there is no significant degradation from using DA. However, ε(LLP) drops by 10–15% when using the DA model for higher lifetimes, suggesting the penalisation of mismodelled input features that provide significant discriminating power at higher jet displacements. The significant performance decrease from using DA indicates that a biased algorithm trained purely on simulation would have unrealistically good performance.



(a) (b)

(c) (d)

Figure 6.10: ROC curves showing the tagger performance for uncompressed split SUSY, GMSB and RPV SUSY scenarios: (a) LLP jets vs light-flavour jets for $c\tau_0 = 1$ mm; (b) LLP jets vs light-flavour jets for $c\tau_0 = 1$ m; (c) LLP jets vs c jets for $c\tau_0 = 1$ mm; (d) LLP jets vs b jets for $c\tau_0 = 1$ mm. For the uncompressed scenario, a reference no DA training is also included for comparison. Figures (a) and (b) taken from Ref. [1].

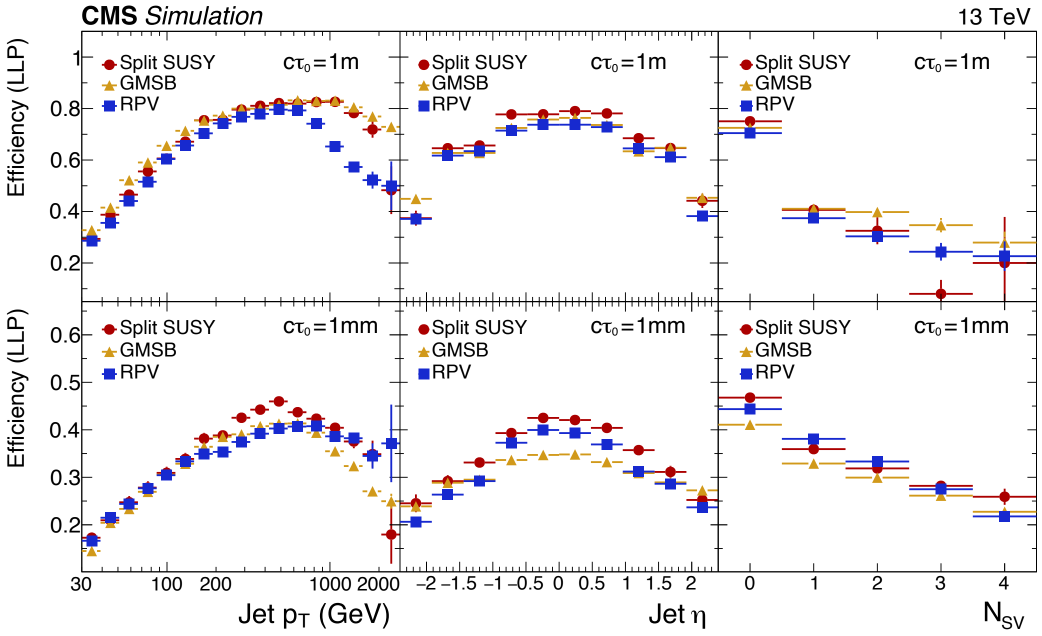


Figure 6.11: ε(LLP) as a function of the jet p_T, η, and N_{SV}. The efficiency curves are shown separately for the split (circular marker), GMSB (triangle marker), and RPV (square marker) SUSY benchmark models, assuming $c\tau_0$ values of 1 m (upper row) and 1 mm (lower row). Figure taken from Ref. [1].

6.5.2 Performance as a function of jet kinematic quantities

For the uncompressed split SUSY model, ε(LLP) is further studied versus jet p_T, η, and the number of SVs. The results are shown for benchmark lifetimes of 1 m and 1 mm in Figure 6.11. Efficiencies are highest for high-p_T, centrally produced jets with $N_{SV} = 0$. The latter observation demonstrates the complementary performance of the tagger compared to the more standard approach of relying on displaced SVs to identify LLPs. Poor SV reconstruction efficiency for highly displaced LLP decays could explain the decrease in performance with an increasing number of SVs.

6.5.3 Parameterisation performance

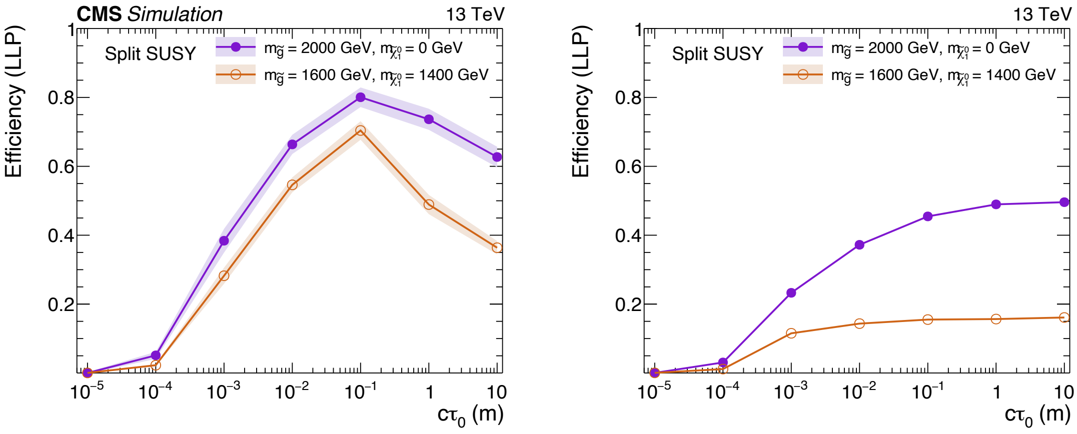
Training a parametric network might hypothetically result in performance degradation compared to a dedicated model for a given $c\tau_0$ value. To assess the generalisation of parametric training, three different scenarios are considered, all trained with no DA. The performance, quantified as ε(LLP) for all three scenarios, is evaluated using an uncompressed split SUSY data set with $c\tau_0 = 1$ mm:

1. Nominal training on all lifetime values, $\varepsilon = 0.39 \pm 0.04$;

2. Training exclusively on $c\tau_0$ = 1 mm, ε = 0.38 ± 0.04;
3. Training with all lifetime values included except for $c\tau_0$ = 1 mm, ε = 0.30 ± 0.03.

No difference between the first two approaches is observed, within the uncertainties, suggesting the performance of the parametric approach is adequate. On the other hand, the last scenario has an ε(LLP) value lower by 0.1, meaning such a model struggles to extrapolate to previously-unseen lifetimes.

6.5.4 Performance versus lifetime



(a) (b)

Figure 6.12: In (a), ε(LLP) as a function of the split SUSY $c\tau_0$ for compressed and uncompressed split SUSY scenarios is shown. In (b), ε(LLP) versus $c\tau_0$ for a shallow network, trained only on global jet features, is shown. Figure (a) taken from Ref. [1].

The tagger performance is investigated as a function of $c\tau_0$ for compressed and uncompressed split SUSY scenarios, as shown in Figure 6.12a. The best performance is achieved for $c\tau_0$ = 0.1 m for both cases. Furthermore, as shown in Figure 6.12b, the performance is evaluated for a shallow network, where all jet constituents are dropped from the training, and only global features are retained. Such a model results in a decrease in performance by ~ 60% for uncompressed scenarios and ~ 30% for compressed ones, showcasing the advantage of using a DNN instead.

6.5.5 Wrong lifetime test

To assess what happens if a “wrong” lifetime value is used when evaluating the network, ε(LLP) for $c\tau_0^{\text{gen}}$ = 1 m and 1 mm uncompressed split SUSY scenarios were evaluated by

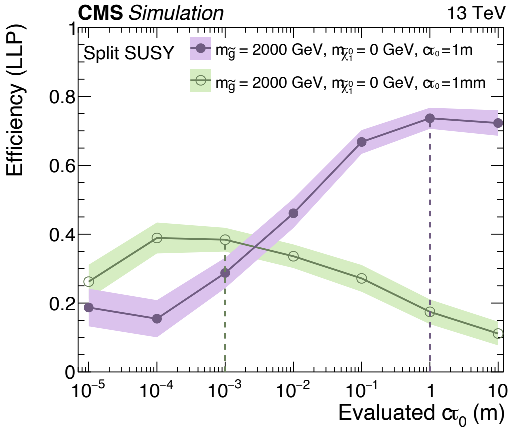


Figure 6.13: The DNN tagger efficiency ε(LLP) as a function of the lifetime used to evaluate the network. The purple (solid filling) and green (hollow filling) lines correspond to uncompressed split SUSY scenarios generated assuming a gluino lifetime of 1 m and 1 mm, respectively. Figure taken from Ref. [1].

providing the DNN tagger with a range of lifetimes $c\tau_0^{\text{eval}}$. As shown in Figure 6.13, the best performance is obtained when approaching the "right" lifetime, i.e. $c\tau_0^{\text{eval}} \approx c\tau_0^{\text{gen}}$, as expected for a parameterised network. Reversing this finding, it would seem that the tagger could predict the jet displacement, which was later on exploited when training the tagger for HNL scenarios, as detailed in Section 7.5.5.

6.6 Tagger validation

The LLP jet probability obtained from simulation can differ significantly from data in the absence of DA. In this Section, the modelling in control regions of $P(\text{LLP}|c\tau_0)$ is investigated. Since a given event has several jet candidates reconstructed, while the tagger is applied on a per-jet basis, an event-level maximum LLP jet probability defined from all selected jets in an event was considered instead:

$$P_{\text{max}}(\text{LLP}|c\tau_0) = \max_{\text{jets}} P(\text{LLP}|c\tau_0) \tag{6.5}$$

A comparison of the $P_{\text{max}}(\text{LLP}|c\tau_0)$ distributions obtained from pp collision data and simulated events in both the μ+jets and $\mu\mu$+jets CRs, using DA and "no DA" models is shown in Figure 6.14. The use of DA in the network leads to a significant improvement in the level of agreement in the distributions of $P_{\text{max}}(\text{LLP}|c\tau_0)$ for the two domains of data and simulation, with only minor residual differences remaining. This improvement is expected for the μ+jets CR, as the same events are used to train, evaluate, and optimise the domain classifier of the DNN. The $\mu\mu$+jets CR, comprising a statistically independent

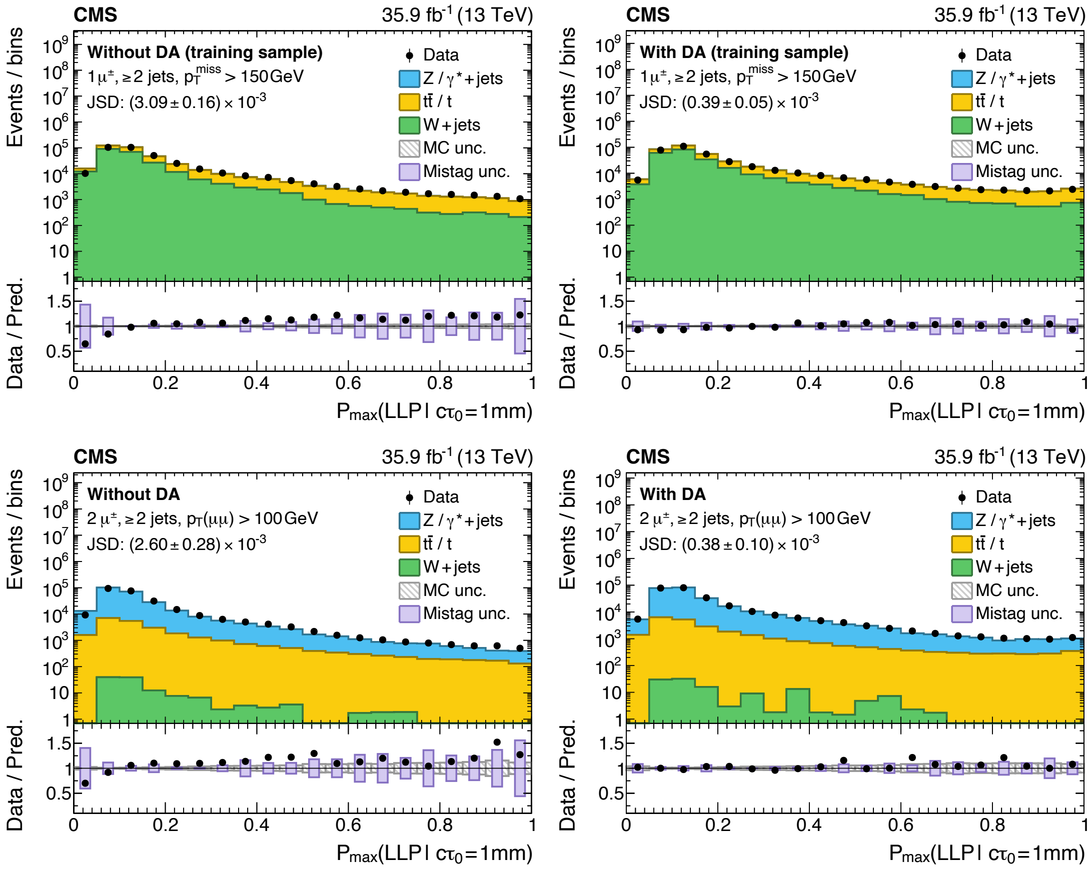


Figure 6.14: Distributions of the maximum probability for the LLP jet class obtained from all selected jets in each event, $P_{\mathrm{max}}(\mathrm{LLP}|c\tau_0)$. The distributions from data (circular marker) and simulated events (histograms) are compared in the μ+jets (upper row) and $\mu\mu$+jets (lower row) CRs, using a DNN trained without (left column) and with (right column) DA. All probabilities are evaluated with $c\tau_0 = 1\,\mathrm{mm}$. The statistical (hatched bands) and systematic (solid bands) uncertainties due to the finite-size simulation samples and the simulation mismodelling of the mistag rate, respectively, are shown. Figures taken from Ref. [1].

event sample dominated by a different physical process, validates the method. The application of DA leads to significantly reduced biases and uncertainties in the modelling of $P_{\mathrm{max}}(\mathrm{LLP}|c\tau_0)$, as the data/MC ratios per bin are closer to unity following the application of DA. The level of agreement between data and simulation is further quantified by the JSD score, which is reduced by an order of magnitude following the application of DA. The quoted uncertainties in JSD reflect the finite sizes of the data and simulated samples. Finally, an estimate of the systematic uncertainty in $P_{\mathrm{max}}(\mathrm{LLP}|c\tau_0)$ due to simulation mismodelling is obtained in a statistically independent sample of $\mu\mu$+jets that satisfy $p_{\mathrm{T}}(\mu\mu) < 100\,\mathrm{GeV}$, while the nominal $\mu\mu$+jets CR contains events with $p_{\mathrm{T}}(\mu\mu) > 100\,\mathrm{GeV}$. The uncertainty is estimated by weighting up or down simulated events by the following weight:

$$w^{\pm} = \Pi_i^{\text{jets}} \left(1 \pm (\xi_i - 1)\right), \tag{6.6}$$

where ξ_i is the ratio of counts from data and simulation in bin i of the $P(\text{LLP}|c\tau_0)$ distribution. The uncertainty band is shown in the bottom panel of Figure 6.14 in purple, and is significantly smaller for a network trained with DA.

6.7 Summary

This Chapter detailed the development of a robust DNN algorithm to identify displaced jets from LLP decays. The need for a new displaced jet labelling scheme was motivated. The DNN was trained on both data and simulation with DA. Overall, the algorithm showed excellent performance, rejecting 99.99% of background jets while keeping displaced jet efficiencies high, typically above 40%. Training with DA for lifetimes greater than 1 mm resulted in a 5–10% degradation in performance while substantially improving the overall agreement between data and simulation. The good modelling in CRs proves the utility of DA as a prophylactic tool to deal with mismodelled phenomena in simulation and improves the confidence with which such an algorithm can be applied in a search for BSM physics. Good performance observed even without a presence of a displaced SV establishes the tagger as an alternative approach to searches for LLP decay using SVs, which was exploited in a search for long-lived HNLs, described in Chapter 7.

Furthermore, the DNN tagger shows good generalisation to other models, which were not used in the training. This makes it a powerful tool able to detect generic displaced jet signatures. Training a single parametric model of the network was validated by not observing a degradation in performance compared to many single-point models. When testing a particular lifetime hypothesis, the best sensitivity was found when specifying that same lifetime as an input to the parametric tagger, confirming the tagger was working as intended.

Chapter 7

Search for heavy neutral leptons

This Chapter presents the analysis strategy of a search for long-lived HNLs coupling to all lepton generations using the DNN displaced jet tagger. First, the properties of a long-lived HNL signature within CMS and the analysis scope are discussed in Section 7.1. The generation of simulated signal samples is described in Section 7.2. The event selection and categorisation requirements are detailed in Section 7.3. The development of an event-level BDT, discriminating between signal and background events, which allows for an optimal preselection of the HNL decay phase space, is described in Section 7.4. Several extensions to the displaced jet tagger which increase the sensitivity to HNL decays are described in Section 7.5.

The analysis is performed as a blind analysis by excluding the signal region (SR) from data for the studies described in this Chapter. The backgrounds are estimated using a data-driven technique which is the subject of Chapter 8.

7.1 Heavy neutral lepton signatures at CMS

As first discussed in Section 2.4.3, long-lived HNL signatures accessible at the LHC occur in the mass range of about 1–20 GeV. It is difficult to directly trigger on HNL decay products at CMS in this mass range, as they are relatively soft. Hence, the potential to trigger on HNL events depends entirely on the production process, specifically, what other particle is produced in association with the HNLs.

HNLs can be produced in on-shell EW boson (W, Z, H) or heavy-flavour meson (B, D) decays. The NC process and gluon-gluon fusion H boson production cannot be triggered on as a neutrino is produced. On the other hand, vector-boson fusion H boson production only becomes important above an HNL mass of around 500 GeV. Heavy-flavour meson production is also challenging to trigger on as the final state particles are soft; in addition, this scenario is only kinematically allowed below the specific heavy-flavour meson mass of a few GeV.

CC production is the primary mode considered in this analysis, as it results in the associated production of a prompt lepton, denoted ℓ_1, which can be triggered on at CMS with relatively high efficiency. The cross-section times the BR for this process, valid for $m_{\mathrm{N}} < m_{\mathrm{W}}$, can be expressed in terms of the CC production cross-section as follows:

$$\sigma_{\mathrm{CC}} = \sigma(pp \longrightarrow \mathrm{W}) \cdot B(\mathrm{W} \longrightarrow \ell\nu) \sum_{\ell} |V_\ell|^2 \left(1 - \frac{m_{\mathrm{N}}^2}{m_{\mathrm{W}}^2}\right)^2 \left(1 + \frac{m_{\mathrm{N}}^2}{2m_{\mathrm{W}}^2}\right). \tag{7.1}$$

The decays of low-mass HNLs take place via off-shell EW gauge bosons. CC semileptonic HNLs decays, which are the subject of this thesis, are the most common. The semileptonic three-body decays via an off-shell W boson $\mathrm{N} \longrightarrow \ell_2\mathrm{W}^* \longrightarrow \ell_2\mathrm{q}\bar{\mathrm{q}}$ result in the production of an additional displaced lepton, denoted as ℓ_2, and displaced jet(s); this occurs almost twice as often as NC decays to two charged leptons and a neutrino $\mathrm{N} \longrightarrow \ell_2\ell_3\nu$. Finally, the HNLs can also decay hadronically via the NC, i.e. $\mathrm{N} \longrightarrow \nu\mathrm{Z}^* \longrightarrow \nu\mathrm{q}\bar{\mathrm{q}}$. However, this would result in a neutrino being produced instead of the ℓ_2, which is a much more experimentally challenging signature, and, thus, not considered in this analysis.

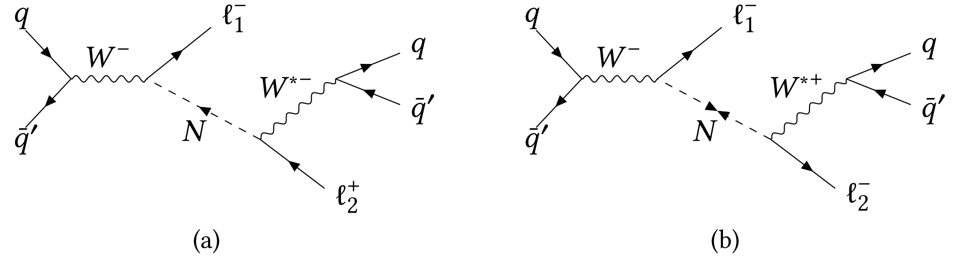


Figure 7.1: CC HNL production in association with a prompt trigger lepton ℓ_1 and decay to a displaced lepton (ℓ_2) and a pair of quarks. In (a), an opposite-sign (OS) dilepton final state is shown, possible for both Dirac and Majorana HNLs. In (b), a same-sign (SS) dilepton final state is shown, possible only for Majorana HNLs. The flavours of $\ell_{1,2}$ can differ.

To summarise, both the HNL production and decay proceed via the CC in this analysis:

$$\mathrm{W} \longrightarrow \ell_1\mathrm{N} \longrightarrow \ell_1\ell_2\mathrm{q}\bar{\mathrm{q}}, \tag{7.2}$$

also shown as a Feynman diagram in Figure 7.1. It is important to note that the semileptonic HNL decays differ kinematically depending on the HNL mass [156]. For $m_{\mathrm{N}} < 1$ GeV, the HNLs decay primarily via a two-body decay $\mathrm{N} \longrightarrow \ell_2\mathrm{M}$, where M is a light meson, instead of the aforementioned three-body decay $N \longrightarrow \ell_2 q\bar{q}$, where each of the quarks hadronise separately. This is not explicitly taken into account, as masses below 1 GeV are not considered in this analysis. In the latter case, the HNL lifetime dependence can be expressed as:

$$\tau_\mathrm{N} \propto m_\mathrm{N}^{-5} \left|V_\ell\right|^{-2}, \tag{7.3}$$

valid when assuming that the HNL couples to a single active neutrino generation. The equation holds for both Dirac and Majorana HNLs, and both scenarios are explicitly targeted in this analysis, allowing to distinguish between Majorana and Dirac HNLs. In the case of discovery, Majorana HNLs would result in observing LNV, while Dirac HNLs would not. For Majorana HNLs, the partial widths are twice as large, as the Majorana HNL is its antiparticle:

$$\Gamma_\mathrm{Majorana} = \Gamma^\mathrm{N}_\mathrm{Dirac} + \Gamma^{\overline{\mathrm{N}}}_\mathrm{Dirac} = 2\,\Gamma_\mathrm{Dirac}. \tag{7.4}$$

In prompt collider HNL searches, only HNL coupling to a single neutrino generation is assumed, while coupling to the other two is considered unconstrained. However, this is not justified in a long-lived HNL search, where the total decay width plays a crucial role in determining where the HNL will decay. Furthermore, as discussed in Section 2.3, single mixing angles might be primarily suppressed in the νMSM, while this does not hold for multiple mixing angles. In this analysis, the HNLs are considered to couple to all three lepton generations simultaneously in the most general case. Non-zero coupling to at least two generations means the flavours of ℓ_1 and ℓ_2 can, generally, differ, corresponding to lepton flavour violation (LFV).

7.2 Simulated signal samples

Samples of simulated signal and background events are used to define the event selection, assess the modelling of key observables in simulation, tune the usage of various MVA techniques to ensure optimal sensitivity, and predict the signal yields. The list of background samples used in this analysis is given in Table 5.1.

The MadGraph event generator is used to simulate the HNL production and decay at LO accuracy with varying HNL mass and V_ℓ by using the HNL model provided in Ref. [157]. In this analysis, only one active HNL, of either Dirac or Majorana nature, with equal couplings to the three lepton generations $V_\mathrm{e} = V_\mu = V_\tau$, is generated. Additional

ME weights, calculated by MadGraph, are used to reweight simulated events to an arbitrary different coupling scenario $(V'_{\mathrm{e}}, V'_{\mu}, V'_{\tau})$, allowing for scanning the coupling space efficiently, as only a single signal sample needs to be generated. The proper lifetime is expressed as a sum of partial widths, Γ_{ℓ}, of the three flavour states:

$$\frac{1}{\tau_0} = \Gamma_{\mathrm{tot}}(m_{\mathrm{N}}, V_{\mathrm{e}}, V_{\mu}, V_{\tau}) = \Gamma_e(m_{\mathrm{N}}, V_{\mathrm{e}}) + \Gamma_{\mu}(m_{\mathrm{N}}, V_{\mu}) + \Gamma_{\tau}(m_{\mathrm{N}}, V_{\tau}). \tag{7.5}$$

Technically, the coupling reweighting can only be performed if the proper HNL lifetime, $c\tau_0$, and the mass remain fixed, in order to keep the HNL kinematics unchanged:

$$\Gamma'_e(m_{\mathrm{N}}, V'_{\mathrm{e}}) + \Gamma'_{\mu}(m_{\mathrm{N}}, V'_{\mu}) + \Gamma'_{\tau}(m_{\mathrm{N}}, V'_{\tau}) = \Gamma_e(m_{\mathrm{N}}, V_{\mathrm{e}}) + \Gamma_{\mu}(m_{\mathrm{N}}, V_{\mu}) + \Gamma_{\tau}(m_{\mathrm{N}}, V_{\tau}). \tag{7.6}$$

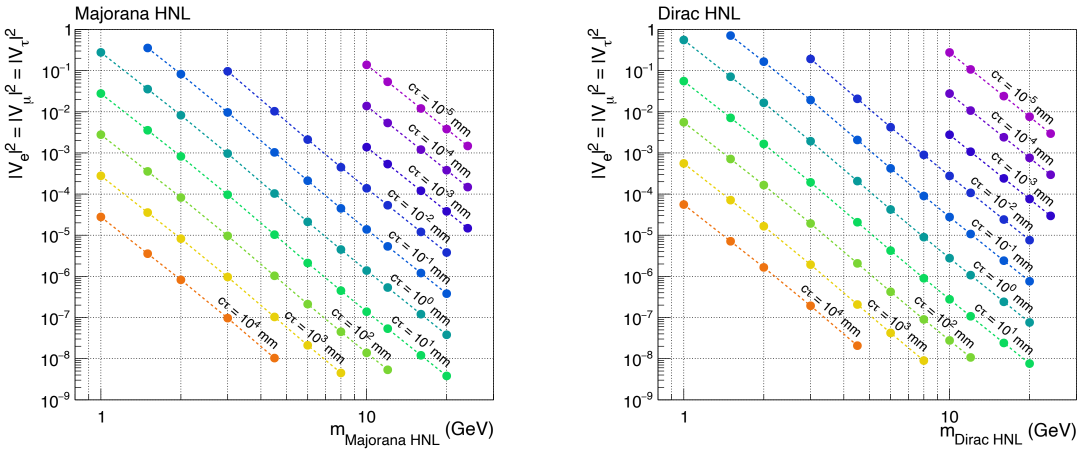


Figure 7.2: The considered signal scenarios with $|V_{\mathrm{e}}| = |V_{\mu}| = |V_{\tau}|$ in the mass-coupling plane for Majorana (left) and Dirac (right) HNLs. Scenarios with the same proper lifetime, $c\tau_0$, as calculated by MadGraph, are colour-coded and connected by dashed lines.

In this analysis, 76 signal samples were generated as shown in Figure 7.2. HNLs are generated at various mass points ranging from 1 to 24 GeV, with proper lifetimes ranging between $10^{-5} < c\tau_0 < 10^{4}$ mm. For each sample, 66 weights are stored for reweighting the sample to different coupling scenarios while keeping the total decay width fixed. The barycentric coordinates f_{e}, f_{μ}, and f_{τ}, with $f_{\mathrm{e}} + f_{\mu} + f_{\tau} = 1$, are introduced, which denote the relative ratios between the couplings $|V_{\mathrm{e}}| : |V_{\mu}| : |V_{\tau}| \equiv f_{\mathrm{e}} : f_{\mu} : f_{\tau}$. The chosen equidistant coupling points for reweighting are shown in Figure 7.3. Six benchmark scenarios are selected as detailed in Table 7.1, corresponding to couplings to single and multiple active neutrino generations. These scenarios are used for hypothesis testing described in Section 8.3.1.

The LO cross-sections were found to be compatible with the NLO cross-sections within

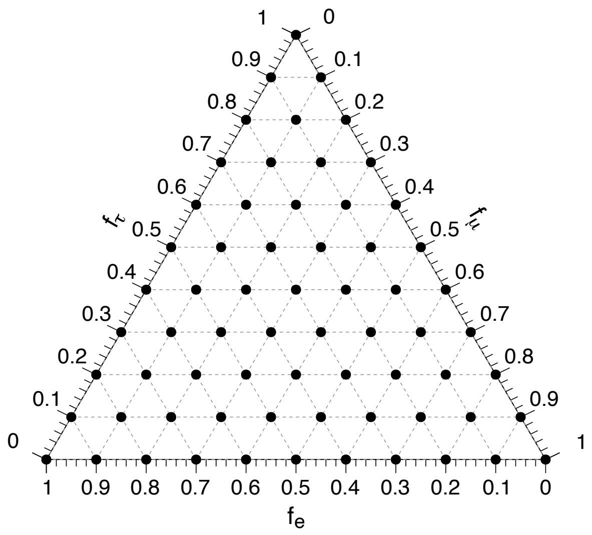


Figure 7.3: The considered 66 relative couplings points in barycentric coordinates ($f_e + f_\mu + f_\tau = 1$) for reweighting each signal sample that has been generated with $|V_e| = |V_\mu| = |V_\tau|$, i.e. $f_e = f_\mu = f_\tau = 1/3$.

Table 7.1: Benchmark HNL coupling scenarios given in barycentric coordinates.

Scenario	f_e	f_μ	f_τ
1	1/3	1/3	1/3
2	1	0	0
3	1/2	1/2	0
4	0	1	0
5	1/2	0	1/2
6	0	1/2	1/2

the estimated scale uncertainty. A $k = \sigma_{\mathrm{NLO}}/\sigma_{\mathrm{LO}} \approx 1.1$ correction factor was applied to the LO cross-sections for normalising each signal sample, and the LO scale uncertainty is taken as the theoretical uncertainty.

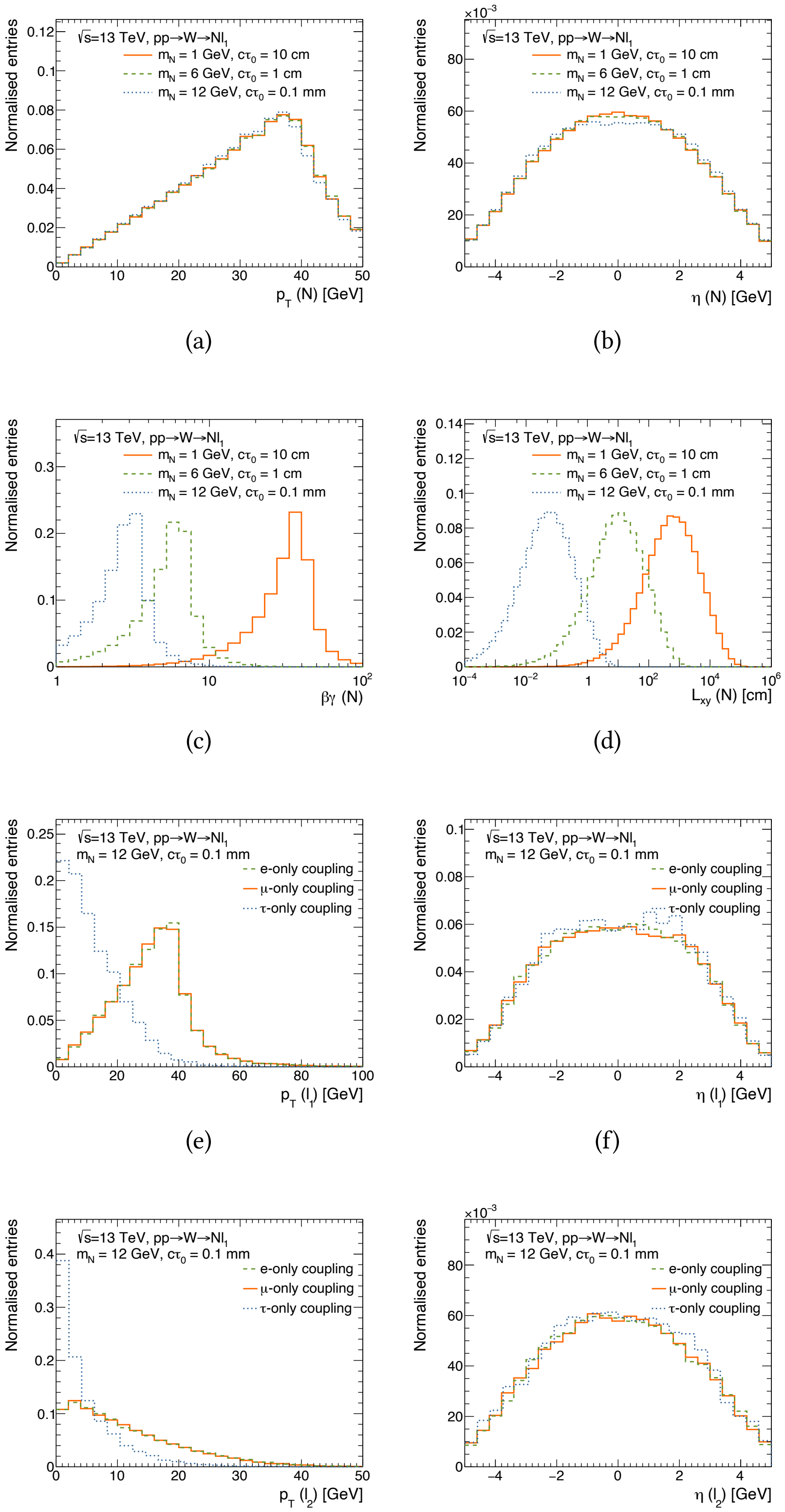
An overview of the BRs for various HNL masses is given in Table 7.2. For completeness, decays to fully leptonic and invisible final states are also listed. The semileptonic decays, which are relevant for this analysis, are further broken down by parton and lepton flavour. The HNL decays to tau leptons are kinematically forbidden when $m_{\mathrm{N}} < m_\tau$. Thus, in the specific case of $V_e = V_\mu = 0$ and $V_\tau > 0$, the only possible decay channel is via NCs that results in a three neutrino final state, $\mathrm{N} \longrightarrow \nu\bar{\nu}\nu$.

Generator-level kinematic properties of the simulated HNLs, namely the transverse momentum p_{T}, pseudorapidity η, boost $\beta\gamma$, and the distance travelled from the pp IP, L_{xy}, are shown for various benchmark signal models in Figure 7.4. The models correspond

Table 7.2: BRs for the tree-level decay modes of a Dirac HNL at various masses as calculated by MadGraph. The BRs are identical for Majorana HNLs if the charge-conjugated decay modes are considered in addition.

	HNL mass			
Decay mode	1 GeV	5 GeV	10 GeV	20 GeV
$\mathrm{N} \longrightarrow \nu \mathrm{q}\bar{\mathrm{q}}$	28.5%	21.1%	18.3%	19.0%
$\mathrm{N} \longrightarrow \nu\bar{\nu}\nu$	12.3%	7.9%	6.1%	5.5%
$\mathrm{N} \longrightarrow \ell^{-}\ell^{+}\nu$	22.0%	22.6%	24.1%	24.2%
$\mathrm{N} \longrightarrow \ell^{-}\mathrm{q}\bar{\mathrm{q}}$	37.2%	48.4%	51.5%	51.4%
$\mathrm{N} \longrightarrow e^{-}\mathrm{u}\bar{\mathrm{d}}$	18.4%	12.1%	9.4%	8.5%
$\mathrm{N} \longrightarrow \mu^{-}\mathrm{u}\bar{\mathrm{d}}$	17.1%	12.1%	9.2%	8.4%
$\mathrm{N} \longrightarrow \tau^{-}\mathrm{u}\bar{\mathrm{d}}$	-	4.8%	7.4%	7.9%
$\mathrm{N} \longrightarrow e^{-}\mathrm{c}\bar{\mathrm{s}}$	-	7.5%	8.3%	8.2%
$\mathrm{N} \longrightarrow \mu^{-}\mathrm{c}\bar{\mathrm{s}}$	-	7.4%	8.2%	8.1%
$\mathrm{N} \longrightarrow \tau^{-}\mathrm{c}\bar{\mathrm{s}}$	-	2.1%	6.4%	7.7%

to different HNL mass and lifetime combinations, which are chosen to be close to this analysis's expected sensitivity reach. The HNLs are typically produced centrally, and their p_{T} distribution peaks at $m_{\mathrm{W}}/2$. Kinematic properties (p_{T}, η) of the prompt and displaced leptons $\ell_{1,2}$ are also shown for various sterile-active neutrino coupling scenarios. Compared with the case where the HNL couples exclusively to the first two lepton generations, the ℓ_1 becomes significantly softer (on average by a factor of three) when produced indirectly via leptonic tau decays for $V_\tau > 0$. The trigger efficiency is very low for such events, as the single lepton trigger threshold is around 30 GeV (see also Table 7.3), and hence a poor sensitivity to the pure tau coupling scenario is expected for this analysis. On the other hand, probing simultaneous coupling to all generations is still possible. For the ℓ_2 candidate, achieving the lowest possible experimental p_{T} threshold is desirable.



(a) (b)

(c) (d)

(e) (f)

(g) (h)

Figure 7.4: (a) HNL transverse momentum; (b) HNL pseudorapidity; (c) HNL boost; (d) HNL transverse displacement; (e) ℓ_1 transverse momentum; (f) ℓ_1 pseudorapidity; (g) ℓ_2 transverse momentum; (h) ℓ_2 pseudorapidity for various benchmark models and coupling scenarios. For the pure-τ coupling, ℓ_2 corresponds to the leptonic tau daughter.

7.3 Event selection and categorisation

In this Section, the strategy to select and categorise events is presented. The selection is primarily dictated by the trigger strategy, which is discussed first. Subsequently, the kinematic requirements on the physics objects (muons, electrons, jets, missing energy) used within this analysis are detailed. The signature of signal events is the presence of one prompt, isolated trigger lepton (ℓ_1) from an on-shell W boson decay, at least one displaced jet, and another displaced lepton (ℓ_2) in the final state.

Requirements on several event-level observables are introduced, which restrict the analysis phase space while remaining sensitive to HNL decays. This also reduces the major SM backgrounds, which are the following in this analysis:

W+jets results in the production of a prompt lepton from $W \longrightarrow \ell\nu$ and an additional non-prompt lepton from ISR or FSR heavy-flavour jet hadronisation.

Z/γ^*+jets results in two prompt leptons from $Z/\gamma^* \longrightarrow \ell\ell$ decay, where the two leptons are treated as $\ell_{1,2}$ candidates, and an additional ISR, FSR, or pileup jet being reconstructed. This background is predominant in the opposite-sign same-flavour (OSSF) categories. If one of the leptons is not reconstructed or identified with the wrong charge (charge flip, mostly relevant for electrons), the background can appear in other categories, similar to W+jets mentioned above.

V$\gamma^\star$+jets results in the production of a prompt photon which can be identified as a jet. Otherwise, the discussion is similar to Z/γ^*+jets, and this background also mainly contributes to OSSF categories.

QCD multijet can produce several leptons from heavy flavour jet decays, one of which appears isolated. Even though this is rare, the process has a very high production cross-section, making this a significant background.

Top background semileptonic or fully leptonic $\mathrm{t\bar{t}}$ decays can fake the signal process, resulting in a prompt lepton(s) in addition to several b jets, which can decay to produce non-prompt leptons in the final state.

A complete understanding of the data modelling in simulation is a crucial step in this analysis. Thus, a CR, where the expected signal contamination is negligible, is introduced to assess the modelling quality.

A categorisation scheme, based on the selected lepton flavour and sign combination, is presented, which allows targeting HNL mixing with different lepton generations and distinguishing between Majorana and Dirac HNLs. Events are also categorised to separately target scenarios where the HNL decay products are collimated (boosted) into a single jet or resolved into separate objects. An additional categorisation, based on the impact parameter significance of the displaced lepton candidate, ℓ_2, is introduced, which allows the search to be sensitive over a broad range of HNL lifetimes.

It is important to note that the presence of a displaced HNL decay vertex is not explicitly required in this analysis. Not requiring a displaced vertex makes the search more inclusive, as SV reconstruction can be inefficient for highly displaced HNLs. Instead, the displaced jet tagger, which for a given jet takes as input reconstructed SVs which share at least one track with the said jet, is leveraged.

7.3.1 Trigger strategy

The analysis makes use of pp collision data, collected at $\sqrt{s} = 13\,\text{TeV}$ using triggers requiring at least one isolated lepton (muon or electron), corresponding to integrated luminosities of $35.9\,\text{fb}^{-1}$, $41.5\,\text{fb}^{-1}$ and $59.7\,\text{fb}^{-1}$ for the three years (2016–2018) of data-taking, respectively. The p_T requirements of the single-lepton triggers are summarised in Table 7.3. Due to a finite lepton p_T resolution of a few GeV, "offline" p_T threshold requirements on the leptons used in the analysis are raised compared with the "online" trigger thresholds to ensure a trigger efficiency close to 100%.

A given event can, in general, be stored in separate data sets collected using both muon and electron triggers if there are several high-p_T, isolated leptons present in the event. A dedicated cleaning procedure, based on the flavour of the highest-p_T lepton is employed to ensure the data events are not used repeatedly.

Because an additional lepton (ℓ_2) from the HNL decay is reconstructed in the final state, dilepton triggers could, in principle, be used to lower the offline lepton p_T thresholds, which was exploited in a previous CMS search for prompt Majorana HNLs [49]. However, these dilepton triggers were found to be inefficient for non-prompt leptons and are therefore not used within this analysis.

Table 7.3: Transverse momentum thresholds for single, isolated lepton triggers used in the analysis for the three years of data taking (2016–2018).

Trigger	2016	2017	2018
Single muon	$p_T > 24\,\text{GeV}$	$p_T > 27\,\text{GeV}$	$p_T > 24\,\text{GeV}$
Single electron/photon	$p_T > 27\,\text{GeV}$	$p_T > 32\,\text{GeV}$	$p_T > 32\,\text{GeV}$

7.3.2 Physics objects

Primary vertex

A description of vertex reconstruction is given in Section 4.1.2. Events with at least one good PV are required. Good primary vertices must be reconstructed from at least five high purity tracks within $|d_z| < 24\,\text{cm}$ and $|d_{xy}| < 2\,\text{cm}$.

Muon candidates

In the analysis, two types of muon candidates are defined. For the prompt trigger lepton from a W boson decay (ℓ_1), tight quality criteria and a tight muon isolation requirement are applied. On the other hand, only loose muon identification requirements are employed to reconstruct events containing a displaced muon candidate (ℓ_2). Tight muon candidates need to satisfy $p_\mathrm{T} > 26, 29, 26\,\text{GeV}$ for 2016, 2017, and 2018 data, respectively, as well as $|\eta| < 2.4$. The p_T thresholds are chosen to be in the single-muon trigger plateau. Loose muons must have a transverse momentum of at least 3 GeV within $|\eta| < 2.4$. A more detailed description of these requirements is provided in Section 4.2.1. Lastly, tight muon candidates must be matched to the HLT muon that fired the trigger within $\Delta R < 0.1$.

Electron candidates

As in the muon case, tight and loose electron candidates are defined to target cases where either ℓ_1 or ℓ_2 is an electron. Electron candidates found within $\Delta R < 0.05$ of a PF muon are not selected, i.e. the muon is given priority. Tight electron candidates must satisfy the tight MVA-based identification working point and be within $|\eta| < 2.4$. The p_T thresholds are 29, 34, 34 GeV for 2016, 2017, and 2018 data, respectively. Finally, tight electron candidates in the barrel (endcap) region need to originate from the vicinity of the PV within $|d_{xy}| < 0.05\ (0.10)\,\text{cm}$ and $|d_z| < 0.10\ (0.20)\,\text{cm}$. Loose electrons must have a transverse momentum of at least 5 GeV within $|\eta| < 2.4$ and fulfil a modified version of the cut-based loose criteria. A more detailed description of these requirements is provided in Section 4.2.2. Finally, tight electron candidates must be matched to the HLT electron that fired the trigger within $\Delta R < 0.3$.

Lepton ordering

Only events containing exactly one tight (ℓ_1) and one loose (ℓ_2) lepton are retained. A requirement of $p_\mathrm{T}(\ell_1) > p_\mathrm{T}(\ell_2)$ is also applied to avoid double counting of events that fire both event triggers in separate data sets. The events are then categorised as $\ell_1\ell_2 \in \{\mu\mu, \mu e, ee, e\mu\}$. The p_T distributions of selected ℓ_1 and ℓ_2 candidates and the invariant mass of the $\ell_1\ell_2$ system are shown in Figure 7.5 for two benchmark categories corresponding to OS $\mu\mu$ and SS $e\mu$ final states. A significant contribution from the Z/γ^*+jets process can be observed for the OS $\mu\mu$ final state, not, however, present for the SS $e\mu$ final state. The distributions for other lepton charge and flavour contributions are not shown here for brevity, as they are qualitatively similar to the two considered benchmark categories.

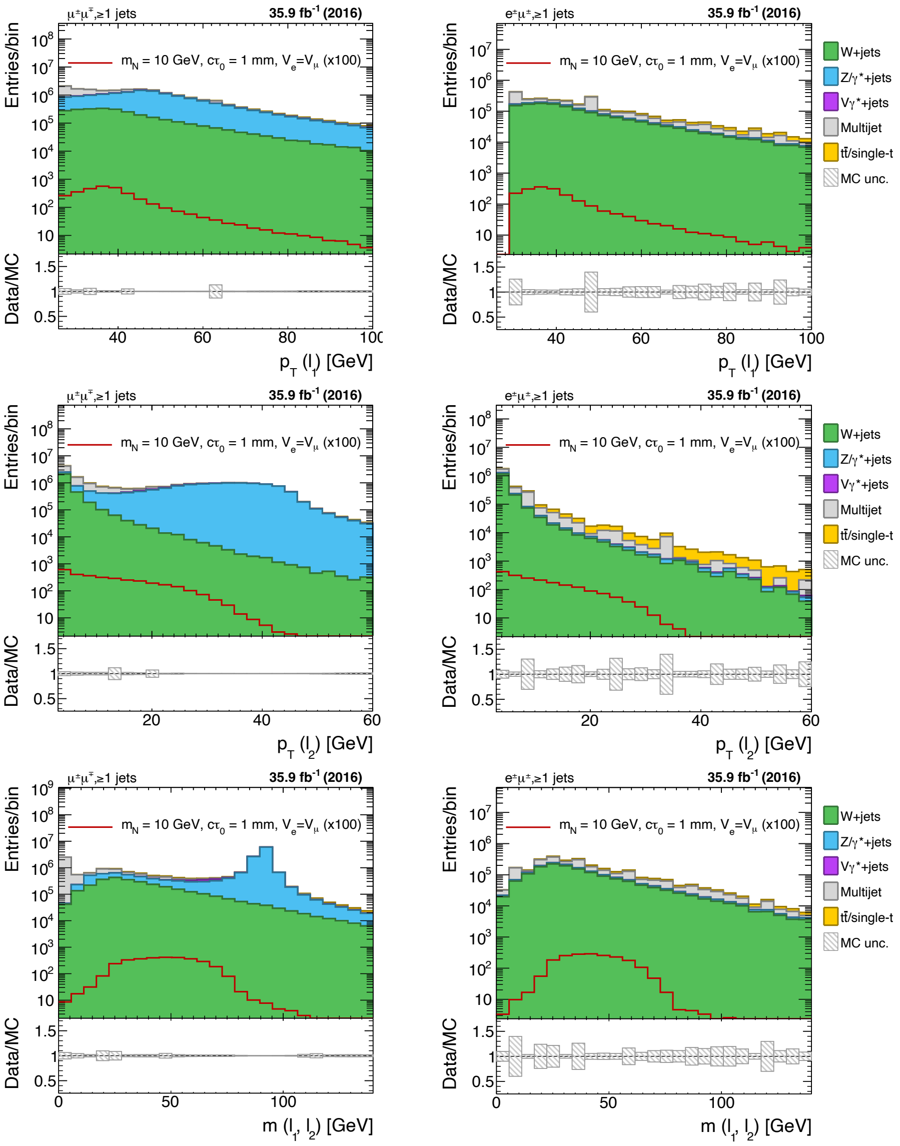
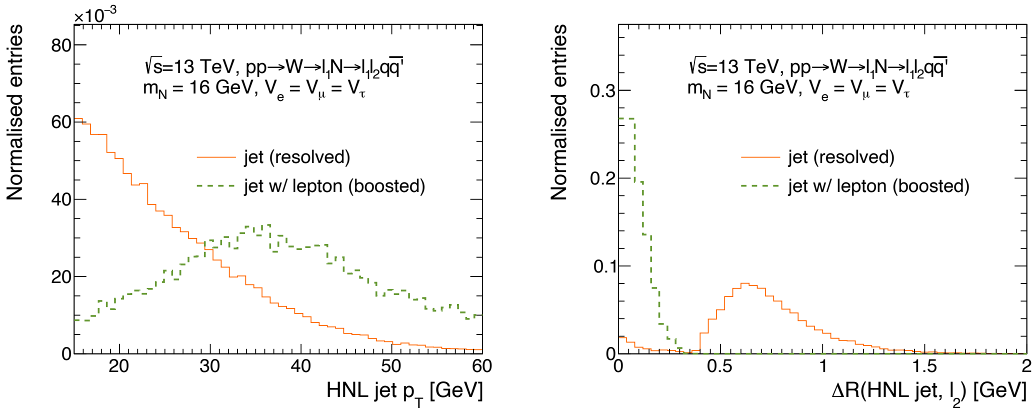


Figure 7.5: The distributions of $p_\mathrm{T}(\ell_1)$ (top row), $p_\mathrm{T}(\ell_2)$ (middle row), and $m(\ell_1, \ell_2)$ (bottom row), shown for benchmark OS $\mu\mu$ (left column) and SS $e\mu$ (right column) final states for the major simulated SM backgrounds after the preselection. A benchmark HNL scenario m_N = 10 GeV, $c\tau_0$ = 1 mm, $V_e = V_\mu$ is overlaid.

Jets

Jets are crucial objects in this analysis as the sensitivity to signal events primarily derives from applying the DNN jet tagger. An in-depth description of the jet clustering algorithm used in this analysis is provided in Section 4.2.3. Any jet overlapping within $\Delta R(\ell_1, \mathrm{jet}) < 0.4$ with the prompt lepton candidate (ℓ_1) is not selected, i.e. the priority is given to the lepton object. A distinguishing feature of this analysis is the treatment of candidate HNL jets, denoted henceforth as $\mathrm{j}^\star$. Depending on the HNL boost, such jets may contain leptons that correspond to a significant energy fraction of the jet. Such jets are explicitly considered in this analysis to leverage the jet tagger, whereas a standard approach taken in CMS is to cross-clean such jets with priority always given to the lepton object.



(a) (b)

Figure 7.6: The distributions of (a) p_{T}; (b) $\Delta R(\mathrm{j}^\star, \ell_2)$ of reconstructed displaced jets from HNL decays. The jets have the ℓ_2 from HNL decay boosted inside the jet or are resolved.

Figure 7.6 shows the p_{T} and $\Delta R(\ell_2, \mathrm{j}^\star)$ distribution for all reconstructed PF jets which possess the LLP label defined in Section 6.3.2. The jets are further split depending on whether they contain the lepton ℓ_2 from the HNL decay at the generator level. Depending on the HNL boost, the HNL decay products can be reconstructed as a single, boosted jet, for which p_{T} peaks at around 30 GeV. On the other hand, jets resolved from the ℓ_2 have a softer p_{T} spectrum. In this case, the jet is still found close to ℓ_2, i.e. the $\Delta R(l_2, \mathrm{j}^\star)$ peaks at ~ 0.6 for an HNL mass of 16 GeV. A small fraction of resolved jets are seen to populate the $\Delta R(\mathrm{j}^\star, \ell_2) < 0.4$ region. This can be explained by scenarios where a spurious lepton is misidentified as the ℓ_2 candidate instead of the lepton from the HNL decay or cases where labelling is imperfect due to a jet with a small ($< 10\%$) lepton energy fraction.

The analysis considers jets in the track acceptance $|\eta| < 2.4$ whose calibrated transverse momentum is greater than 20 GeV. Several additional criteria are applied to the jet if it contains the ℓ_2 candidate. The jet is rejected in case of overlap with an isolated lepton, specifically, $I_{\mathrm{rel.}}^{\ell_2} < 0.15$, or if the jet $p_{\mathrm{T}} < 30\,\mathrm{GeV}$. The primary motivation for the isolation requirement is not to consider jets that are primarily clustered from isolated leptons, as the hadronic activity in the jet likely originates from pileup; in addition, such jets were found not to be well-calibrated.

The jet that is the closest in ΔR to ℓ_2 is chosen as the HNL jet candidate, $\mathrm{j}^{\star}$. The efficiency of this selection is studied in Section 7.5.1. Figure 7.7 shows the overall number of selected jets, the $\Delta R(\ell_2, \mathrm{j}^{\star})$, and $p_{\mathrm{T}}(\mathrm{j}^{\star})$.

An alternative jet selection approach could have been to consider a larger jet clustering radius, e.g. $R = 0.8$, which would have allowed to simultaneously account for both boosted and resolved HNL decay scenarios. However, JECs for such jets are only supported by CMS for $p_{\mathrm{T}} > 150\,\mathrm{GeV}$, far greater than the typical HNL jet energy. In addition, the DNN jet tagger setup would require substantial modifications. The approach adopted within this thesis was to retrain the DNN jet tagger to target boosted and resolved scenarios separately by defining appropriate jet labels, which is further detailed in Section 7.5.

Missing transverse energy

Assuming $V_{\tau} = 0$, no genuine missing energy is expected in the final state, which can be exploited to reject background events. Even for the case of $V_{\tau} > 0$, the expected amount of missing energy from the neutrinos present in the final state is comparable to the typical missing-energy resolution (~ 30 GeV). The definition of missing energy used in this analysis is given in Section 4.2.4. The distribution of $p_{\mathrm{T}}^{\mathrm{miss}}$ is shown in Figure 7.8.

Event observables

Various event observables are defined based on the kinematics of the selected objects. These are used to define the SR and suppress remaining backgrounds by training an MVA event-level classifier, described in Section 7.4.

$m(\ell_1, \ell_2)$ The invariant mass of the dilepton system.

$\Delta\phi(\ell_1, \ell_2)$ The angular separation between ℓ_1 and ℓ_2.

$\Delta\phi(\ell_1, \vec{p}_{\mathrm{T}}^{\,\mathbf{miss}})$ The angular separation between ℓ_1 and $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$.

$m_{\mathrm{T}}(\ell_1, \vec{p}_{\mathrm{T}}^{\,\mathbf{miss}})$ The transverse mass of the ℓ_1 and $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$, defined as:

$$m_{\mathrm{T}}(\ell_1, \vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}) = \sqrt{2 \times p_{\mathrm{T}}(\ell_1) \times \vec{p}_{\mathrm{T}}^{\,\mathrm{miss}} \times (1 - \cos(\Delta\phi(\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}, \ell_1))}. \tag{7.7}$$

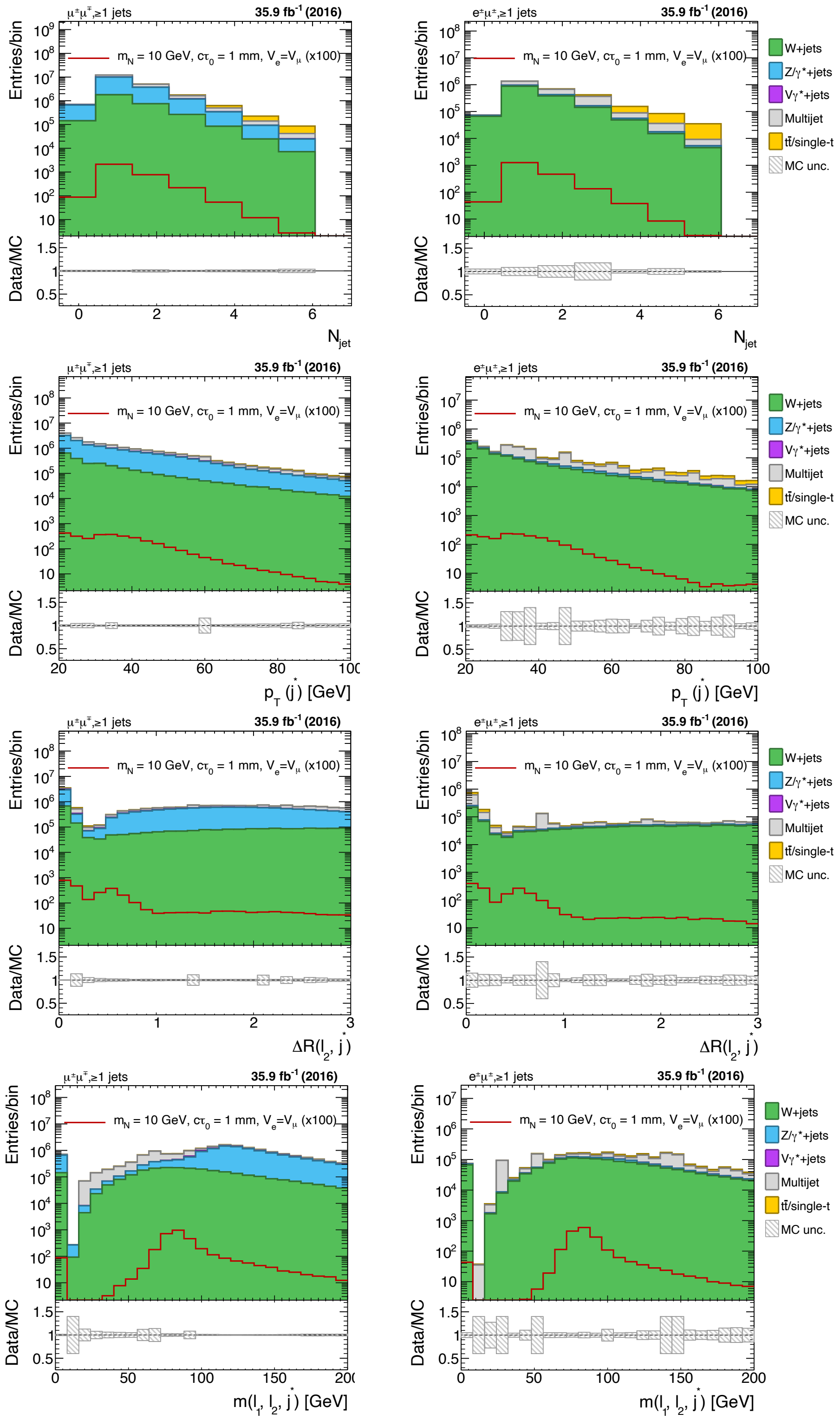


Figure 7.7: The distributions of the number of selected jets (top row), $p_{\mathrm{T}}(\mathrm{j}^{\star})$ (second row), the angular separation $\Delta R(\ell_2, \mathrm{j}^{\star})$ (third row), and $m(\ell_1, \ell_2, \mathrm{j}^{\star})$ (bottom row) shown for benchmark OS $\mu\mu$ (left column) and SS $e\mu$ (right column) categories for the major simulated SM backgrounds after the preselection. A benchmark HNL scenario $m_{\mathrm{N}} = 10\,\mathrm{GeV}, c\tau_0 = 1\,\mathrm{mm}, V_e = V_\mu$ is overlaid.

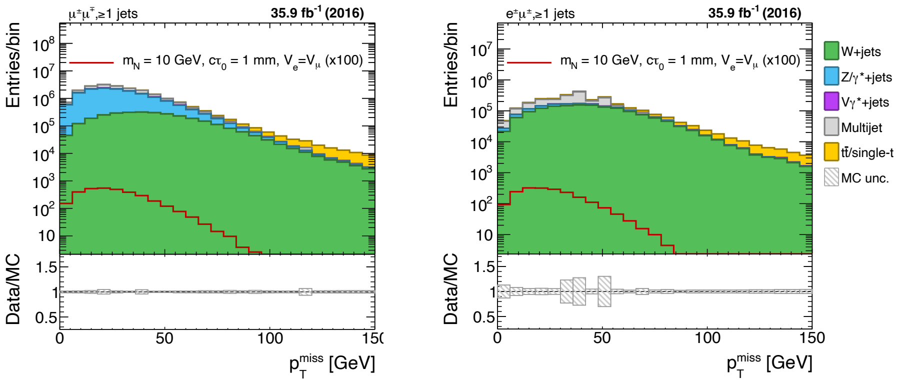


Figure 7.8: The distributions of $p_{\mathrm{T}}^{\mathrm{miss}}$ shown for benchmark OS $\mu\mu$ (left column) and SS $e\mu$ (right column) final states for the major simulated SM backgrounds. A benchmark HNL scenario $m_{\mathrm{N}} = 10\,\mathrm{GeV}, c\tau_0 = 1\,\mathrm{mm}, V_e = V_\mu$ is overlaid.

$m(\ell_1, \ell_2, \mathrm{j}^\star)$ The invariant mass of the dilepton and $\mathrm{j}^\star$ candidate system, defined as:

$$m(\ell_1, \ell_2, \mathrm{j}^\star) = \begin{cases} \sqrt{(p(\ell_1) + p(\mathrm{j}^\star))^2} & \text{if } \Delta R(\ell_2, \mathrm{j}^\star) \leq 0.4, \\ \sqrt{(p(\ell_1) + p(\ell_2) + p(\mathrm{j}^\star))^2} & \text{if } \Delta R(\ell_2, \mathrm{j}^\star) > 0.4. \end{cases} \tag{7.8}$$

where p denotes the four-momentum of a particle. This definition ensures that there is no double-counting of the ℓ_2 energy in the case of overlap with $\mathrm{j}^\star$.

Event sphericity, aplanarity, circularity Several “event shape” variables are calculated based on the eigenvalues of the sphericity tensor, defined as [158]:

$$S^{\alpha\beta} = \frac{\sum_i p_i^\alpha \cdot p_i^\beta}{\sum_i \left|\mathbf{p}_i^2\right|}, \tag{7.9}$$

where $\alpha, \beta = x, y, z$, and $\mathbf{p}_i$ is the momentum of a physics object i obtained from a collection of all selected jets and leptons in the event. The eigenvalues of the tensor express geometrical correlations between selected objects in an event. For a perfect QCD dijet event, all eigenvalues should be zero, while for three-jet events, one of the eigenvalues should be zero. For events with a spherical distribution, all eigenvalues should be equal. From the eigenvalues of this matrix, the event shape variables are calculated as follows:

$$S = \frac{3}{2}(\lambda_2 + \lambda_3), \tag{7.10}$$

$$A = \frac{3}{2}(\lambda_3), \tag{7.11}$$

$$C = 3(\lambda_0\lambda_1 + \lambda_0\lambda_2 + \lambda_1\lambda_2). \tag{7.12}$$

7.3.3 Preselection

The preselection is a set of fiducial requirements that restrict the phase space to possible HNL final states with as high as possible signal efficiency and substantially reduce the SM backgrounds. The preselection is constructed to retain sufficient flexibility to be used for defining the SR and the CR. Events containing exactly one ℓ_1 and ℓ_2 candidate are selected. Only events having at least one but no more than five selected jets are retained. Events must pass the missing energy filters defined in Section 4.2.4. This requirement limits hadronic activity to suppress $\mathrm{t\bar{t}}$ events while having a negligible effect on signal efficiency. Finally, a requirement of $p_{\mathrm{T}}^{\mathrm{miss}} < 100\,\mathrm{GeV}$ is imposed, which reduces events originating from the W+jets process.

The SR is defined by a requirement of $\Delta R(\ell_2, \mathrm{j}^{\star}) < 1.3$ as the HNL products are expected to be collimated. In addition, $m(\ell_1, \ell_2) < 80\,\mathrm{GeV}$ is required to suppress events originating from the Z/γ^{*}+jets process, and $m(\ell_1, \ell_2) > 20\,\mathrm{GeV}$ to suppress QCD multijet events. By inverting the $m(\ell_1, \ell_2)$ requirements, the CR is defined. An overview of each region's various selection requirements is summarised in Table 7.4. Finally, a requirement imposing the signal CC production hypothesis, $70 < m(\ell_1, \ell_2, \mathrm{j}^{\star}) < 90\,\mathrm{GeV}$, is applied in the SR. The efficiency of these requirements for signal and background processes are summarised in Table 7.5.

It is worth exploring the scenario of a tau particle being produced. Both ℓ_1 and ℓ_2 can originate from a leptonically-decaying tau particle; however, in this case, the resulting p_{T} of the final state leptons will be a factor of three smaller on average, as some of the energy will be carried away by neutrinos, as shown in Figure 7.4. For ℓ_1, this results in severely reduced trigger efficiency. However, the analysis is still sensitive to scenarios where the ℓ_2 originates from a leptonic tau decay. Finally, in case either tau particle decays hadronically, the event will not be selected, as only a single lepton will be present in the final state.

7.3.4 Event categorisation

In this Section, the categorisation of SR events is described. Only events that pass the preselection requirements, as summarised in Section 7.3.3, are considered. The events are categorised to enhance sensitivity to different HNL mass, $c\tau_0$, and mixing scenarios. Ultimately, the categorisation results in $4 \times 2 \times 3 \times 2 = 48$ independent SR categories:

- lepton flavour, i.e. $\mu\mu$, μe, $e\mu$, ee allows to probe different HNL-active neutrino mixing scenarios;

Table 7.4: Overview of selection requirements in SR and CR for the dilepton categories. The prompt lepton requirements are shown per year (2016, 2017, 2018) of data-taking.

	SR	CR
Lepton selection		
Tight, isolated lepton (ℓ_1), $\lvert\eta\rvert < 2.4$ Muon: $p_T > 26,\ 29,\ 26\,\text{GeV}$ Electron: $p_T > 29,\ 34,\ 34\,\text{GeV}$	✓	✓
Loose lepton (ℓ_2), $\lvert\eta\rvert < 2.4$ Muon: $p_T > 3\,\text{GeV}$ Electron: $p_T > 5\,\text{GeV}$	✓	✓
$p_T(\ell_1) > p_T(\ell_2)$ Third lepton veto	✓	✓
Jet selection ($p_T > 20\,\text{GeV}, \lvert\eta\rvert < 2.4$)		
Jets w/leptons ($p_T > 30\,\text{GeV}, I^{\ell}_{\text{rel}} > 0.15$) $N_{\text{jet}} < 5$	✓	✓
Event observables		
$p_T^{\text{miss}} < 100\,\text{GeV}$	✓	
p_T^{miss} filter	✓	✓
$\min \Delta R(\ell_2, \text{jets}) < 1.3$	✓	
$m(\ell\ell) > 20\,\text{GeV}$	✓	✓
$m(\ell\ell) < 80\,\text{GeV}$	✓	inverted
$70\,\text{GeV} < m(\ell_1, \ell_2, \text{j}^{\star}) < 90\,\text{GeV}$	✓	

Table 7.5: Cumulative efficiency of various object- and event-level preselection requirements shown for a benchmark signal model of $m_N = 10\,\text{GeV}, c\tau_0 = 1\,\text{mm}, V_e = V_\mu = V_\tau$ and the major SM background (W+jets, Z/γ^*+jets) processes.

	signal	W+jets	Z/γ^*+jets
Prompt lepton (ℓ_1)	18.65%	21.86%	29.24%
Trigger requirement	15.26%	18.12%	25.04%
One loose lepton (ℓ_2)	5.73%	1.29%	17.83%
Third lepton veto	5.21%	1.22%	16.16%
At least one jet	3.77%	0.68%	9.17%
$N_{\text{jets}} < 5$	3.64%	0.62%	8.80%
$p_T^{\text{miss}} < 100\,\text{GeV}$, p_T^{miss} filter	3.64%	0.61%	8.77%
$\min \Delta R(\ell_2, \text{jets}) < 1.3$	2.65%	0.17%	2.81%
$20 < m(\ell\ell) < 80\,\text{GeV}$	2.59%	0.13%	0.79%
$70 < m(\ell_1, \ell_2, \text{j}^{\star}) < 90\,\text{GeV}$	1.87%	0.03%	0.18%

- dilepton system charge allows to distinguish between Dirac and Majorana HNLs;
- ℓ_2 transverse impact parameter significance, $d_{xy}^{\mathrm{sig}} = d_{xy}/\sigma(d_{xy})$, allows separate optimisation for prompt and displaced HNL scenarios;
- $\Delta R(\ell_2, \mathrm{j}^{\star}) < 0.4$ is used to identify boosted topologies. Events with $\Delta R(\ell_2, \mathrm{j}^{\star}) > 1.3$, on the other hand, are considered resolved. The resolved category is populated more with increasing HNL mass.

Displacement categorisation

Several variables were initially considered to separate HNL events based on their displacement, including d_{xy}, d_{xy}^{sig}, and the profiled tagger parameter $\hat{L}_{xy}$ (see Section 7.5.5). Both d_{xy} and $\hat{L}_{xy}$ were found to be mismodelled in simulation. The advantage of using d_{xy}^{sig} is that potential bias in both the impact parameter measurement and its uncertainty cancel out when taking the ratio, and as a result this variable was found to be modelled better than the other two candidates. Figure 7.9 shows the d_{xy}^{sig} distribution for background and two benchmark signal processes of different $c\tau_0$. The events are divided into the following three subcategories:

Prompt $d_{xy}^{\mathrm{sig}} < 1$, corresponding primarily to prompt leptons.

Intermediate $1 < d_{xy}^{\mathrm{sig}} < 10$, to target heavy flavour-like signatures.

Displaced $d_{xy}^{\mathrm{sig}} > 10$, corresponding to significantly displaced leptons.

As can be seen from Figure 7.9, the contribution from the low-mass HNL model is primarily in the boosted topology due to the significant HNL boost, while the higher-mass HNL model populates both the boosted and the resolved event topologies. The categorisation allows for efficient separation of background processes depending on the ℓ_2 origin (prompt, from heavy-flavour decay, mismeasured) and separation between signal models of different $c\tau_0$.

7.3.5 Corrections

Several corrections are applied to physics objects and events to improve the agreement between the data and simulation and are summarised below. The corrections are also considered potential sources of systematic uncertainty, further described in Section 8.2.

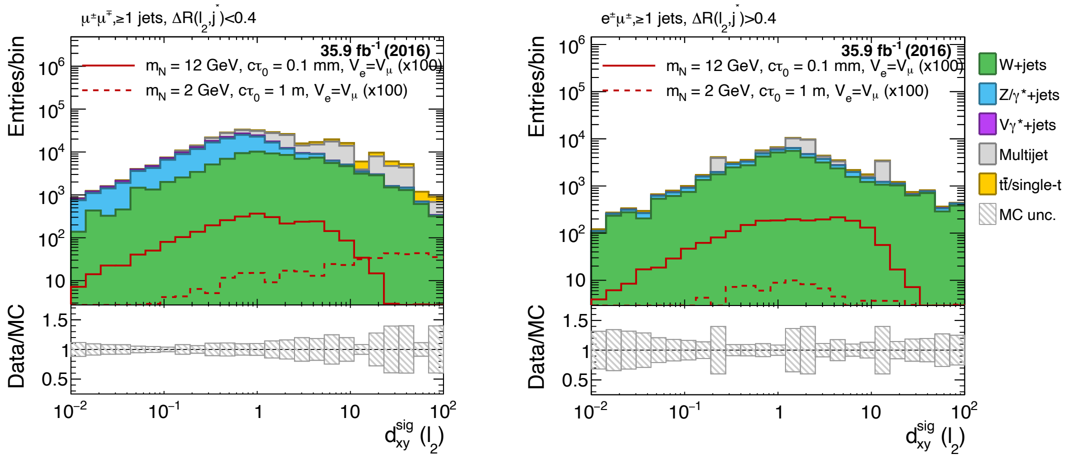


Figure 7.9: The distributions of d_{xy}^{sig} shown for benchmark OS $\mu\mu$ boosted (left) and SS $e\mu$ resolved (right) categories for the major simulated SM backgrounds in the SR. Two benchmark HNL scenarios, $m_{\text{N}} = 10\,\text{GeV}, c\tau_0 = 1\,\text{mm}$, and $m_{\text{N}} = 2\,\text{GeV}, c\tau_0 = 1\,\text{m}$, are overlaid.

Jet energy scale (JES) The momenta of reconstructed jets in both data and simulation are calibrated by applying offset and response corrections. Residual corrections are additionally applied to data only. The corrections are also propagated to $p_{\text{T}}^{\text{miss}}$.

JER The jet energy is modified in simulation such that its resolution matches the one in the data. For a jet that can be matched to a corresponding particle level jet, the deviation between the two jet energies is either decreased or increased using data-to-simulation scale factors.

Pileup Simulated events are reweighted to obtain a pileup profile matching the actual distribution in data. The corresponding distribution in data is derived using the instantaneous luminosity for each year of data taking and a minimum bias cross-section of 69 mb.

Muons The efficiency of a prompt muon candidate to pass trigger, identification, and isolation requirements is corrected by applying data-to-simulation scale factors.

Electrons Prompt electron reconstruction and identification data-to-simulation scale factors are applied.

7.4 Event-level boosted decision tree

In order to ensure an optimal event selection, a BDT was trained to discriminate between HNL and background events. The BDT is used as an additional event-level discriminant orthogonal to the jet-level DNN tagger. In this Section, the input features, training setup and performance of the BDT are detailed.

7.4.1 Input features

The BDT was parameterised in terms of the dilepton system charge and the lepton flavours to obtain a single model for all flavour and sign combinations. The input features were carefully chosen to ensure a minimal correlation with the tagger score, as the two variables are used in a data-driven background estimation method detailed in Section 8. In addition, good modelling of all input features was ensured by inspecting the data/MC agreement in the CR. The selected input features are listed in Table 7.6. Additional studies of the modelling of the BDT input features are shown in Appendix A.

Table 7.6: BDT input features.

feature
$p_T^{\ell_1}, \lvert\eta^{\ell_1}\rvert$
p_T^{miss}
$\Delta\phi(\ell_1, \vec{p}_T^{\text{miss}})$
$m_T(\ell_1, \vec{p}_T^{\text{miss}})$
Event shape variable A
ℓ_1, ℓ_2 flavour, $q(\ell_1) \times q(\ell_2)$

7.4.2 Training setup

Only simulated events passing the SR preselection requirements detailed in Table 7.4 were used for the training. A separate model was trained for each year of data taking. Independent signal HNL events were used in training, with a 50–50% split between the signal and background events. Backgrounds events from W+jets, Z/γ^*+jets, and $V\gamma^\star$+jets processes were subsampled according to their excepted yield in the SR. On the other hand, QCD multijet events were not included in the training due to the low event yield of the simulated samples.

The `XGBoost` [131] library was used with the categorical cross-entropy loss function, a learning rate of 0.05 and up to 1000 estimators (weak learners). The data set was separated into training (80%) and validation samples (20%). The BDT model loss was monitored, and early stopping was applied if the validation sample loss ceased decreasing for ten epochs to avoid overtraining.

7.4.3 Performance

The BDT performance was evaluated on background events entering the SR. A ROC curve is shown in Figure 7.10 for three different HNL benchmark scenarios. To benchmark the BDT performance, a logistic regression model was trained on the same input features as the BDT, yet was significantly outperformed by the BDT. The step-like behaviour results from the limited QCD multijet events with high average event weights omitted from the training but included in the BDT validation. For a threshold requirement of 0.4 on the BDT score, the backgrounds are reduced by more than 50%, while the signal efficiency remains around 80%. A working point with a relatively high signal efficiency is chosen as the analysis sensitivity is mostly driven by the application of the DNN jet tagger.

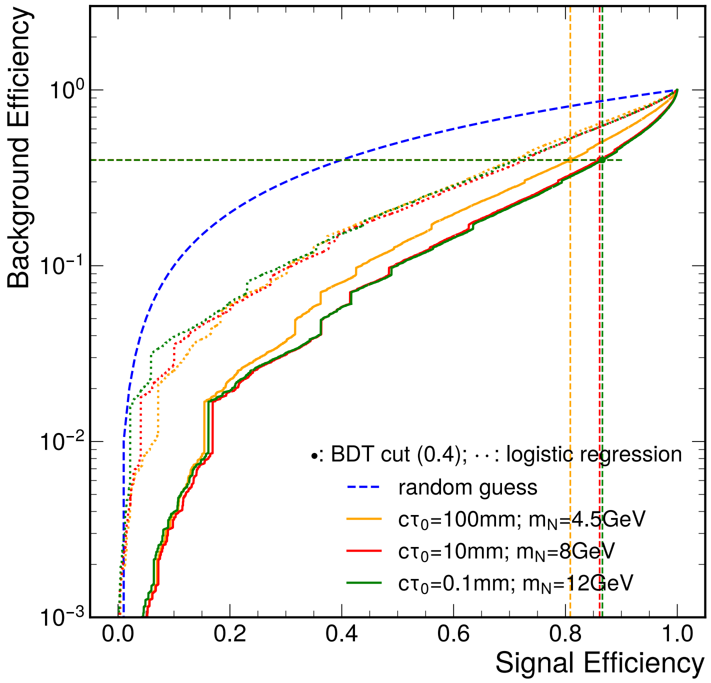


Figure 7.10: ROC curve showing the BDT performance for the 2016 scenario. The signal efficiency for three benchmark HNL scenarios, indicated in the legend, is compared against the expected SM backgrounds in the SR. The efficiency corresponding to a working point threshold of 0.4 is also indicated by dashed lines. Additional reference ROC curves obtained by training a logistic regression discriminant are also shown as dotted lines.

7.4.4 Validation

The modelling of the BDT discriminant in the CR (Figure 7.11) is validated for boosted and resolved event topologies in OS $\mu\mu$ and SS $e\mu$ final states. Overall, a good agreement is observed between data and simulation for the OS $\mu\mu$ final state. However, the modelling for SS $e\mu$ is limited due to low event yields of simulated QCD multijet events. The mismodelling highlights a need to use a data-driven technique to estimate the backgrounds.

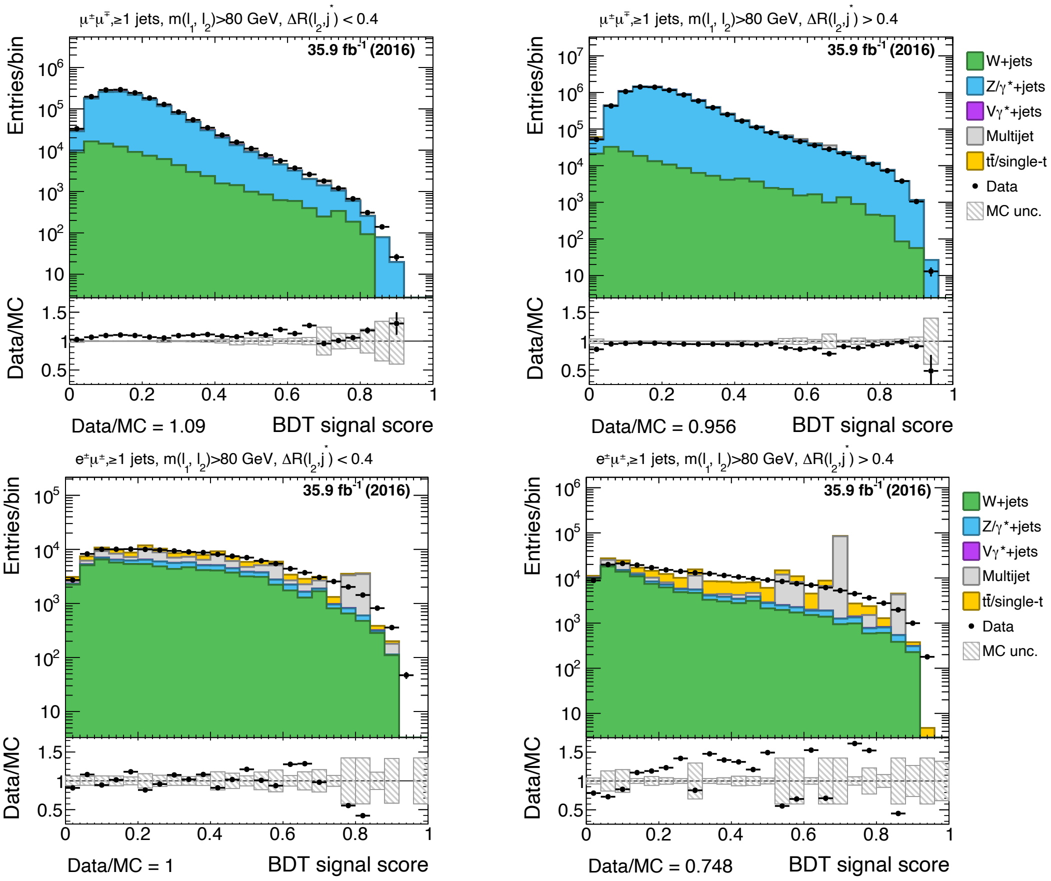


Figure 7.11: Data/MC comparison of the BDT signal class probability for OS $\mu\mu$ (top) and $e\mu$ (bottom) categories, broken down by boosted (left) and resolved (right) topology, shown for simulation and CR data collected in 2016.

7.5 Displaced jet tagger

The displaced jet tagger was extended to improve performance when targeting semileptonic HNL decays. In this Section, several of these new features are discussed. First, a categorisation of displaced jets depending on their flavour is introduced. Extensions to the tagger architecture, which primarily involve including lepton input features, are detailed. A change in the tagger parameterisation procedure from lifetime to displacement is described. Subsequently, the displacement profiling approach, picking the optimal input parameter for a given jet, is introduced. Finally, the tagger performance to discriminate jets from HNL decay from background jets, and tagger modelling in the CR is studied.

7.5.1 Displaced jet subclasses

The inclusive displaced jet definition, introduced in Section 6.3.2, was extended to accommodate different flavours of displaced jets. Jets meeting the generic displaced jet definition were subcategorised to target different topologies of HNL decays ($N \longrightarrow \ell_2 q\bar{q}$). The labelling was explicitly designed with HNL decays in mind, although it generalises to any LLP decay. A crucial feature of low-mass HNL is that they are significantly boosted, and their decay products can be collimated into a single jet or reconstructed as several resolved objects. Based on this, the following exclusive displaced jet categories were adopted:

Quark-lepton jets contain both the lepton and the quark from the LLP decay at the generator level. These jets are labelled as `LLP_QMU` and `LLP_QE`.

Quark jets contain a parton-level quark from the LLP decay but no leptons or jets originating from the FSR of the displaced quark. The latter case is important if the LLP is a massive gluino, but it is negligible for the soft partons from the HNL decay. These jets are labelled as `LLP_Q`.

Lepton jets contain the lepton but no quarks from the LLP decay and are labelled as `LLP_MU` or `LLP_E`.

Tau jets represent a subclass of hadronic tau decays, which is labelled as `LLP_TAU`. Leptonic tau decays, on the other hand, populate the previously mentioned subclasses.

Heavy-flavour displaced jets are not considered in the training since these are heavily suppressed kinematically and through the small off-diagonal CKM MEs, compared with light-flavour displaced jets. The following signal jet classes were used in training: `LLP_Q`, `LLP_QMU`, `LLP_QE`, `LLP_MU`, `LLP_E`, `LLP_TAU`. Studies of tagger performance later revealed that the tagger struggled to differentiate between displaced quark jets, `LLP_Q`, and displaced hadronic tau decays, `LLP_TAU`. Identifying displaced hadronic tau decays is a challenging task worth further study; in this thesis, the class probabilities for `LLP_Q` and `LLP_TAU` classes were subsequently merged to recover any possible sensitivity loss.

The fractions of different jet classes for two different HNL masses are shown in Figure 7.12. The `LLP_MU` or `LLP_E` overall comprised a negligible fraction and hence are not shown. With increasing HNL mass and decreasing boost, the HNL decay products become more resolved and hence the fraction of resolved quark jets increases. A purity matrix, shown in Figure 7.13, quantifies how the truth level HNL jets populate the reconstruction-level boosted and resolved categories. Overall, the matrix is highly diagonal. There is inefficiency of ~ 15% for correctly identifying `LLP_Q` as also seen in Figure 7.6(b).

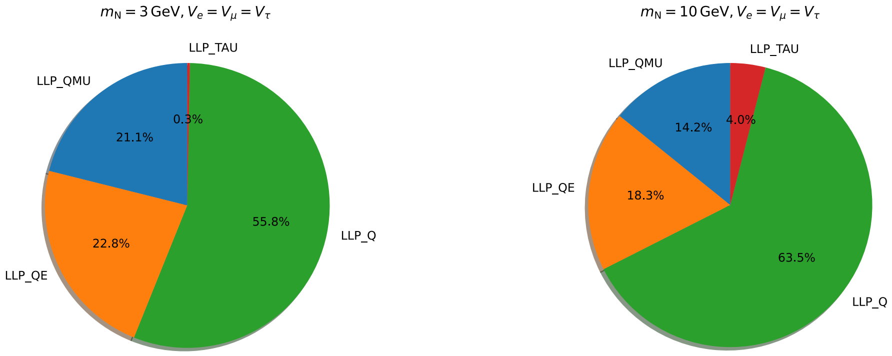


Figure 7.12: Displaced jet subclass fractions for 3 GeV (left) 10 GeV (right) HNL models, assuming equal couplings to all three active neutrino generations.

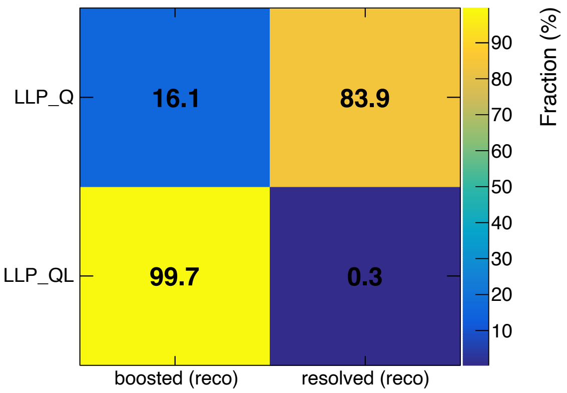


Figure 7.13: Probability matrix of reconstructing a truth-level quark-lepton or quark HNL jet in the boosted or resolved categories.

7.5.2 Background jet classes

Initial tests in training the extended DNN tagger revealed that the performance for displaced quark-lepton jet classes (`LLP_QMU` and `LLP_QE`) was inadequate when discriminating against specific background jets containing a prompt lepton (photon). Such jets can occur accidentally when a prompt lepton (photon) is clustered inside a jet containing some spurious soft hadronic activity. Hence, it was deemed necessary also to introduce a corresponding background jet class for training the DNN. Prompt lepton (photon) jets are defined if the majority of the jet momentum can be associated with a corresponding prompt lepton (photon) at the generator level, and if they do not originate from the hadronisation of a quark, a gluon, or from the subsequent decay of a hadron. Furthermore, pileup jets were also included in the training, as they become an important background source for $p_T < 50$ GeV. Overall, the following background jet classes were used: uds, g, c, b, μ, e, τ, γ, pileup.

The displaced jets of classes `LLP_MU`, `LLP_E`, just like the prompt jet classes μ, e, τ, γ, can only be called jets in the strict sense of the word that they are clustered into PF jets by the anti-k_T algorithm. Typically, such jets, which have a high lepton energy fraction (>50%), and for which JECs are not applicable, would not be selected in a standard analysis based on overlap with an isolated lepton. In this case, only the prompt, isolated lepton constituent would be retained. For completeness, these classes were still included in the training.

7.5.3 Tagger architecture

The DNN tagger architecture was extended by including additional PF muon and electron groups, as shown in Figure 7.14, to help with classifying jets with leptons. Several new global jet features were introduced compared with the ones listed in Section 6.4. The tagger was trained with CR events using DA similarly as described in Section 6. The complete list of extended input features can be found in Appendix B.

7.5.4 Displacement parameterisation

In the first version of the tagger, the DNN was parameterised using the proper lifetime, $c\tau_0$. In the extended version, we opted to parameterise the tagger using the transverse displacement of the HNL, L_{xy}, instead. This is a more natural approach, as it directly connects LLP and jet constituent displacement. During the training, the displacement is generated at random for background jet classes by sampling from the displaced jet

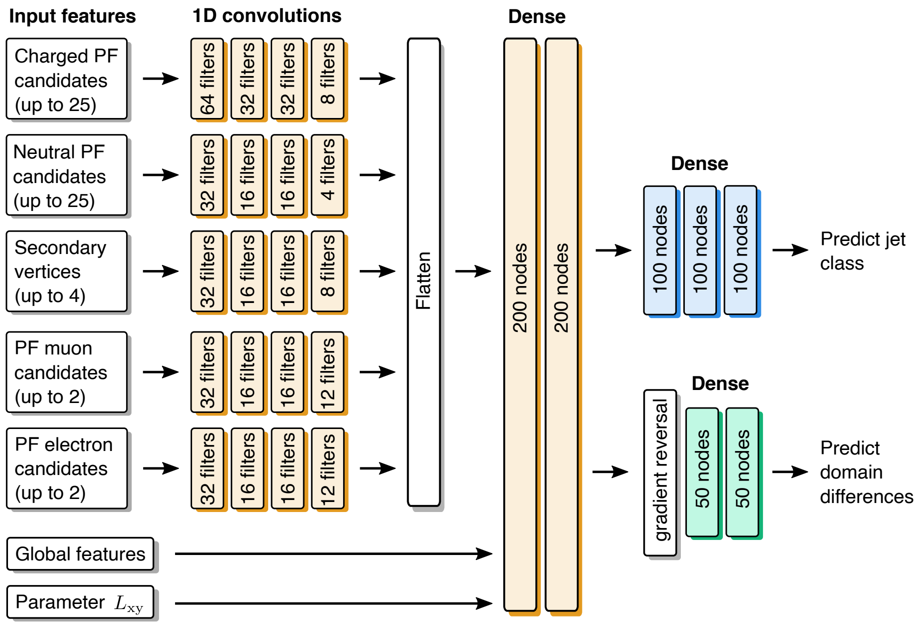

Input features
1D convolutions
Dense
Charged PF candidates (up to 25)
64 filters
32 filters
32 filters
8 filters
Neutral PF candidates (up to 25)
32 filters
16 filters
16 filters
4 filters
Secondary vertices (up to 4)
32 filters
16 filters
16 filters
8 filters
PF muon candidates (up to 2)
32 filters
16 filters
16 filters
12 filters
PF electron candidates (up to 2)
32 filters
16 filters
16 filters
12 filters
Global features
Parameter L_{xy}
Flatten
200 nodes
200 nodes
Dense
100 nodes
100 nodes
100 nodes
Predict jet class
Dense
gradient reversal
50 nodes
50 nodes
Predict domain differences

Figure 7.14: The architecture of the NN for identifying displaced jets from the HNL decay. Compared with Figure 6.5, additional PF muon and electron input features have been introduced. In addition, the tagger is parameterised as a function of the jet displaced instead of the LLP $c\tau_0$.

distribution, as was previously done for $c\tau_0$ as described in Section 6.4.3. One drawback of the new parameterisation approach is that it is a priori unclear what displacement value L_{xy} to use when evaluating the tagger for a given signal model.

7.5.5 Tagger profiling

As shown in Figure 6.13, the tagger has to be evaluated by specifying the “right” model parameter to obtain the best sensitivity. The first approach tried in this analysis was, for a given HNL model and jet flavour, to marginalise the displacement distribution for a given HNL scenario:

$$P(\mathrm{c}|m_N, c\tau_0) = \int_0^\infty P(L_{xy}|m_N, c\tau_0)P(\mathrm{c}|L_{xy})\mathrm{d}L_{xy}, \, \mathrm{c} \in \{\mathrm{q}, \mathrm{q}\mu, \mathrm{q}e\}. \tag{7.13}$$

The prior displacement distribution, $P(L_{xy}|m_N, c\tau_0)$, is exponential. For integrals of the form $\int_0^{+\infty} e^{-x} f(x) dx$, the Gauss-Laguerre [159] quadrature can be used to approximate the definite integral to the desired accuracy. This approach would result in optimal sensitivity; however, it was computationally infeasible due to the CPU-intensive nature of running the tagger, compounded by the wide range of mass and coupling scenarios.

Ultimately, the displacement parameter L_{xy} was chosen to be profiled in this analysis. For a given jet, the parameter is chosen within the interval $10^{-1} < L_{xy} < 10^{3}$ mm to obtain the highest likelihood (discriminant score) for a given single jet class:

$$P_{\text{single}}(\text{c}) = \max_{L_{xy}} P(\text{c}|L_{xy}). \tag{7.14}$$

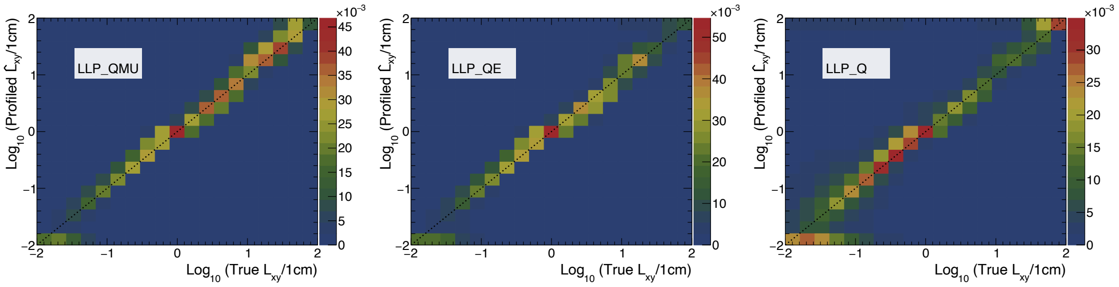
It is important to note that the resulting likelihoods are not probabilities, i.e.:

$$\sum_{\text{c}}^{\text{SM+LLP}} P_{\text{single}}(\text{c}) \neq 1. \tag{7.15}$$

As the tagger is used to identify displaced jets, a likelihood ratio for each displaced jet subclass was used instead:

$$P_{\text{c}} = \max_{L_{xy}} \left[\frac{P(\text{c}|L_{xy})}{\sum_{j}^{\text{SM}} P(\text{j}|L_{xy})} \right]. \tag{7.16}$$

The profiled parameter corresponding to the maximum likelihood ratio is denoted as $\hat{L}_{xy}$. The likelihood ratio increases the performance to discriminate signal versus background jets at the expense of losing discrimination between displaced jet subclasses. The likelihood ratio discriminant was transformed $[0; \infty] \longrightarrow [0;\ 1]$ for simplicity and easier visualisation. The chosen mapping was monotonic and did not modify the performance.



(a) quark-muon (b) quark-electron (c) quark

Figure 7.15: True vs profiled displacement for signal (a) quark-muon; (b) quark-electron; (c) quark jet classes.

To verify the performance of the profiling approach, the dependence of the profiled $\hat{L}_{xy}$ versus the true L_{xy} for reconstructed signal jets was studied, and the results are shown in Figure 7.15. A strong correlation between the two quantities is observed, thus validating the method. Some differences between jet classes are observed due to different reconstruction efficiencies as a function of L_{xy}.

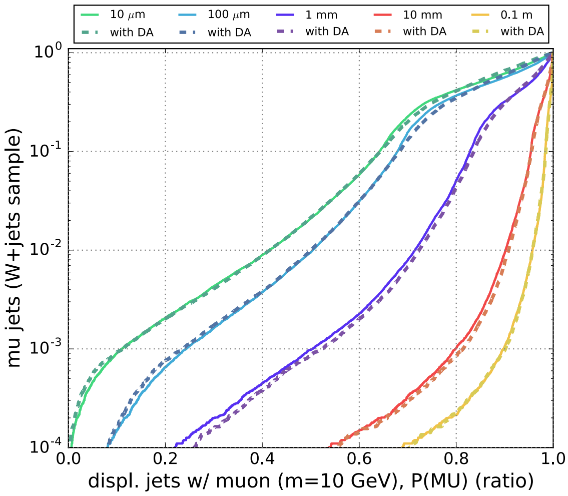
In the analysis, the use of a specific discriminant depends on the lepton flavour and event topology. For resolved events, the P_{q} score is used. For events classified as boosted, the $P_{\mathrm{q}\mu}$ or $P_{\mathrm{q}e}$ scores are used instead, depending on the flavour of the ℓ_2 candidate.

7.5.6 Performance and validation

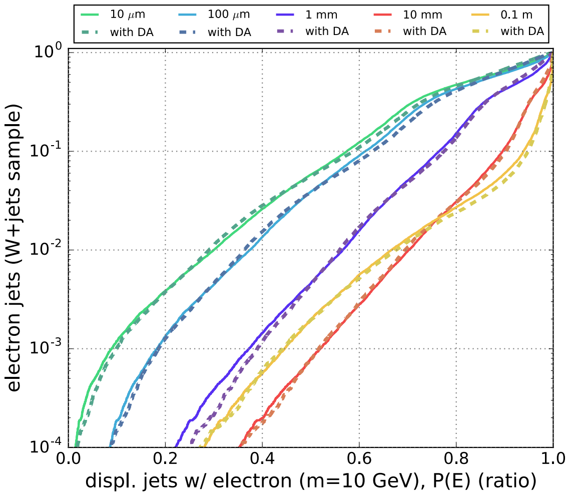
ROC curves were obtained using profiled likelihood-ratio discriminants $P_{\mathrm{q}\mu}$, $P_{\mathrm{q}e}$, and P_{q} for the corresponding combination of displaced jet classes (`LLP_QMU`, `LLP_QE`, `LLP_Q`), taken from a sample of HNL production with $m_{\mathrm{N}} = 10\,\mathrm{GeV}$. For each of the three cases, the equivalent SM jet class (μ, e, udsg) is considered, taken from an inclusive sample of W+jets production, as shown in Figure 7.16. As expected, the discrimination power increases with increasing jet displacement for all classes. Displaced muon-quark jets `LLP_QMU` exhibit a further increase at high lifetimes (~ 100 mm) due to information from the muon systems that helps with identifying very displaced tracks. For lifetimes of 10–100 mm, an ε(LLP) of about 20–30% for displaced quark jets is found. An even higher ε(LLP) is observed for displaced muon (electron)-quark jets, of about 80–90% (40–50%). The tagger modelling in the CR (Figure 7.17) is validated for resolved and boosted jets in OS $\mu\mu$ and SS $e\mu$ final states. Overall, a good agreement is observed between data and simulation for the discriminants used in the analysis. Additional studies of the modelling of the DNN jet tagger score are shown in Appendix A. Unlike the initial split SUSY scenario, using DA does not significantly degrade the performance for HNL jets or increase the data/MC agreement in the CR. This suggests that the features of soft $\mathcal{O}(10\,\mathrm{GeV})$ jets might be better described in simulation than those of energetic $\mathcal{O}(100\,\mathrm{GeV})$ jets. Another possible factor that could explain the difference is the inclusion of leptons with relatively well modelled properties inside jets for the HNL scenario.

7.6 Summary

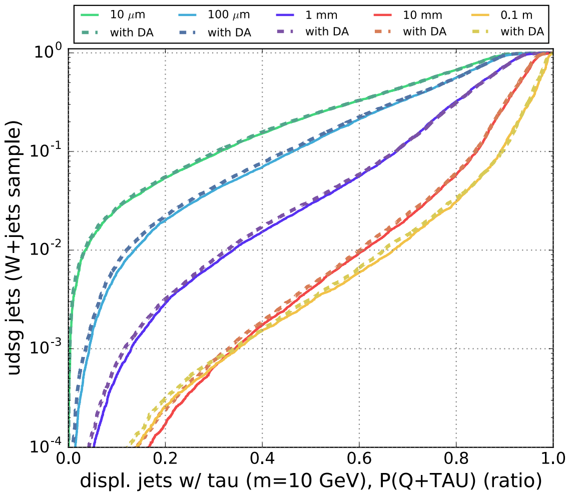
The analysis targets the CC production and decay of long-lived HNLs, resulting in two leptons and at least one jet, and the events are categorised to identify various HNL signal scenarios. An event-level BDT was developed for additional discrimination between signal and background events. The DNN jet tagger was retrained to enhance the sensitivity to HNL decays, which potentially include an displaced lepton. The residual backgrounds must be robustly estimated to extract a hypothetical signal contribution, which is described in Chapter 8.



(a) quark-muon

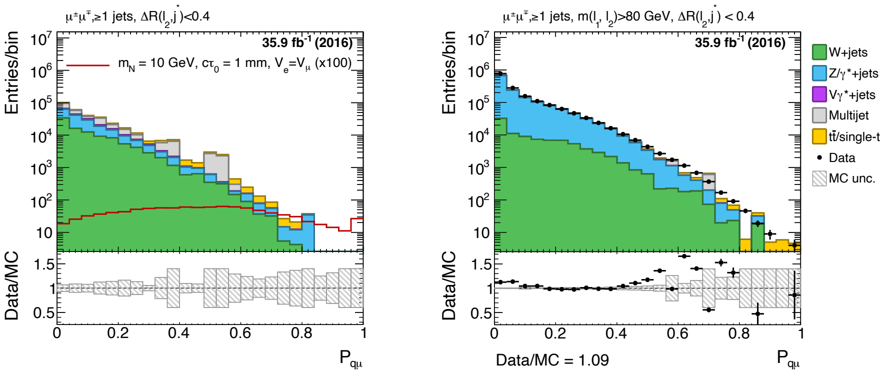


(b) quark-electron

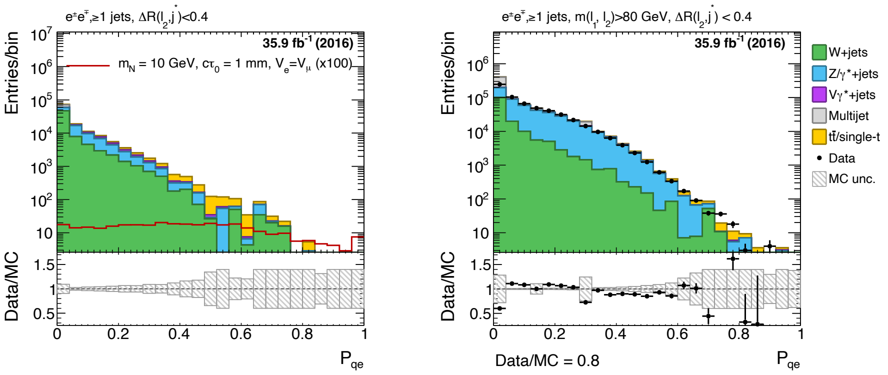


(c) quark

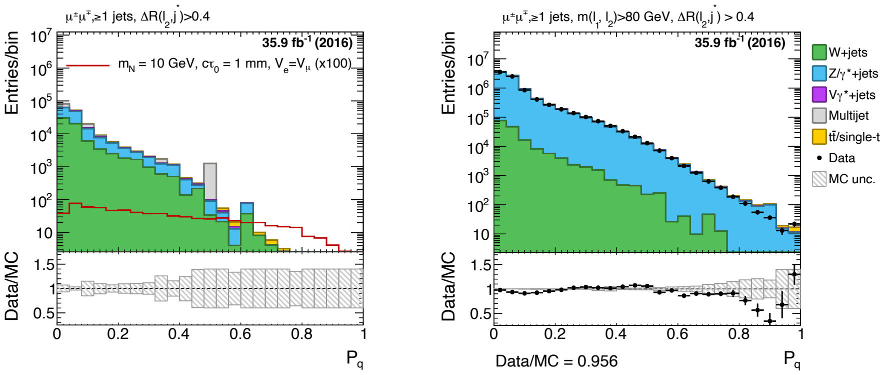
Figure 7.16: Tagger ROC curves for signal (a) quark-muon; (b) quark-electron; (c) quark jets versus the equivalent SM background jet class obtained from a W+jets sample, measured for an HNL with $m_N = 10\,\text{GeV}$ with a range of proper lifetimes.



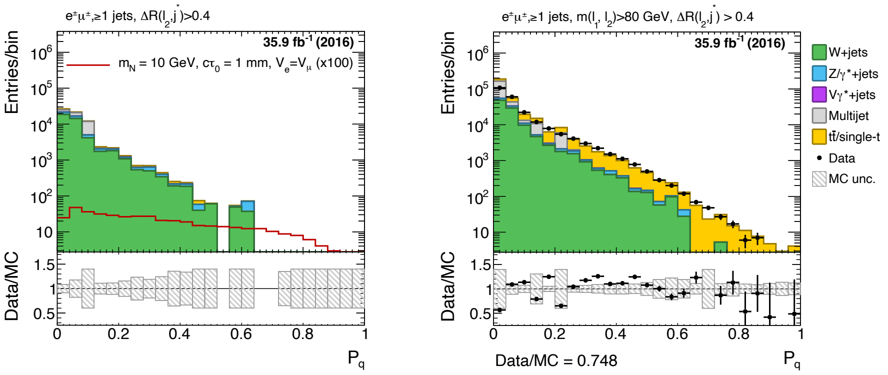
(a) OS $\mu\mu$, boosted



(b) OS ee, boosted



(c) OS $\mu\mu$, resolved



(d) SS $e\mu$, resolved

Figure 7.17: Expected yields in simulation in the SR (left, blinded) and the data/MC comparison in the CR (right) for the 2016 data-taking scenario of the profiled ratio likelihood (a) $P_{q\mu}$ in OS $\mu\mu$ boosted events; (b) P_{qe} in OS ee boosted events; (c) P_q in OS $\mu\mu$ resolved events; (d) P_q in SS $e\mu$ resolved events.

Chapter 8

Background estimation and results

In this Chapter, a data-driven background estimation technique is described. The resulting background yields are used together with the predicted simulated signal yields to extract the search results. Finally, various sources of systematic uncertainty are considered.

8.1 Background estimation

The backgrounds are estimated in an entirely data-driven fashion by using the ABCD method with the BDT and DNN discriminants. A data-driven background estimation is crucial as the number of simulated events was found to be insufficient and the simulation accuracy inadequate, as observed, for example, in the modelling of the event-level BDT score in Figure 7.11.

8.1.1 The ABCD technique

The ABCD method is a simple technique regularly employed to estimate backgrounds directly from data. ABCD methods exploit two discriminating random variables (observables), denoted X and Y, which are statistically independent for the background processes [160]. By applying thresholds on these observables, all events are partitioned into four regions. Three of these regions, called A, B and C, are background-dominated sidebands, while the fourth, D, is signal enriched. If the observables are independent, the background in region D can be predicted from the three sideband regions as:

$$\tilde{N}_{\mathrm{D}}^{\mathrm{bkg}} = N_{\mathrm{B}}^{\mathrm{bkg}} \times N_{\mathrm{C}}^{\mathrm{bkg}} / N_{\mathrm{A}}^{\mathrm{bkg}}, \tag{8.1}$$

where N_i is the number of observed events in regions $i = \mathrm{A,B,C}$. The background yield is described by the following individual likelihood terms for each of the ABCD regions:

$$\mathscr{L}_i^m(\mu, \boldsymbol{\theta} \mid n_i^m) = \text{Poisson}\left(n_i^m \mid b_i^m(\boldsymbol{\theta}) + \mu \cdot s_i^m(\boldsymbol{\theta})\right) \cdot p(\boldsymbol{\theta}), \tag{8.2}$$

$$\mathscr{L}_i^d(\mu, \boldsymbol{\theta} \mid n_i^d) = \text{Poisson}\left(n_i \mid \frac{b_i^b \cdot b_i^c}{b_i^a} + \mu \cdot s_i^d(\boldsymbol{\theta})\right) \cdot p(\boldsymbol{\theta}), \tag{8.3}$$

where $m = a, b, c$ and $p(\boldsymbol{\theta})$ is defined in Equation 5.35. In this formulation, possible signal contamination in sidebands A, B, and C is explicitly considered in the likelihood fit. The total likelihood is the product of likelihoods for individual independent channels:

$$\mathscr{L}(\mu, \boldsymbol{\theta} \mid \boldsymbol{n}) = \prod_{i=1}^{N} \prod_{j=a,b,c,d} \mathscr{L}_i^j(\mu, \boldsymbol{\theta} \mid n_i^j). \tag{8.4}$$

Assuming the counts are in the Poissonian regime, the intrinsic pre-fit uncertainty on the predictor $\tilde{N}_\mathrm{D} \equiv \tilde{N}_\mathrm{D}^\mathrm{bkg}$ can be expressed as:

$$\frac{\tilde{\sigma}_\mathrm{D}}{\tilde{N}_\mathrm{D}} = \sqrt{\frac{1}{N_\mathrm{A}} + \frac{1}{N_\mathrm{B}} + \frac{1}{N_\mathrm{C}}}. \tag{8.5}$$

The real challenge in using the ABCD method is to find a pair of independent variables. One way of quantifying the independence between two variables is the MI (Equation 5.13). However, there are several shortcomings associated with this approach. In the first instance, the MI is computed from simulation as the SR is blinded; in addition, a binned approximation has to be used as the exact probability density cannot be determined. Thus, a small value of MI is a necessary but not sufficient condition for the ABCD method to hold, as the estimation of the MI itself may be biased. To further validate the method, closure tests have to be performed. Such tests aim to compare the predicted yield, $\tilde{N}_\mathrm{D}$, with the true yield, N_D, in a signal-depleted validation region (VR).

8.1.2 Candidate variable studies

In practice, a trial-and-error approach must be adopted to find a pair of variables from a pool of several candidates. The profiled tagger likelihood score, $P \in \{P_{\mathrm{q}\mu}, P_{\mathrm{q}e}, P_\mathrm{q}\}$, the most discriminating variable used in this analysis, was naturally chosen as the first candidate for the ABCD method. For the second variable, several candidates were initially considered. As the tagger performance was correlated with displaced jet kinematic quantities, as shown in Figure 6.11, only variables not directly depending on such features were considered. These variables were the angular separation $\Delta\phi(\ell_1, \ell_2)$, the $p_{\mathrm{T}}^{\mathrm{miss}}$, the m_{T}, the aplanarity, and the BDT signal likelihood. A systematic study was performed using simulation to find the best candidate based on MI. The study was done per event topology and dilepton category, using all significant background sources except for QCD multijet production, which would have highly biased the results due to the relatively large events weights. As can be seen from Table 8.1, a similar level of correlation is observed for all variables, and they all performed well during initial tests. All other things being equal, the BDT was chosen as it is the most discriminating variable. Two-dimensional histograms of the BDT and P distributions for simulated background events are shown in Figure 8.2 for different event categories.

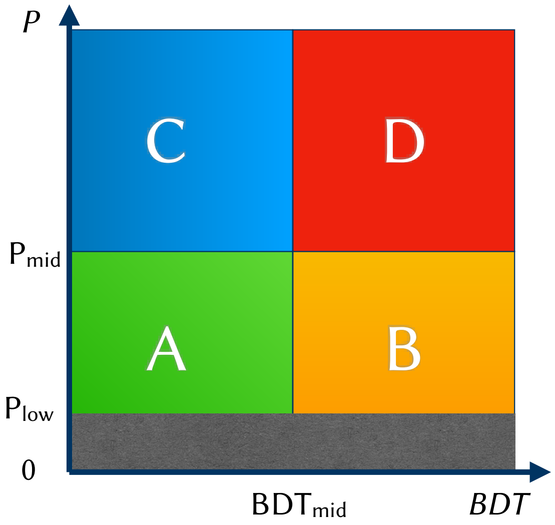


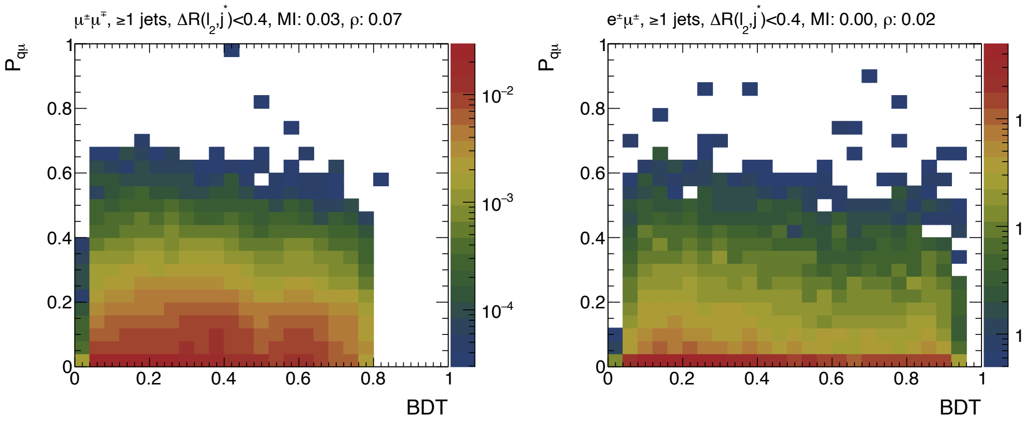
Figure 8.1: A sketch of the ABCD region definitions. The region populated below P_{low} is discarded to ensure only signal-like jets are selected.

8.1.3 Simulation checks

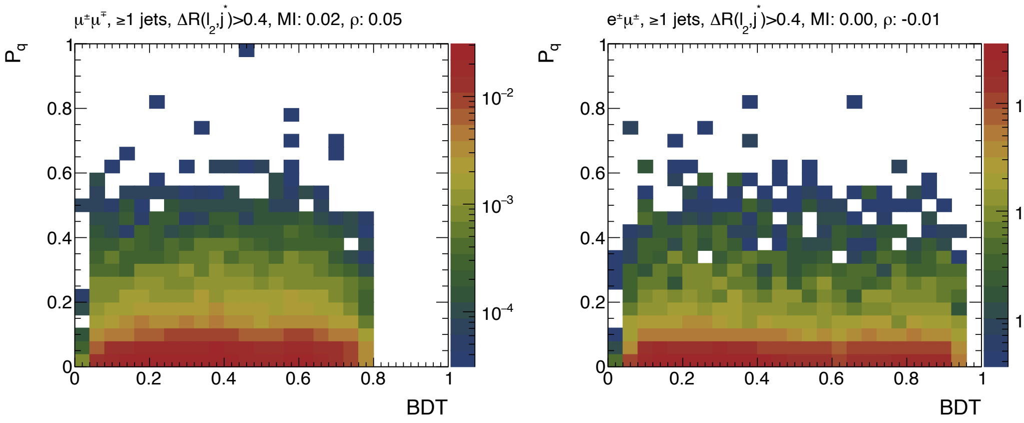
In this analysis, all sources of background are estimated simultaneously. Even if the two variables are independent for different background processes, the statement might not hold when considering all sources of background simultaneously. Specifically, the statement of independence implies that the probability density function factorises:

$$\rho(x, y) = p(x)q(y); \tag{8.6}$$

in other words, the distribution of one variable is the same regardless of any selection on



(a) OS $\mu\mu$, boosted (b) SS $e\mu$, boosted



(c) OS $\mu\mu$, resolved (d) SS $e\mu$, resolved

Figure 8.2: Two-dimensional histograms (heatmaps) of P and the BDT score for simulated background events in the SR for boosted (top-row) and resolved (bottom row) categories, OS $\mu\mu$ (left) and SS $e\mu$ (right) final states. The MI and the Pearson correlation coefficient between the two variables are indicated in the plots.

the second one. This is not, in general, the case for the sum of two background sources with differing densities:

$$\rho(x,y) = \rho_1(x,y) + \rho_2(x,y) = p_1(x)q_1(y) + p_2(x)q_2(y) \neq p(x)q(y). \tag{8.7}$$

For the ABCD assumption to hold, the background distribution in at least one of the ABCD variables has to be the same for all background processes, i.e. $p_1(x) = p_2(x) = p(x)$:

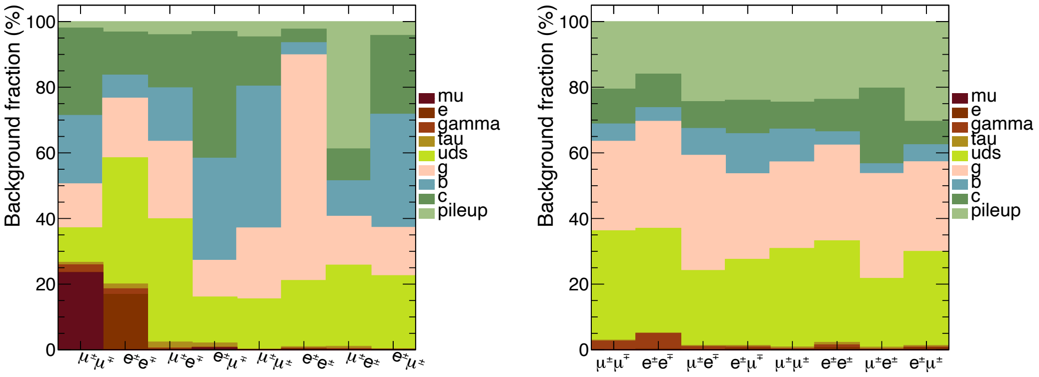
$$\rho(x,y) = p(x)(q_1(y) + q_2(y)) = p(x)q(x). \tag{8.8}$$

In our case, it can be seen from Figure 8.3 that different jet classes populate the SR, and each of them will have a different tagger score distribution. For instance, the tagger

Table 8.1: The MI per category between the jet tagger score P and different candidate variables per category.

$\ell_1\ell_2$	topology	M.I. (x100)				
		$\Delta\phi(l_1,l_2)$	$p_{\mathrm{T}}^{\mathrm{miss}}$	m_{T}	aplanarity	BDT
$\mu^{\pm}\mu^{\mp}$	resolved	5.0	5.5	5.2	4.6	5.9
	boosted	4.9	4.3	4.4	2.3	4.2
$e^{\pm}e^{\mp}$	resolved	8.9	7.2	7.7	5.8	7.5
	boosted	5.2	3.8	4.5	2.4	5.1
$\mu^{\pm}e^{\mp}$	resolved	9.2	9.6	10.1	6.5	10.5
	boosted	3.6	3.8	3.9	2.6	4.7
$e^{\pm}\mu^{\mp}$	resolved	7.7	8.5	9.6	5.6	10.4
	boosted	5.6	6.3	6.8	3.4	6.3
$\mu^{\pm}\mu^{\pm}$	resolved	8.2	7.4	7.5	4.8	7.8
	boosted	6.1	6.3	6.7	3.4	6.5
$e^{\pm}e^{\pm}$	resolved	12.4	13.6	12.1	6.3	12.4
	boosted	4.5	4.4	4.0	3.0	4.6
$\mu^{\pm}e^{\pm}$	resolved	12.2	11.8	13.7	8.0	12.6
	boosted	3.9	4.1	4.4	2.4	3.9
$e^{\pm}\mu^{\pm}$	resolved	8.2	9.3	9.4	6.1	7.9
	boosted	7.5	6.4	7.4	4.2	5.9

discriminates better against light-flavour than heavy-flavour jets as the latter case results in significantly displaced non-prompt leptons.

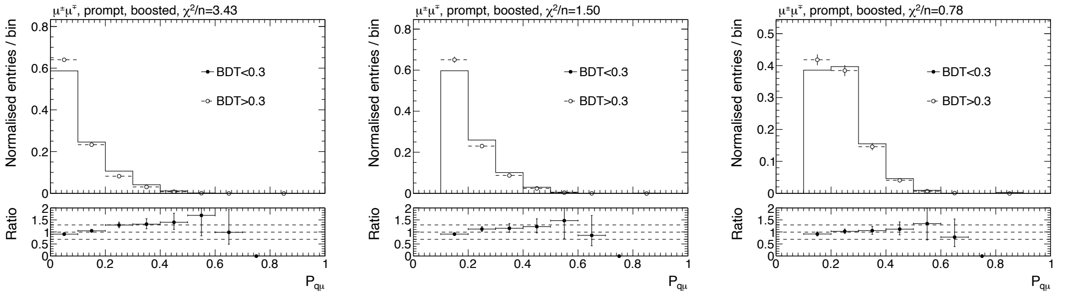


(a) boosted

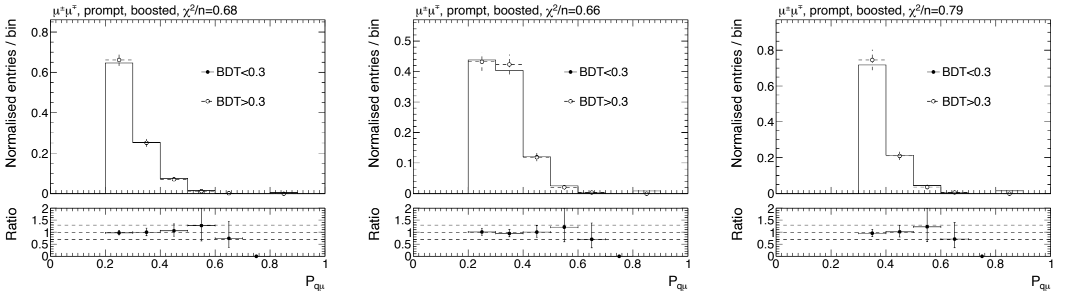
(b) resolved

Figure 8.3: Background composition by j⋆ truth label for (a) boosted and (b) resolved categories, broken down by dilepton flavour and sign final states.

As an additional check, the tagger score distributions are examined in two bins of the BDT score, as shown in Figure 8.4. There is a reasonable level of agreement between the two normalised distributions. The compatibility of the two histograms is quantified using a χ^2 test-statistic for different P_{min} thresholds. Increasing the lower threshold means the ABCD prediction is performed further from the bulk of the distribution. In other words, only events that are somewhat signal-like are selected. However, the statistical precision of the method is reduced as a result. The lower threshold on the tagger score was set to $P_{\text{min}} = 0.3$ (0.2) for boosted (resolved) scenarios to improve the closure of the ABCD method, as indicated by reduced values of χ^2. To ensure the robustness of this choice and check for any bias, P_{min} was varied from its nominal value in closure tests.



(a) $P_{\text{min}} = 0.00$ (b) $P_{\text{min}} = 0.10$ (c) $P_{\text{min}} = 0.15$



(d) $P_{\text{min}} = 0.20$ (e) $P_{\text{min}} = 0.25$ (f) $P_{\text{min}} = 0.30$

Figure 8.4: Distributions of $P_{\text{q}\mu}$ in low- and high-BDT score regions for the dimuon OS prompt boosted category, shown with differing P_{min} thresholds.

8.1.4 Threshold optimisation

The optimal thresholds (P_{mid}, BDT_{mid}) were derived per category. This is necessary as the background yields, the tagger performance, and the BDT performance vary significantly depending on the SR category. Optimal thresholds were considered separately for each year due to different data-taking conditions, different DNN models used, and different trigger thresholds. The thresholds were chosen using a procedure to optimise the discovery significance of Equation 5.50, while requiring at least ten expected background events in the three sidebands A, B, and C. Several signal samples, listed in Table 8.2, expected to be close to the expected analysis sensitivity reach, were used in the optimisation to derive the best average significance, to not bias the thresholds for any particular signal model. The optimal thresholds BDT_{mid} were typically found in the range 0.5–0.8. The optimal P_{mid} thresholds were usually 0.6–0.8 for boosted and 0.4–0.5 for resolved event topologies.

Table 8.2: Signal samples used to optimise the tagger and BDT thresholds. For each model, coupling scenarios summarised in Table 7.1 were considered in the optimisation.

mass (GeV)	$c\tau_0$ (mm)
2	1000
4.5	100
8	10
10	1
12	0.1
16	0.001

8.1.5 Closure tests

To verify the validity of the ABCD method, two VRs were defined to perform closure tests in data:

VR1 (mass sideband) The $m(\ell_1, \ell_2, \mathrm{j}^\star)$ requirement is inverted to $m(\ell_1, \ell_2, \mathrm{j}^\star) < 70\,\mathrm{GeV}$. The optimal tagger threshold P_{mid} is reduced by 20% in each category to increase the statistical precision of the test. The results are shown in Figure 8.5.

VR2 (tagger sideband) The low tagger score sideband is subdivided into four new regions, i.e. $AB \longrightarrow A'B'C'D'$. The results are shown in Figure 8.6.

To check for bias in the ABCD method, a free parameter, κ, scaling the predicted yield was introduced:

$$\tilde{N}_{\mathrm{D}}^{\mathrm{bkg}} = \kappa N_{\mathrm{B}}^{\mathrm{bkg}} \times N_{\mathrm{C}}^{\mathrm{bkg}} / N_{\mathrm{A}}^{\mathrm{bkg}}. \tag{8.9}$$

Six such parameters were considered, one for each combination of year of data taking and topology (boosted and resolved). Each parameter was simultaneously constrained in the VR closure test from a maximum-likelihood fit to 24 independent SR categories. This choice was made considering a separately trained tagger and BDT model for each year and the different data-taking conditions. Different parameters were needed for each topology due to significant differences in tagger performance between boosted and resolved jets. For each category group, the best-fit value of κ is indicated in the respective Figure; the values are also summarised in Table 8.3. Overall, the ABCD hypothesis holds up well. Based on the most significant deviation from unity, a conservative 15% systematic uncertainty was assigned for each of the six category groups.

A closure test was also performed for each macrocategory, i.e. without the categorisation of events based on dilepton sign, flavour, and d_{xy}^{sig}. The results of this check are also shown in Table 8.3, which shows the overall ratio of observed to predicted events, obtained by using the same threshold for all categories. In addition, the lower threshold, P_{low}, was varied by ±10% from its nominal value, and the closure test was repeated to assess the stability of the method. Overall, a good agreement between the three results is observed, suggesting the prediction does not strongly depend on the particular choice of P_{low}.

Finally, a GOF test, described in Section 5.3.4, is performed for the VRs. The probability distribution of the toy statistic is generated using MC toys and compared to the observed value. The results are shown in Figure 8.7, and a p-value is derived. Overall, a high p-value is observed for both VRs, validating the ABCD method.

Table 8.3: Best-fit values κ and ratios of predicted to observed yields for a combined closure test performed per year of data taking and event topology (boosted, resolved). In addition, the ratio results are shown when varying the lower DNN tagger threshold P_{low} by ±10% from its nominal value.

Closure test	Year	κ	Ratio	Ratio (+10%)	Ratio (−10%)
VR1, resolved	2016	0.92 ± 0.05	0.96 ± 0.02	0.98 ± 0.02	0.95 ± 0.02
	2017	0.96 ± 0.03	0.93 ± 0.02	0.94 ± 0.02	0.93 ± 0.02
	2018	1.00 ± 0.05	0.95 ± 0.02	0.97 ± 0.02	0.95 ± 0.02
VR2, resolved	2016	0.99 ± 0.03	0.98 ± 0.03	0.97 ± 0.03	0.98 ± 0.03
	2017	1.03 ± 0.03	1.02 ± 0.03	1.05 ± 0.03	1.03 ± 0.02
	2018	1.03 ± 0.02	1.03 ± 0.02	1.03 ± 0.02	1.04 ± 0.02
VR1, boosted	2016	0.87 ± 0.04	0.98 ± 0.02	0.98 ± 0.02	0.98 ± 0.02
	2017	1.04 ± 0.05	0.99 ± 0.03	0.98 ± 0.03	0.98 ± 0.03
	2018	0.86 ± 0.03	0.95 ± 0.02	0.95 ± 0.02	0.93 ± 0.02
VR2, boosted	2016	0.98 ± 0.02	0.97 ± 0.02	0.94 ± 0.03	0.99 ± 0.02
	2017	1.01 ± 0.03	1.00 ± 0.03	0.99 ± 0.04	1.01 ± 0.02
	2018	1.01 ± 0.02	1.01 ± 0.02	0.99 ± 0.03	1.02 ± 0.02

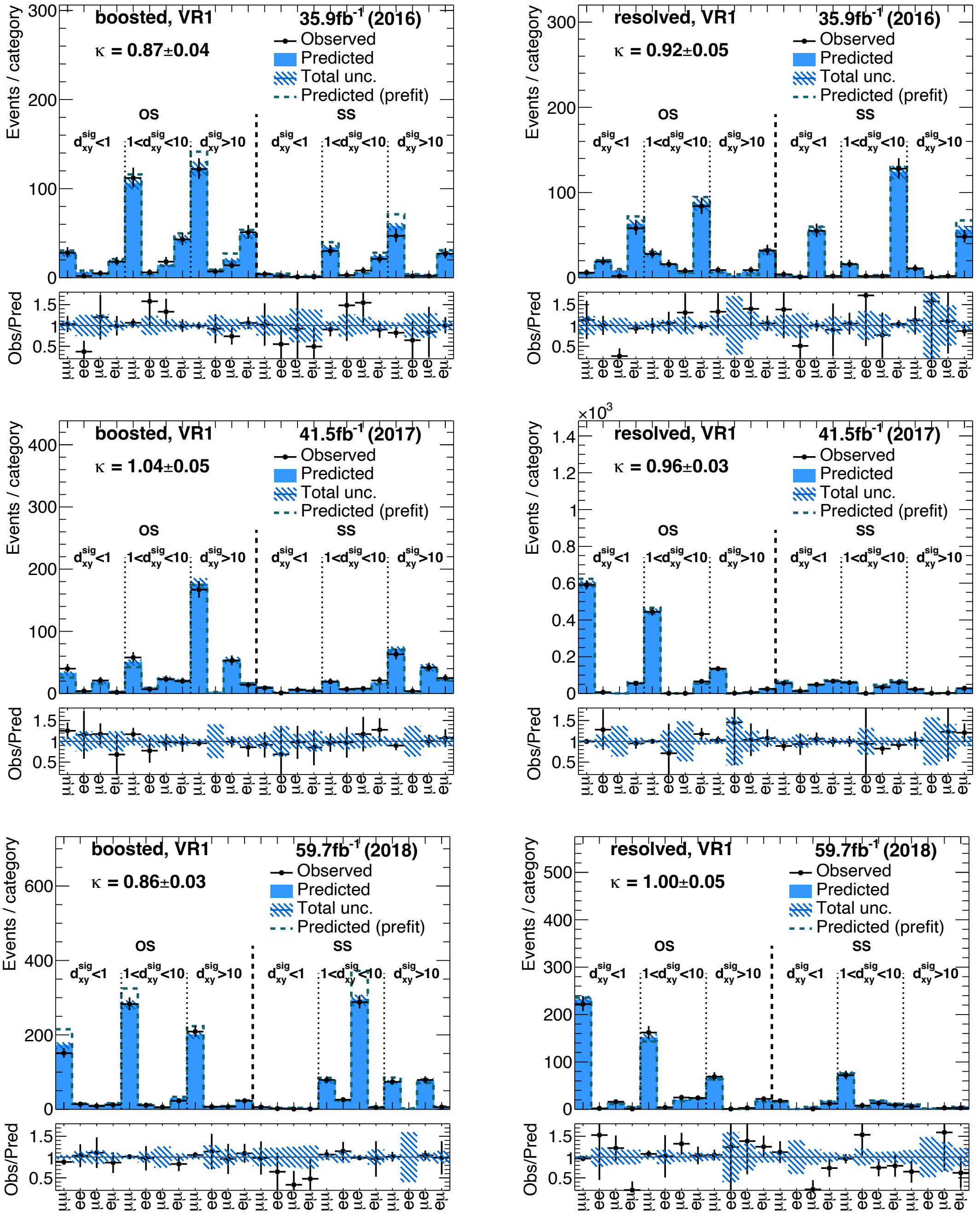


Figure 8.5: Closure test in VR1 ($m(\ell_1, \ell_2, \mathrm{j}^\star) < 70\,\mathrm{GeV}$) for boosted (left) and resolved (right) categories for 2016 (top row), 2017 (middle row), and 2018 (bottom row) data-taking scenarios.

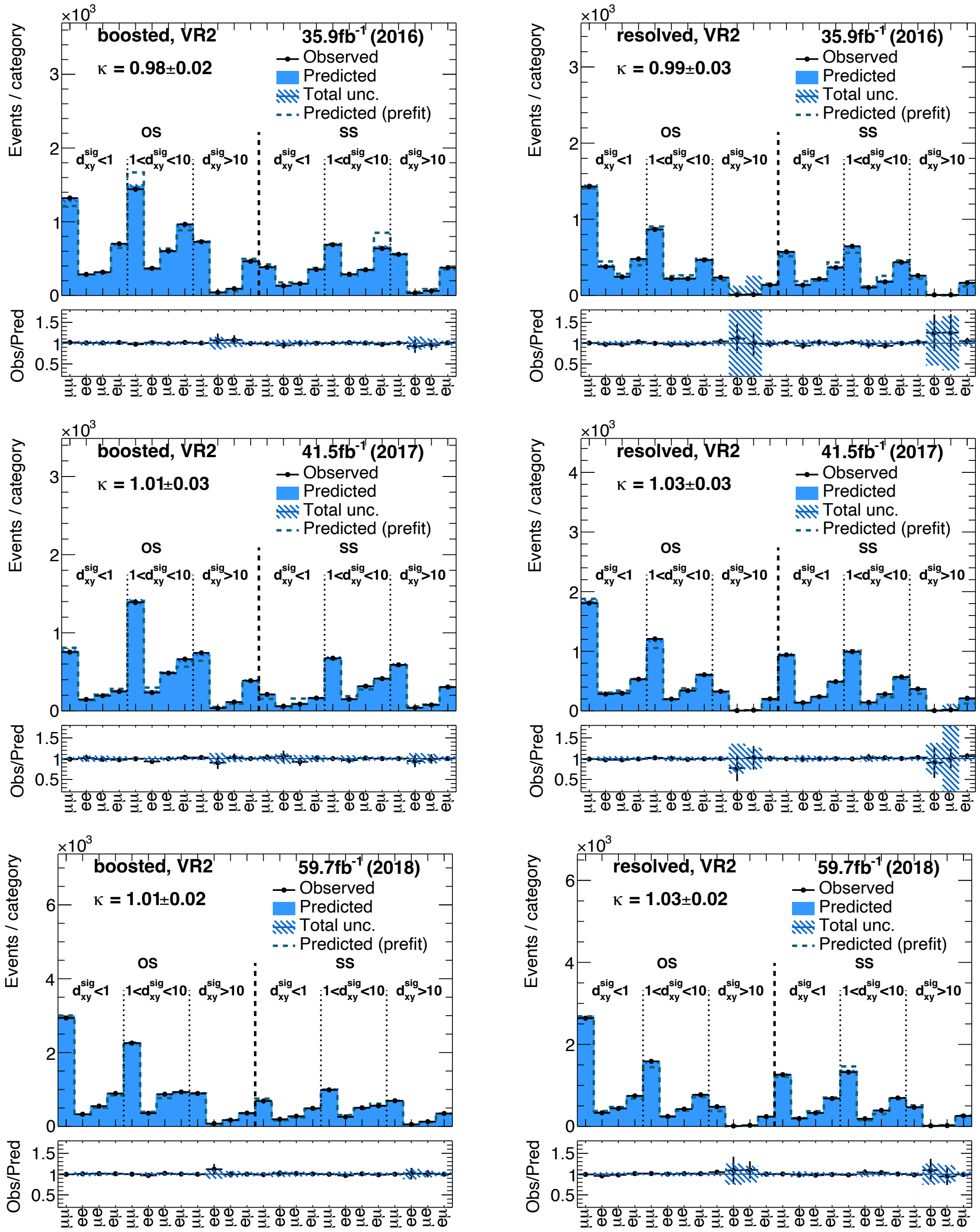


Figure 8.6: Closure test in VR2 (sideband AB) for boosted (left) and resolved (right) categories for 2016 (top row), 2017 (middle row), and 2018 (bottom row) data-taking scenarios.

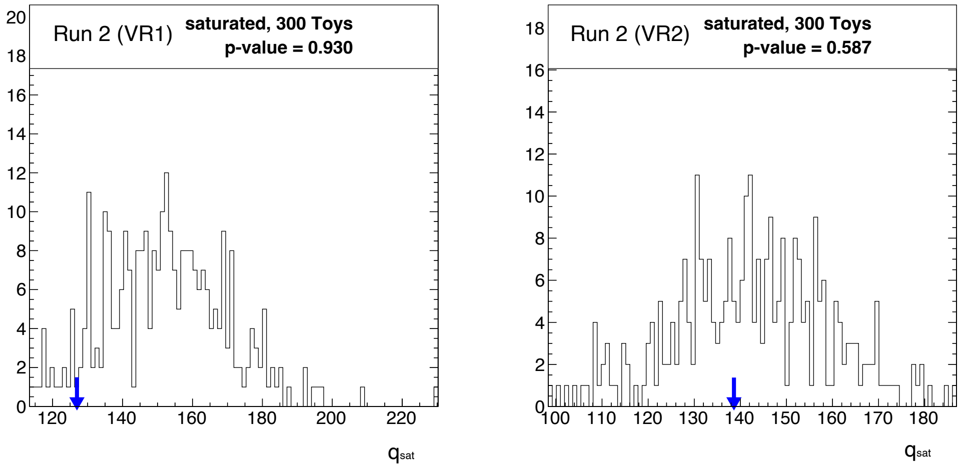


Figure 8.7: The predicted saturated likelihood test statistic distribution and the observed value (in blue) for VR1 (left) and (b) VR2 (right) GOF test.

8.2 Systematic uncertainties

In this Section, the considered sources of systematic uncertainty are detailed. Each systematic uncertainty is treated as a NP in the statistical model and profiled in background-only, and signal-plus-background maximum likelihood fits to data. Most uncertainties only impact the signal prediction since it is taken from simulation, whereas the number of expected background events is determined from data. The systematic uncertainties are considered to be uncorrelated between the three years of data taking. The magnitudes of systematic uncertainties are discussed together with the search results in Section 8.3.

8.2.1 Signal systematic uncertainties

Prompt lepton efficiencies Systematic uncertainties on the prompt lepton reconstruction, identification and trigger efficiencies are accounted for by varying the p_T and η-dependent scale factors independently within one standard deviation of their measured uncertainties.

Displaced lepton efficiencies A 10% systematic uncertainty is assigned to account for potential mismodelling in displaced lepton reconstruction and identification efficiency, based on the data/MC agreement observed in the CR.

Displaced track efficiency The distribution of track $|d_{xy}|$ in data and simulation is used to derive an uncertainty on the track reconstruction efficiency, similarly to the approach described in Equation 6.6, and acts as a proxy to account for potential mismodelling of the DNN jet tagger score for signal events. The weights are measured in single-lepton and jet (e/μ + jets) samples, with no ℓ_2 reconstructed, and with a further requirement of $p_\mathrm{T}^\mathrm{miss} > 100\,\mathrm{GeV}$. This selection ensures that the samples are enriched in b-jet topologies from the $\mathrm{t\bar{t}}$ process and contain genuine displaced tracks and SVs. The track-level weights are then propagated to the jet level by varying the uncertainties for the three tracks with largest d_{xy}^sig values in a given jet.

Pileup An uncertainty of 5% on the minimum bias cross-section is taken as a systematic uncertainty when calculating the expected data pileup distribution.

JES In simulated events, the JES is modified for all jets using the uncertainty associated with the p_T and η-dependent scale factors. The resulting differences are propagated to the missing transverse momentum.

JER The JER is modified in simulated events by either increasing or reducing the difference in p_{T} with respect to the matched jet at the generator level. Random smearing is used when no generator-level jet is found within $\Delta R < 0.4$.

Unclustered energy The component of the missing transverse momentum, which is not clustered inside jets and thus remains uncorrected by the JES and resolution, is varied within its uncertainty.

Luminosity An uncertainty of 1.2%, 2.3%, and 2.5% on the measured integrated luminosity is included for 2016, 2017, and 2018 data-taking periods, respectively.

Parton distribution function Uncertainties in the PDF are determined by reweighting each event according to the weights of 100 replicas (sets of MC representations of the experimental data) that provide the PDF probability distribution [13].

Factorisation and renormalisation scale The factorisation and renormalisation scale are varied by factors of 2 or 0.5 in simulated samples using precomputed weights by the event generators. Only residual changes after the selection are taken as uncertainty for the signal samples while the overall effect on the normalisation (approximately ±12%) is considered part of the theoretical uncertainty instead.

W boson $\mathbf{p_{T}}$ correction Higher-order corrections to the p_{T} spectrum of the W boson in W+jets production that are also applicable to signal events have been investigated, but were found to be already covered by the factorisation and renormalisation scale uncertainties. Hence, no further correction was applied.

8.2.2 Background systematic uncertainties

Data-driven background yield As detailed in Section 8.1.5, a 15% non-closure uncertainty is assigned, and simultaneously constrained by several categories in the fit.

8.3 Results

After extensive validation of the background estimation method, the observed SR yields in region D were unblinded. The results are shown in Figure 8.8. The expected yields for two benchmark HNL models are also included: one low-mass, high-lifetime and another higher-mass, low-lifetime scenario, which are expected to be excluded by this analysis. No deviation from the background prediction can be seen. A GOF test of the background-only hypothesis is shown in Figure 8.9, and the observed data is consistent with the background-only hypothesis ($p > 0.05$). Given the agreement between predicted and observed yields, no evidence for HNLs is observed, and ULs are set.

8.3.1 Interpretation

The results are interpreted by deriving ULs on the signal strength $\sigma_{\mathrm{UL}}(m_{\mathrm{N}}, V_\ell^2)$, as described in Section 5.3. Observed (expected) ULs on the signal cross-section are obtained by performing the fit in the asymptotic approximation with the observed counts (Asimov data set). The UL is compared with the theoretical cross-section $\sigma_{\mathrm{th}}(m_{\mathrm{N}}, V_\ell^2)$ obtained using MadGraph at LO, multiplied by a k-factor of 1.1 as described in Section 7.2. The ULs are presented in the mass-coupling plane for various benchmark scenarios, and additionally, in the whole coupling space for a fixed HNL $c\tau_0$.

8.3.2 Mass-coupling plane limits

A given point is excluded if the signal strength parameter satisfies the following:

$$\mu(m_{\mathrm{N}}, V_\ell^2) = \frac{\sigma_{\mathrm{UL}}(m_{\mathrm{N}}, V_\ell^2)}{\sigma_{\mathrm{th}}(m_{\mathrm{N}}, V_\ell^2)} < 1. \tag{8.10}$$

A grid of $\sigma_{\mathrm{UL}}(m_{\mathrm{N}}, V_\ell^2)$ values is obtained by cubic spline interpolation between the available benchmark signal points. This allows finding the exclusion contour corresponding to $\mu(m_{\mathrm{N}}, V_\ell^2) = 1$. The obtained limits are shown per year of data taking and mixing scenario in Figure 8.10.

8.3.3 Flavour-generic limits

Limits in the barycentric space are given in Figure 8.11 for a Dirac HNL scenario with $c\tau_0 = 1\,\mathrm{mm}$ and $c\tau_0 = 10\,\mathrm{mm}$. For each coupling scenario, the maximum excluded HNL

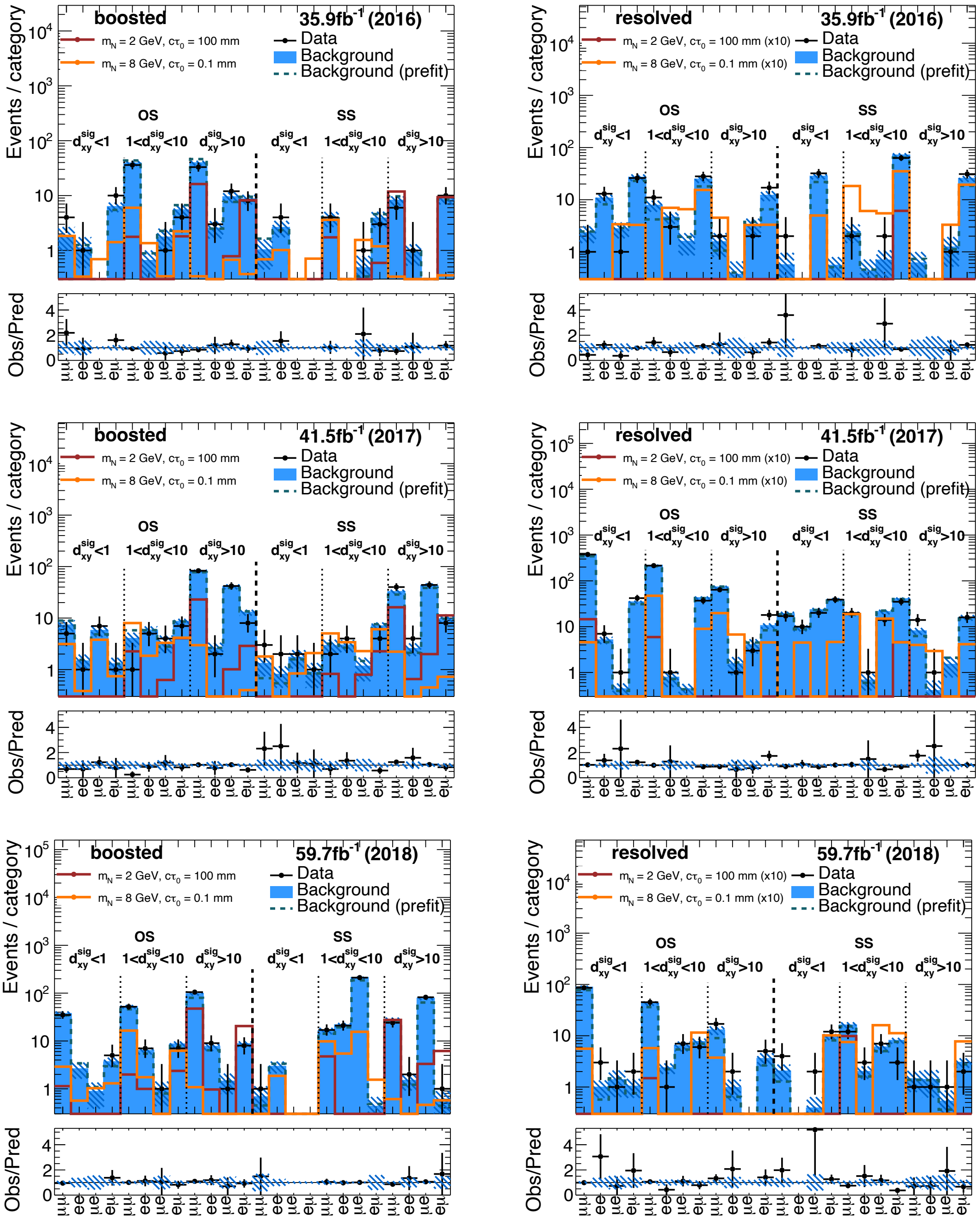


Figure 8.8: Observed and predicted yields in the SR for boosted (left) and resolved (right) categories for 2016 (top row), 2017 (middle row), and 2018 (bottom row) data-taking scenarios.

mass is shown colour-coded. These plots are obtained by interpolating the cross-sections linearly in the 5D-space spanned by (f_e, f_μ, f_τ, $c\tau_0$, m_N) within the 5-simplex constructed from the closest points. For fixed values of f_e, f_μ, f_τ, and $c\tau_0$, the maximum m_N is found where $\mu < 1$.

There is a direct connection between the limits shown in Figure 8.10 and those in Figure 8.11. From each of the six benchmark scenarios (plots) of Figure 8.10, a single point

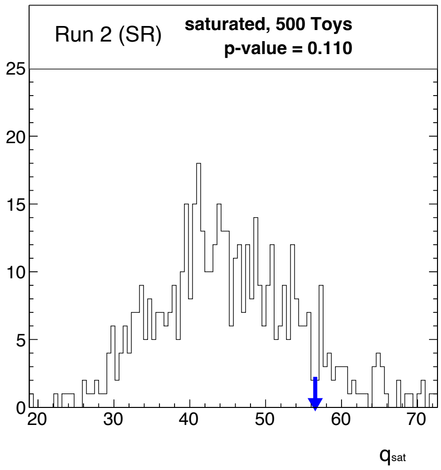


Figure 8.9: The predicted saturated likelihood test statistic distribution and the observed value (in blue) for the SR GOF test.

on the triangle plot can be obtained. This can be achieved by finding the HNL mass at which the observed limit curve crosses a diagonal constant-$c\tau_0$ line.

8.3.4 Discussion

The analysis is sensitive to a large portion of mixed-coupling scenarios involving all three lepton flavours. The best relative sensitivity is obtained for the pure-muon coupling. Due to the lower efficiency of reconstructing displaced electrons, the limits are, as expected, worse for the pure-electron coupling scenario at low HNL masses. The analysis is not sensitive to the pure-tau coupling scenario due to the low trigger efficiency.

The most stringest ULs on the HNL production cross-section are set for tracker lifetimes of 1–10 mm and the mass range of 1–20 GeV, corresponding to $\sigma_{\mathrm{UL}} \lesssim 0.5$ pb for $V_e = V_\mu = V_\tau$. Outside of this range, σ_{UL} increases as either the signal becomes too prompt or too displaced. In general, as the HNL mass increases, the exclusion sensitivity increases up to the turning point, which corresponds approximately to 8 GeV, and starts to decrease for higher HNL masses. This can be explained by decreased HNL displacement and increased reliance on the resolved event topology; as for both of these scenarios, DNN tagger performance is somewhat reduced compared with displaced, boosted events. In other words, the exclusion line evolves from covering low-mass, displaced scenarios to covering high-mass, prompt scenarios. Some fluctuations in the exclusion contour shape can be observed for $m_{\mathrm{N}} > 8$ GeV, resulting from reduced signal sample granularity at higher masses, as shown in Figure 7.2.

The 2D-space limits for pure-muon and pure-electron scenarios are compared with recent CMS [48], ATLAS [47] and LHCb [68] results. The CMS results are significantly worse for

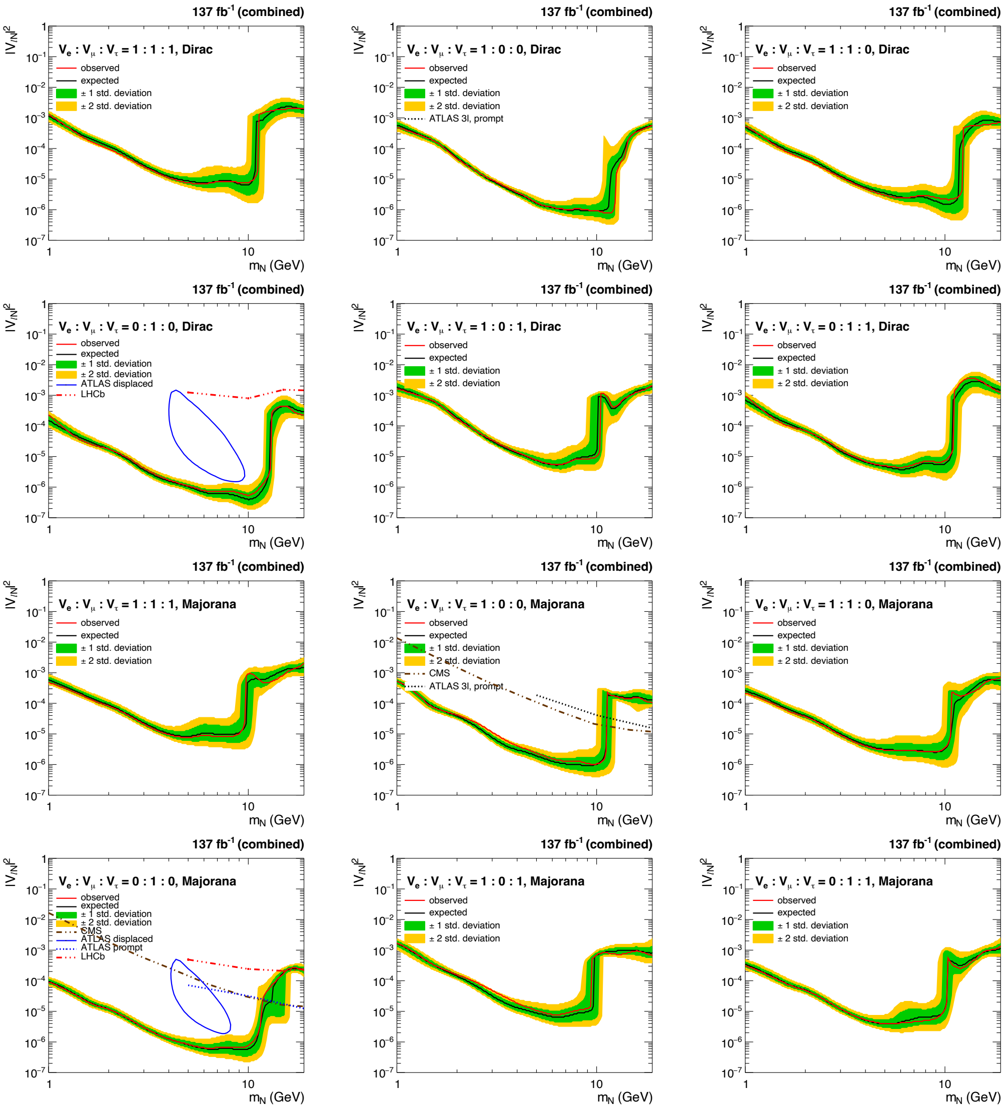
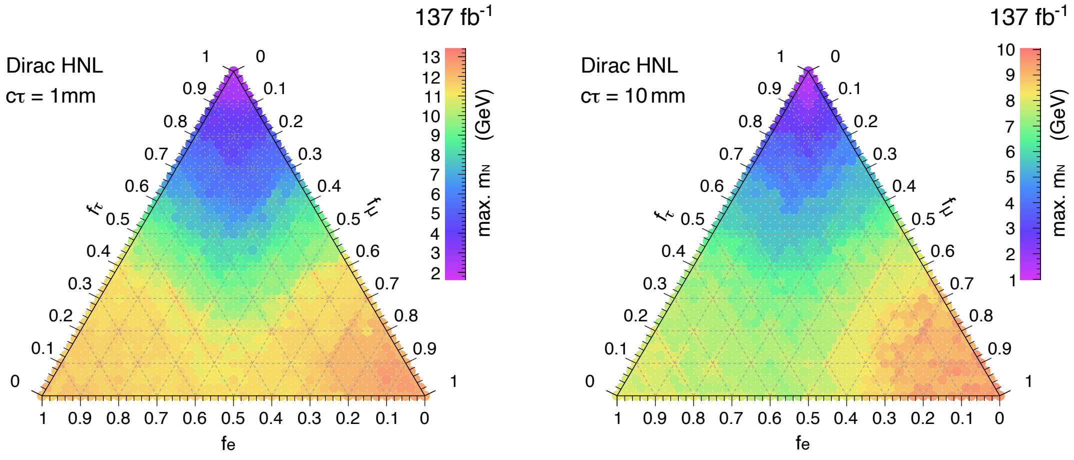


Figure 8.10: Expected and observed limits for Dirac and Majorana HNLs for different benchmark coupling scenarios (listed in Table 7.1).

long-lived than prompt HNLs, as the search is focused on prompt leptons. The ATLAS results have a closed, ellipsoid-like shape in the mass-coupling plane, as the analysis is not sensitive to prompt scenarios. This is due to a displaced vertex requirement that is inefficient for prompt HNL decays. For the novel interpretation, corresponding to universal HNL coupling to all three generations, no results exist with which a direct comparison could be made.

Table 8.4 shows the impacts (defined in Equation 5.51) of systematic uncertainties. Low sideband counts in data, which limit the accuracy of the data-driven background prediction, primarily drive the uncertainty, considering the large number of SR categories. In other words, the analysis is statistically limited. ABCD non-closure followed by displaced lepton identification are the most significant systematic uncertainties. JES and JER uncertainties are considerable for soft signal jets and affect the DNN jet tagger score due to



(a)

(b)

Figure 8.11: Maximum excluded mass for each flavour combination for Dirac HNLs with (a) $c\tau_0 = 1$ mm and (b) $c\tau_0 = 10$ mm.

jets migrating in and out of selection above a given p_{T} threshold. Likewise, unclustered energy is an important source of uncertainty when considering low-p_{T} final states, and especially affects the event-level BDT discriminant as it relies on $p_{\mathrm{T}}^{\mathrm{miss}}$ as one of its inputs.

8.4 Summary

The data-driven technique described in this Chapter provides a robust estimation of the background rates, validated in both simulation and data. The predictions are compared against observed data in the SR, corresponding to 137 fb^{-1}, and the negative search results are interpreted by setting ULs on various HNL production scenarios. The results are interpreted for long-lived and prompt HNLs of Dirac and Majorana nature in the 1–20 GeV mass range. Good agreemeent between the observed and expected ULs is observed, and the limits are competitive with previous measurements. For the first time in the LHC, flavour-universal limits are obtained by considering HNLs that couple to all three lepton generations simultaneously. The analysis is statistically limited by the low background sideband yields of the data-driven prediction.

Table 8.4: Impacts of systematic uncertainties on the signal cross-section. For the background prediction sideband yield impacts, an average over 48 categories is provided.

Systematic uncertainty	Impact (%)		
	2016	2017	2018
Prompt electron reco.	0.0	0.1	0.0
Prompt electron id.	0.1	0.3	0.4
Prompt muon id.	0.0	0.1	0.0
Prompt muon iso.	0.0	0.0	0.0
Trigger	0.1	0.0	0.0
Displaced electron reco.	0.0	0.1	0.3
Displaced lepton id.	7.2	7.7	8.8
Displaced track	0.4	3.7	4.0
Pileup	2.1	3.5	5.8
Jet energy scale	2.6	1.5	5.6
Jet energy resolution	1.0	2.5	1.9
Unclustered energy	3.6	4.3	9.0
Luminosity	1.1	0.5	0.8
PDF	2.8	2.7	3.0
Scale	0.1	0.3	0.4
Non-closure, boosted	24.6	21.2	15.5
Non-closure, resolved	25.2	22.8	40.4
Sideband A yield (average)	0.8	0.7	1.4
Sideband B yield (average)	4.1	4.5	5.5
Sideband C yield (average)	11.4	7.9	9.7

Conclusions

Chapter 6 presented a novel DNN displaced jet tagging algorithm. The tagger was developed to target long-lived gluino decays and achieved excellent performance for tracker and calorimeter displacements, rejecting 99.99% of background jets while retaining around half of the signal jets. The tagger was trained on jets from data and simulated samples with DA, penalising mismodelled input features. As a result, the modelling of the tagger output improved significantly without a major loss in performance. Finally, the tagger was found to generalise well to different flavours of displaced jets.

Chapter 7 presented a search for long-lived HNLs decaying to displaced jets and leptons using a data set of $131.7\,\mathrm{fb}^{-1}$ of pp collision data. The search exploits the displaced jet tagger to identify signal jets. The analysis benefits from a finely segmented categorisation scheme, which enhances the sensitivity to different signal scenarios. An event-level BDT was developed for additional discrimination between signal and background events.

The jet tagger and BDT are used together to perform a data-driven background prediction described in Chapter 8. Given the agreement between the predicted and observed yields, ULs on HNL production cross-section were set for various HNL mass, lifetime, and coupling scenarios. This is the first search at the LHC considering universal coupling of HNLs to all active neutrino generations, while the results for single-generation couplings remain competitive with existing analyses. For the universal coupling $V_e = V_\mu = V_\tau$ in the mass range of 1–20 GeV, the best exclusion $\sigma_{\mathrm{UL}} \lesssim 0.5\,\mathrm{pb}$ is observed for tracker lifetimes of 1–10 mm. Good sensitivity for prompt HNL scenarios is also achieved, as there is no explicit requirement on a displaced HNL decay vertex.

Several possible improvements outside the scope of this thesis are presented in the following. The analysis only achieves weak sensitivity for pure HNL coupling to the third generation, and future work in this area, such as dedicated displaced hadronic tau decay reconstruction techniques, would be of great interest. In addition, the analysis would benefit from (re-)interpretability studies of the MVA discriminants used for signal identification, to facilitate the derivation of signal efficiencies of various hypothesised HNL models in future phenomenological efforts.

Abbreviations

νSSM Neutrino Minimal Standard Model 26, 29

Adam Adaptive Moment Estimation 77, 109

ATLAS ATLAS 28, 35, 94, 166, 167

BBN Big Bang nucleosynthesis 26, 28

BDT boosted decision tree xiii, xv, xvi, xix, xx, 2, 59, 60, 67, 82, 83, 94, 120, 139–141, 147, 151, 153–157, 168, 171, 194

BR branching ratio xii, 3, 32, 121, 124, 125

BSM beyond Standard Model 23, 31–33, 36, 38, 45–47, 64, 67, 89, 118

CC charged-current xviii, 14, 121, 135, 147

CERN the European Organisation for Nuclear Research xiv, xv, 38, 39, 41, 94

CMB cosmic microwave background 21, 23

CPU central processing unit 109

CR control region xii, xiii, xviii–xxi, 51, 83, 107, 108, 116–118, 127, 135, 136, 139, 141, 142, 144, 147, 149, 162, 194–201

CSC cathode strip chamber xv, 46, 47, 55

DA domain adaptation xii, xvi–xviii, 3, 83–85, 95, 107, 111–114, 116–118, 144, 147, 171

DAQ data acquisition system 38, 47

DL deep learning 2, 81, 94

DM dark matter 20, 23, 24, 26, 34, 45

DNN deep neural network xiii, xvi–xviii, 2, 51, 79, 94–96, 101, 105, 106, 115–118, 120, 131, 132, 139, 140, 144, 147, 151, 156, 158, 162, 166, 167, 171, 194, 202

DT drift tube xv, 46, 47, 55

DUNE Deep Underground Neutrino Experiment 28

EB electromagnetic barrel 43, 44, 60, 207

ECAL electromagnetic calorimeter xv, 40, 43–45, 54–56, 61, 65, 206–208

EE electromagnetic endcap 43, 60, 207

ES electromagnetic pre-shower xv, 44

EW electroweak xii, 6, 13, 15, 16, 20, 25, 29, 46, 48, 120, 121

EWSB electroweak symmetry breaking 15–18

FPGA field-programmable gate array 48

FSR final state radiation 68, 69, 96, 127, 142

GMSB gauge-mediated symmetry breaking xvii, xviii, 34, 95, 96, 113, 114

GOF goodness-of-fit xx, 90, 158, 161, 164, 166

GPU graphical processing unit 108, 109

GRL gradient reversal layer xvi, 84, 85

GSW Glashow-Salam-Weinberg 13, 14, 20

GT global trigger 47, 48

GUT Grand Unified Theory 20

HB hadronic barrel 45

HCAL hadronic calorimeter xv, 40, 44–46, 55, 56, 61, 103, 204, 206–208

HE hadronic endcap 45

HEP high energy physics 20, 32, 72, 83

HF hadronic forward 45

HLT high-level trigger 47, 48, 52, 129

HNL heavy neutral lepton xii, xiv, xv, xviii–xx, 2, 3, 26–30, 35, 36, 51, 58, 60, 62, 63, 70, 106, 116, 118, 120–128, 130–135, 137–140, 142–145, 147, 148, 164, 166–168, 171, 194

IP interaction point xv, 40–42, 47, 52, 54, 113, 124, 206

ISR initial state radiation 68, 97, 112, 127

JEC jet energy correction xvi, 62–65, 107, 132, 144

JER jet energy resolution 62, 138, 163, 167

JES jet energy scale 138, 162, 163, 167

JSD Jensen-Shannon divergence xvii, 76, 111, 112, 117

KL Kullback-Leibler 74, 76

KS Kolmogorov-Smirnov 79

L1 Level-1 47, 48

LEP Large Electron-Positron Collider 28

LFV lepton flavour violation 122

LHC Large Hadron Collider xv, 2, 11–13, 24, 28, 31, 35, 36, 38–40, 46, 48, 49, 52, 70, 84, 94, 120, 168, 171

LHCb LHCb 11, 35, 166

LLP long-lived particle xvii–xix, 2, 3, 20, 31–33, 35, 36, 51, 53, 95, 97, 99–101, 105, 106, 110, 112–114, 116–118, 131, 142, 144, 145

LNV lepton number violation 22, 122

LO leading-order 7, 70, 71, 95, 122–124, 164

LSP lightest supersymmetric particle xiv, 24, 25, 33, 34, 99, 107

MB muon barrel xv, 47

MC Monte Carlo xvi, xvii, xix, xx, 12, 54, 62, 67–70, 92, 99, 106, 108, 112, 117, 139, 141, 147, 149, 158, 163

ME matrix element 7, 31, 32, 68–72, 123, 142

MI mutual information xiii, xx, 75, 152–155

ML machine learning 2, 61, 67, 70, 72, 75, 83, 94

MSSM Minimal Supersymmetric Standard Model 24

MVA multivariate xii, 51, 54, 58–60, 67, 70–72, 80, 122, 129, 132, 171

NC neutral-current 15, 120, 121, 124

NLO next-to-leading-order 7, 70, 71, 123

NLSP next-to-lightest supersymmetric particle 33

NN neural network xvi, xvii, xix, 3, 12, 67, 73–77, 79–82, 84, 85, 100, 106–110, 145

NNPDF neural network parton distribution function xiv, 12, 13

NP nuisance parameter 86–88, 92, 162

OS opposite-sign xviii–xxi, 121, 130, 133, 134, 138, 141, 147, 149, 154, 156, 199, 200

OSSF opposite-sign same-flavour xxi, 127, 195–198, 201

PDF parton distribution function xiv, 12, 13, 163

PF Particle Flow xii, xiii, xv, xvi, xix, 2, 40, 51, 54–65, 96–98, 101–105, 129, 131, 144, 145, 203–208

POI parameter of interest 86–88, 92

pp proton-proton xv, 3, 12, 38, 40, 47–49, 52, 54, 61, 65, 67, 68, 70, 83, 84, 95, 101, 108, 112, 116, 124, 128, 171

PS parton shower xvi, 67, 69–71

PSB Proton Synchrotron Booster 38

PV primary vertex 53, 54, 57, 58, 61, 102, 103, 129, 203–205, 207

QCD quantum chromodynamics 10–12, 14, 34, 47, 69, 71, 72, 95, 96, 100, 107, 108, 134, 135, 139, 140, 153

QED quantum electrodynamics 9, 10, 14

QFT quantum field theory 5–7, 9

ReLU rectified linear unit xvi, 79–81, 104

ROC receiver operating characteristic xvii, xix, xx, 112, 113, 140, 147, 148

RPC resistive plate chamber xv, 46, 47, 55

RPV R-parity violating xvii, xviii, 35, 95, 96, 113, 114

SGD stochastic gradient descent xvi, 77, 78, 81, 108

SHiP SHiP 28

SM Standard Model xii–xiv, xix, xx, 2, 5, 6, 8, 9, 15, 17, 18, 20–24, 26, 28, 29, 31–34, 36, 45, 47, 67, 70, 89, 96, 97, 100, 110, 122, 127, 130, 133–136, 138, 140, 147, 148

SPS Super Proton Synchrotron 38

SR signal region xiii, xix, xx, 120, 132, 135, 136, 138–140, 149, 152, 154, 156, 157, 164–168

SS same-sign xviii–xxi, 121, 130, 133, 134, 138, 141, 147, 149, 154, 195–201

SUSY supersymmetry xiv, xvi–xviii, 2, 3, 24, 33–35, 51, 70, 95, 96, 102, 107, 112–116, 147

SV secondary vertex xii, 53, 101–105, 114, 118, 128, 162, 203–205

TEC tracker endcap 42

TIB tracker inner barrel 42

TID tracker inner disk 42

TOB tracker outer barrel 42

TP trigger primitive 47

UE underlying event xvi, 68, 69, 97, 100

UL upper limit xiv, 3, 21, 30, 86, 88, 89, 164, 166, 168, 171

VR validation region xx, 152, 157–161

WIMP weakly interacting massive particle 24

Bibliography

[1] A. M. Sirunyan *et al.*, "A deep neural network to search for new long-lived particles decaying to jets," *Mach. Learn. Sci. Tech.*, vol. 1, p. 035 012, 2020. DOI: `10.1088/2632-2153/ab9023`. arXiv: `1912.12238 [hep-ex]`.

[2] M. Thomson, *Modern Particle Physics.* Cambridge University Press, 2013. DOI: `10.1017/CBO9781139525367`.

[3] P. A. Zyla *et al.*, "Review of Particle Physics," *PTEP*, vol. 2020, no. 8, p. 083C01, 2020. DOI: `10.1093/ptep/ptaa104`.

[4] G. 't Hooft and M. J. G. Veltman, "Regularization and Renormalization of Gauge Fields," *Nucl. Phys. B*, vol. 44, pp. 189–213, 1972. DOI: `10.1016/0550-3213(72)90279-9`.

[5] E. Noether, "Invariante variationsprobleme," ger, *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, vol. 1918, pp. 235–257, 1918. [Online]. Available: `http://eudml.org/doc/59024`.

[6] K. G. Wilson, "Confinement of Quarks," *Phys. Rev. D*, vol. 10, J. C. Taylor, Ed., pp. 2445–2459, 1974. DOI: `10.1103/PhysRevD.10.2445`.

[7] R. Aaij *et al.*, "Observation of New Resonances Decaying to $J/\psi K^+$ and $J/\psi\phi$," *Phys. Rev. Lett.*, vol. 127, no. 8, p. 082 001, 2021. DOI: `10.1103/PhysRevLett.127.082001`. arXiv: `2103.01803 [hep-ex]`.

[8] R. Aaij *et al.*, "Observation of a narrow pentaquark state, $P_c(4312)^+$, and of two-peak structure of the $P_c(4450)^+$," *Phys. Rev. Lett.*, vol. 122, no. 22, p. 222 001, 2019. DOI: `10.1103/PhysRevLett.122.222001`. arXiv: `1904.03947 [hep-ex]`.

[9] Y. L. Dokshitzer, "Calculation of the Structure Functions for Deep Inelastic Scattering and e^+e^- Annihilation by Perturbation Theory in Quantum Chromodynamics.," *Sov. Phys. JETP*, vol. 46, pp. 641–653, 1977.

[10] V. N. Gribov and L. N. Lipatov, "Deep inelastic ep scattering in perturbation theory," *Sov. J. Nucl. Phys.*, vol. 15, pp. 438–450, 1972.

[11] G. Altarelli and G. Parisi, "Asymptotic Freedom in Parton Language," *Nucl. Phys. B*, vol. 126, pp. 298–318, 1977. DOI: `10.1016/0550-3213(77)90384-4`.

[12] R. D. Ball *et al.*, “Parton distributions for the LHC Run II,” *JHEP*, vol. 04, p. 040, 2015. DOI: `10.1007/JHEP04(2015)040`. arXiv: `1410.8849 [hep-ph]`.

[13] R. D. Ball *et al.*, “Parton distributions from high-precision collider data,” *Eur. Phys. J. C*, vol. 77, no. 10, p. 663, 2017. DOI: `10.1140/epjc/s10052-017-5199-5`. arXiv: `1706.00428 [hep-ph]`.

[14] S. L. Glashow, “The renormalizability of vector meson interactions,” *Nucl. Phys.*, vol. 10, pp. 107–117, 1959. DOI: `10.1016/0029-5582(59)90196-8`.

[15] A. Salam, “Weak and Electromagnetic Interactions,” *Conf. Proc. C*, vol. 680519, pp. 367–377, 1968. DOI: `10.1142/9789812795915_0034`.

[16] S. Weinberg, “A Model of Leptons,” *Phys. Rev. Lett.*, vol. 19, pp. 1264–1266, 1967. DOI: `10.1103/PhysRevLett.19.1264`.

[17] C. S. Wu, E. Ambler, R. W. Hayward, *et al.*, “Experimental Test of Parity Conservation in β Decay,” *Phys. Rev.*, vol. 105, pp. 1413–1414, 1957. DOI: `10.1103/PhysRev.105.1413`.

[18] M. A. Bouchiat and C. Bouchiat, “1. Parity Violation Induced by Weak Neutral Currents in Atomic Physics,” *J. Phys. (France)*, vol. 35, pp. 899–927, 1974. DOI: `10.1051/jphys:019740035012089900`.

[19] P. J. Mohr, D. B. Newell, and B. N. Taylor, “CODATA Recommended Values of the Fundamental Physical Constants: 2014,” *Rev. Mod. Phys.*, vol. 88, no. 3, p. 035 009, 2016. DOI: `10.1103/RevModPhys.88.035009`. arXiv: `1507.07956 [physics.atom-ph]`.

[20] N. Aghanim *et al.*, “Planck 2018 results. VI. Cosmological parameters,” *Astron. Astrophys.*, vol. 641, A6, 2020, [Erratum: Astron.Astrophys. 652, C4 (2021)]. DOI: `10.1051/0004-6361/201833910`. arXiv: `1807.06209 [astro-ph.CO]`.

[21] M. Aker *et al.*, “Improved Upper Limit on the Neutrino Mass from a Direct Kinematic Method by KATRIN,” *Phys. Rev. Lett.*, vol. 123, no. 22, p. 221 802, 2019. DOI: `10.1103/PhysRevLett.123.221802`. arXiv: `1909.06048 [hep-ex]`.

[22] F. T. Avignone, S. R. Elliott, and J. Engel, “Double beta decay, Majorana neutrinos, and neutrino mass,” *Rev. Mod. Phys.*, vol. 80, pp. 481–516, 2 Apr. 2008. DOI: `10.1103/RevModPhys.80.481`.

[23] V. C. Rubin, N. Thonnard, and W. K. Ford Jr., “Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 (R = 4 kpc) to UGC 2885 (R = 122 kpc),” *Astrophys. J.*, vol. 238, p. 471, 1980. DOI: `10.1086/158003`.

[24] V. Trimble, “Existence and Nature of Dark Matter in the Universe,” *Ann. Rev. Astron. Astrophys.*, vol. 25, pp. 425–472, 1987. DOI: `10.1146/annurev.aa.25.090187.002233`.

[25] P. Natarajan *et al.*, “Mapping substructure in the HST Frontier Fields cluster lenses and in cosmological simulations,” *Mon. Not. Roy. Astron. Soc.*, vol. 468, no. 2, pp. 1962–1980, 2017. DOI: `10.1093/mnras/stw3385`. arXiv: `1702.04348 [astro-ph.GA]`.

[26] A. D. Sakharov, “Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe,” *Pisma Zh. Eksp. Teor. Fiz.*, vol. 5, pp. 32–35, 1967. DOI: `10.1070/PU1991v034n05ABEH002497`.

[27] P. Huet and E. Sather, “Electroweak baryogenesis and standard model CP violation,” *Phys. Rev. D*, vol. 51, pp. 379–394, 1995. DOI: `10.1103/PhysRevD.51.379`. arXiv: `hep-ph/9404302`.

[28] Y. A. Golfand and E. P. Likhtman, “Extension of the Algebra of Poincare Group Generators and Violation of p Invariance,” *JETP Lett.*, vol. 13, pp. 323–326, 1971.

[29] D. V. Volkov and V. P. Akulov, “Is the Neutrino a Goldstone Particle?” *Phys. Lett. B*, vol. 46, pp. 109–110, 1973. DOI: `10.1016/0370-2693(73)90490-5`.

[30] J. Wess and B. Zumino, “Supergauge Transformations in Four-Dimensions,” *Nucl. Phys. B*, vol. 70, A. Salam and E. Sezgin, Eds., pp. 39–50, 1974. DOI: `10.1016/0550-3213(74)90355-1`.

[31] S. P. Martin, “A Supersymmetry primer,” *Adv. Ser. Direct. High Energy Phys.*, vol. 18, G. L. Kane, Ed., pp. 1–98, 1998. DOI: `10.1142/9789812839657_0001`. arXiv: `hep-ph/9709356`.

[32] P. Fayet, “Supersymmetry and Weak, Electromagnetic and Strong Interactions,” *Phys. Lett. B*, vol. 64, p. 159, 1976. DOI: `10.1016/0370-2693(76)90319-1`.

[33] E. Witten, “Constraints on Supersymmetry Breaking,” *Nucl. Phys. B*, vol. 202, p. 253, 1982. DOI: `10.1016/0550-3213(82)90071-2`.

[34] S. Dimopoulos and H. Georgi, “Softly Broken Supersymmetry and SU(5),” *Nucl. Phys. B*, vol. 193, pp. 150–162, 1981. DOI: `10.1016/0550-3213(81)90522-8`.

[35] G. Jungman, M. Kamionkowski, and K. Griest, “Supersymmetric dark matter,” *Phys. Rept.*, vol. 267, pp. 195–373, 1996. DOI: `10.1016/0370-1573(95)00058-5`. arXiv: `hep-ph/9506380`.

[36] S. Dimopoulos, S. Raby, and F. Wilczek, “Supersymmetry and the Scale of Unification,” *Phys. Rev. D*, vol. 24, pp. 1681–1683, 1981. DOI: `10.1103/PhysRevD.24.1681`.

[37] The CMS Collaboration, “CMS Supersymmetry summary plots,” [Online]. Available: `https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsSUS`.

[38] S. Weinberg, “Baryon and Lepton Nonconserving Processes,” *Phys. Rev. Lett.*, vol. 43, pp. 1566–1570, 1979. DOI: `10.1103/PhysRevLett.43.1566`.

[39] T. Asaka and M. Shaposhnikov, “The νMSM, dark matter and baryon asymmetry of the universe,” *Physics Letters B*, vol. 620, no. 1, pp. 17–26, 2005, ISSN: 0370-2693. DOI: `https://doi.org/10.1016/j.physletb.2005.06.020`.

[40] D. Gorbunov and M. Shaposhnikov, “How to find neutral leptons of the νMSM?” *JHEP*, vol. 10, p. 015, 2007, [Erratum: JHEP 11, 101 (2013)]. DOI: `10.1088/1126-6708/2007/10/015`. arXiv: `0705.1729 [hep-ph]`.

[41] P. F. Smith, “Proposed experiments to detect keV range sterile neutrinos using energy-momentum reconstruction of beta decay or K-capture events,” *New J. Phys.*, vol. 21, no. 5, p. 053 022, 2019. DOI: `10.1088/1367-2630/ab1502`. arXiv: `1607.06876 [physics.ins-det]`.

[42] A. Pilaftsis and T. E. J. Underwood, “Resonant leptogenesis,” *Nucl. Phys. B*, vol. 692, pp. 303–345, 2004. DOI: `10.1016/j.nuclphysb.2004.05.029`. arXiv: `hep-ph/0309342`.

[43] A. Boyarsky, O. Ruchayskiy, and M. Shaposhnikov, “The Role of sterile neutrinos in cosmology and astrophysics,” *Ann. Rev. Nucl. Part. Sci.*, vol. 59, pp. 191–214, 2009. DOI: `10.1146/annurev.nucl.010909.083654`. arXiv: `0901.0011 [hep-ph]`.

[44] A. Boyarsky, M. Ovchynnikov, O. Ruchayskiy, *et al.*, “Improved big bang nucleosynthesis constraints on heavy neutral leptons,” *Phys. Rev. D*, vol. 104, no. 2, p. 023 517, 2021. DOI: `10.1103/PhysRevD.104.023517`. arXiv: `2008.00749 [hep-ph]`.

[45] P. S. B. Dev and A. Pilaftsis, “Minimal Radiative Neutrino Mass Mechanism for Inverse Seesaw Models,” *Phys. Rev. D*, vol. 86, p. 113 001, 2012. DOI: `10.1103/PhysRevD.86.113001`. arXiv: `1209.4051 [hep-ph]`.

[46] F. F. Deppisch, P. S. Bhupal Dev, and A. Pilaftsis, “Neutrinos and Collider Physics,” *New J. Phys.*, vol. 17, no. 7, p. 075 019, 2015. DOI: `10.1088/1367-2630/17/7/075019`. arXiv: `1502.06541 [hep-ph]`.

[47] G. Aad *et al.*, “Search for heavy neutral leptons in decays of W bosons produced in 13 TeV pp collisions using prompt and displaced signatures with the ATLAS detector,” *JHEP*, vol. 10, p. 265, 2019. DOI: `10.1007/JHEP10(2019)265`. arXiv: `1905.09787 [hep-ex]`.

[48] A. M. Sirunyan *et al.*, “Search for heavy neutral leptons in events with three charged leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV,” *Phys. Rev. Lett.*, vol. 120, no. 22, p. 221 801, 2018. DOI: `10.1103/PhysRevLett.120.221801`. arXiv: `1802.02965 [hep-ex]`.

[49] A. M. Sirunyan *et al.*, “Search for heavy Majorana neutrinos in same-sign dilepton channels in proton-proton collisions at $\sqrt{s} = 13$ TeV,” *JHEP*, vol. 01, p. 122, 2019. DOI: `10.1007/JHEP01(2019)122`. arXiv: `1806.10905 [hep-ex]`.

[50] P. Abreu *et al.*, "Search for neutral heavy leptons produced in Z decays," *Z. Phys. C*, vol. 74, pp. 57–71, 1997, [Erratum: Z.Phys.C 75, 580 (1997)]. DOI: `10.1007/s002880050370`.

[51] D. Liventsev *et al.*, "Search for heavy neutrinos at Belle," *Phys. Rev. D*, vol. 87, no. 7, p. 071 102, 2013, [Erratum: Phys.Rev.D 95, 099903 (2017)]. DOI: `10.1103/PhysRevD.87.071102`. arXiv: `1301.1105 [hep-ex]`.

[52] F. Bergsma *et al.*, "A Search for Decays of Heavy Neutrinos," *Phys. Lett. B*, vol. 128, J. Tran Thanh Van, Ed., p. 361, 1983. DOI: `10.1016/0370-2693(83)90275-7`.

[53] S. Alekhin *et al.*, "A facility to Search for Hidden Particles at the CERN SPS: the SHiP physics case," *Rept. Prog. Phys.*, vol. 79, no. 12, p. 124 201, 2016. DOI: `10.1088/0034-4885/79/12/124201`. arXiv: `1504.04855 [hep-ph]`.

[54] E. Cortina Gil *et al.*, "Search for K^+ decays to a muon and invisible particles," *Phys. Lett. B*, vol. 816, p. 136 259, 2021. DOI: `10.1016/j.physletb.2021.136259`. arXiv: `2101.12304 [hep-ex]`.

[55] G. Bernardi *et al.*, "Further Limits on Heavy Neutrino Couplings," *Phys. Lett. B*, vol. 203, pp. 332–334, 1988. DOI: `10.1016/0370-2693(88)90563-1`.

[56] R. Acciarri *et al.*, "Long-Baseline Neutrino Facility (LBNF) and Deep Underground Neutrino Experiment (DUNE): Conceptual Design Report, Volume 2: The Physics Program for DUNE at LBNF," Dec. 2015. arXiv: `1512.06148 [physics.ins-det]`.

[57] M. Chrzaszcz, M. Drewes, T. E. Gonzalo, *et al.*, "A frequentist analysis of three right-handed neutrinos with GAMBIT," *Eur. Phys. J. C*, vol. 80, no. 6, p. 569, 2020. DOI: `10.1140/epjc/s10052-020-8073-9`. arXiv: `1908.02302 [hep-ph]`.

[58] L. Lee, C. Ohm, A. Soffer, *et al.*, "Collider Searches for Long-Lived Particles Beyond the Standard Model," *Prog. Part. Nucl. Phys.*, vol. 106, pp. 210–255, 2019. DOI: `10.1016/j.ppnp.2019.02.006`. arXiv: `1810.12602 [hep-ph]`.

[59] E. Bagnaschi *et al.*, "Likelihood Analysis of the pMSSM11 in Light of LHC 13-TeV Data," *Eur. Phys. J. C*, vol. 78, no. 3, p. 256, 2018. DOI: `10.1140/epjc/s10052-018-5697-0`. arXiv: `1710.11091 [hep-ph]`.

[60] N. Arkani-Hamed and S. Dimopoulos, "Supersymmetric unification without low energy supersymmetry and signatures for fine-tuning at the LHC," *JHEP*, vol. 06, p. 073, 2005. DOI: `10.1088/1126-6708/2005/06/073`. arXiv: `hep-th/0405159`.

[61] P. Gambino, G. F. Giudice, and P. Slavich, "Gluino decays in split supersymmetry," *Nucl. Phys. B*, vol. 726, pp. 35–52, 2005. DOI: `10.1016/j.nuclphysb.2005.08.011`. arXiv: `hep-ph/0506214`.

[62] G. R. Farrar and P. Fayet, "Phenomenology of the Production, Decay, and Detection of New Hadronic States Associated with Supersymmetry," *Phys. Lett. B*, vol. 76, pp. 575–579, 1978. DOI: `10.1016/0370-2693(78)90858-4`.

[63] The CMS Collaboration, "CMS Exotica summary plots for 13 TeV data," [Online]. Available: `https://twiki.cern.ch/twiki/bin/view/CMSPublic/SummaryPlotsEXO13TeV`.

[64] G. F. Giudice and R. Rattazzi, "Theories with gauge mediated supersymmetry breaking," *Phys. Rept.*, vol. 322, pp. 419–499, 1999. DOI: `10.1016/S0370-1573(99)00042-3`. arXiv: `hep-ph/9801271`.

[65] L. J. Hall and M. Suzuki, "Explicit R-Parity Breaking in Supersymmetric Models," *Nucl. Phys. B*, vol. 231, pp. 419–444, 1984. DOI: `10.1016/0550-3213(84)90513-3`.

[66] R. Barbier *et al.*, "R-parity violating supersymmetry," *Phys. Rept.*, vol. 420, pp. 1–202, 2005. DOI: `10.1016/j.physrep.2005.08.006`. arXiv: `hep-ph/0406039`.

[67] M. Chemtob, "Phenomenological constraints on broken R parity symmetry in supersymmetry models," *Prog. Part. Nucl. Phys.*, vol. 54, pp. 71–191, 2005. DOI: `10.1016/j.ppnp.2004.06.001`. arXiv: `hep-ph/0406029`.

[68] R. Aaij *et al.*, "Search for heavy neutral leptons in $W^+ \longrightarrow \mu^+\mu^\pm$jet decays," *Eur. Phys. J. C*, vol. 81, no. 3, p. 248, 2021. DOI: `10.1140/epjc/s10052-021-08973-5`. arXiv: `2011.05263 [hep-ex]`.

[69] A. Abada, N. Bernal, M. Losada, *et al.*, "Inclusive Displaced Vertex Searches for Heavy Neutral Leptons at the LHC," *JHEP*, vol. 01, p. 093, 2019. DOI: `10.1007/JHEP01(2019)093`. arXiv: `1807.10024 [hep-ph]`.

[70] L. Evans and P. Bryant, "LHC Machine," *JINST*, vol. 3, S08001, 2008. DOI: `10.1088/1748-0221/3/08/S08001`.

[71] S. Chatrchyan *et al.*, "The CMS Experiment at the CERN LHC," *JINST*, vol. 3, S08004, 2008. DOI: `10.1088/1748-0221/3/08/S08004`.

[72] The CMS Collaboration, "Technical proposal for the upgrade of the CMS detector through 2020," Tech. Rep. CERN-LHCC-2011-006. LHCC-P-004, Jun. 2011. [Online]. Available: `http://cds.cern.ch/record/1355706`.

[73] *The CMS magnet project: Technical Design Report*, ser. Technical design report. CMS. Geneva: CERN, 1997. [Online]. Available: `http://cds.cern.ch/record/331056`.

[74] V. Karimaki *et al.*, *The CMS tracker system project: Technical Design Report*, ser. Technical design report. CMS. Geneva: CERN, 1997. [Online]. Available: `http://cds.cern.ch/record/368412`.

[75] S. Chatrchyan *et al.*, “Description and performance of track and primary-vertex reconstruction with the CMS tracker,” *JINST*, vol. 9, no. 10, P10009, 2014. DOI: `10.1088/1748-0221/9/10/P10009`. arXiv: `1405.6569 [physics.ins-det]`.

[76] *The CMS electromagnetic calorimeter project: Technical Design Report*, ser. Technical design report. CMS. Geneva: CERN, 1997. [Online]. Available: `http://cds.cern.ch/record/349375`.

[77] P. Adzic *et al.*, “Energy resolution of the barrel of the CMS electromagnetic calorimeter,” *JINST*, vol. 2, P04004, 2007. DOI: `10.1088/1748-0221/2/04/P04004`.

[78] *The CMS hadron calorimeter project: Technical Design Report*, ser. Technical design report. CMS. Geneva: CERN, 1997. [Online]. Available: `http://cds.cern.ch/record/357153`.

[79] S. Abdullin *et al.*, “The CMS barrel calorimeter response to particle beams from 2-GeV/c to 350-GeV/c,” *Eur. Phys. J. C*, vol. 60, pp. 359–373, 2009, [Erratum: Eur.Phys.J.C 61, 353–356 (2009)]. DOI: `10.1140/epjc/s10052-009-0959-5`.

[80] A. M. Sirunyan *et al.*, “Calibration of the CMS hadron calorimeters using proton-proton collision data at $\sqrt{s} = 13\,\text{TeV}$,” *JINST*, vol. 15, no. 05, P05002, 2020. DOI: `10.1088/1748-0221/15/05/P05002`. arXiv: `1910.00079 [physics.ins-det]`.

[81] J. G. Layter, *The CMS muon project: Technical Design Report*, ser. Technical design report. CMS. Geneva: CERN, 1997. [Online]. Available: `http://cds.cern.ch/record/343814`.

[82] A. M. Sirunyan *et al.*, “Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$,” *JINST*, vol. 13, no. 06, P06015, 2018. DOI: `10.1088/1748-0221/13/06/P06015`. arXiv: `1804.04528 [physics.ins-det]`.

[83] G. L. Bayatyan *et al.*, *CMS TriDAS project: Technical Design Report, Volume 1: The Trigger Systems*, ser. Technical design report. CMS. [Online]. Available: `http://cds.cern.ch/record/706847`.

[84] S. Cittolin, A. Racz, and P. Sphicas, *CMS The TriDAS Project: Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger. CMS trigger and data-acquisition project*, ser. Technical design report. CMS. Geneva: CERN, 2002. [Online]. Available: `http://cds.cern.ch/record/578006`.

[85] V. Khachatryan *et al.*, “The CMS trigger system,” *JINST*, vol. 12, no. 01, P01020, 2017. DOI: `10.1088/1748-0221/12/01/P01020`. arXiv: `1609.02366 [physics.ins-det]`.

[86] “CMS Luminosity Measurements for the 2016 Data Taking Period,” CERN, Geneva, Tech. Rep. CMS-PAS-LUM-17-001, 2017. [Online]. Available: `https://cds.cern.ch/record/2257069`.

[87] "CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13$ TeV," CERN, Geneva, Tech. Rep. CMS-PAS-LUM-17-004, 2018. [Online]. Available: `https://cds.cern.ch/record/2621960`.

[88] "CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13$ TeV," CERN, Geneva, Tech. Rep. CMS-PAS-LUM-18-002, 2019. [Online]. Available: `https://cds.cern.ch/record/2676164`.

[89] A. M. Sirunyan *et al.*, "Particle-flow reconstruction and global event description with the CMS detector," *JINST*, vol. 12, no. 10, P10003, 2017. DOI: `10.1088/1748-0221/12/10/P10003`. arXiv: `1706.04965 [physics.ins-det]`.

[90] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," *Journal of Basic Engineering*, vol. 82, no. 1, pp. 35–45, Mar. 1960, ISSN: 0021-9223. DOI: `10.1115/1.3662552`. eprint: `https://asmedigitalcollection.asme.org/fluidsengineering/article-pdf/82/1/35/5518977/35\_1.pdf`.

[91] W. Elmetenawee, "CMS track reconstruction performance during Run 2 and developments for Run 3," *PoS*, vol. ICHEP2020, p. 733, 2021. DOI: `10.22323/1.390.0733`. arXiv: `2012.07035 [physics.ins-det]`.

[92] K. Rose, "Deterministic annealing for clustering, compression, classification, regression, and related optimization problems," *IEEE Proc.*, vol. 86, no. 11, pp. 2210–2239, 1998. DOI: `10.1109/5.726788`.

[93] R. Fruhwirth, W. Waltenberger, and P. Vanlaer, "Adaptive vertex fitting," *J. Phys. G*, vol. 34, N343, 2007. DOI: `10.1088/0954-3899/34/12/N01`.

[94] A. M. Sirunyan *et al.*, "Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC," *JINST*, vol. 16, no. 05, P05014, 2021. DOI: `10.1088/1748-0221/16/05/P05014`. arXiv: `2012.06888 [hep-ex]`.

[95] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm," *JHEP*, vol. 04, p. 063, 2008. DOI: `10.1088/1126-6708/2008/04/063`. arXiv: `0802.1189 [hep-ph]`.

[96] A. M. Sirunyan *et al.*, "Pileup mitigation at CMS in 13 TeV data," *JINST*, vol. 15, no. 09, P09018, 2020. DOI: `10.1088/1748-0221/15/09/P09018`. arXiv: `2003.00503 [hep-ex]`.

[97] V. Khachatryan *et al.*, "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV," *JINST*, vol. 12, no. 02, P02014, 2017. DOI: `10.1088/1748-0221/12/02/P02014`. arXiv: `1607.03663 [hep-ex]`.

[98] "Jet energy scale and resolution performances with 13 TeV data," Jun. 2016. [Online]. Available: `http://cds.cern.ch/record/2160347`.

[99] A. M. Sirunyan *et al.*, "Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$ using the CMS detector," *JINST*, vol. 14, no. 07, P07004, 2019. DOI: `10.1088/1748-0221/14/07/P07004`. arXiv: `1903.06078 [hep-ex]`.

[100] "Mitigation of anomalous missing transverse momentum measurements in data collected by CMS at $\sqrt{s} = 13\,\text{TeV}$ during the LHC Run 2," Apr. 2020. [Online]. Available: `https://cds.cern.ch/record/2714938`.

[101] A. Buckley *et al.*, "General-purpose event generators for LHC physics," *Phys. Rept.*, vol. 504, pp. 145–233, 2011. DOI: `10.1016/j.physrep.2011.03.005`. arXiv: `1101.2599 [hep-ph]`.

[102] T. Sjöstrand, S. Ask, J. R. Christiansen, *et al.*, "An introduction to PYTHIA 8.2," *Comput. Phys. Commun.*, vol. 191, pp. 159–177, 2015. DOI: `10.1016/j.cpc.2015.01.024`. arXiv: `1410.3012 [hep-ph]`.

[103] R. K. Ellis, W. J. Stirling, and B. R. Webber, *QCD and collider physics.* Cambridge University Press, Feb. 2011, vol. 8.

[104] S. Höche, "Introduction to parton-shower event generators," in *Theoretical Advanced Study Institute in Elementary Particle Physics: Journeys Through the Precision Frontier: Amplitudes for Colliders*, Nov. 2014. DOI: `10.1142/9789814678766_0005`. arXiv: `1411.4085 [hep-ph]`.

[105] V. Khachatryan *et al.*, "Event generator tunes obtained from underlying event and multiparton scattering measurements," *Eur. Phys. J. C*, vol. 76, no. 3, p. 155, 2016. DOI: `10.1140/epjc/s10052-016-3988-x`. arXiv: `1512.00815 [hep-ex]`.

[106] B. Andersson, G. Gustafson, G. Ingelman, *et al.*, "Parton Fragmentation and String Dynamics," *Phys. Rept.*, vol. 97, pp. 31–145, 1983. DOI: `10.1016/0370-1573(83)90080-7`.

[107] J. Alwall *et al.*, "Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions," *Eur. Phys. J. C*, vol. 53, pp. 473–500, 2008. DOI: `10.1140/epjc/s10052-007-0490-5`. arXiv: `0706.2569 [hep-ph]`.

[108] S. Agostinelli *et al.*, "GEANT4–a simulation toolkit," *Nucl. Instrum. Meth. A*, vol. 506, pp. 250–303, 2003. DOI: `10.1016/S0168-9002(03)01368-8`.

[109] R. Frederix and S. Frixione, "Merging meets matching in MC@NLO," *JHEP*, vol. 12, p. 061, 2012. DOI: `10.1007/JHEP12(2012)061`. arXiv: `1209.6215 [hep-ph]`.

[110] S. Frixione, P. Nason, and G. Ridolfi, "A Positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction," *JHEP*, vol. 09, p. 126, 2007. DOI: `10.1088/1126-6708/2007/09/126`. arXiv: `0707.3088 [hep-ph]`.

[111] S. Alioli, P. Nason, C. Oleari, *et al.*, "NLO single-top production matched with shower in POWHEG: s- and t-channel contributions," *JHEP*, vol. 09, p. 111, 2009, [Erratum: JHEP 02, 011 (2010)]. DOI: `10.1088/1126-6708/2009/09/111`. arXiv: `0907.4076 [hep-ph]`.

[112] E. Re, "Single-top Wt-channel production matched with parton showers using the POWHEG method," *Eur. Phys. J. C*, vol. 71, p. 1547, 2011. DOI: `10.1140/epjc/s10052-011-1547-z`. arXiv: `1009.2450 [hep-ph]`.

[113] K. Hornik, "Approximation capabilities of multilayer feedforward networks," *Neural Networks*, vol. 4, no. 2, pp. 251–257, 1991, ISSN: 0893-6080. DOI: `https://doi.org/10.1016/0893-6080(91)90009-T`.

[114] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948. DOI: `10.1002/j.1538-7305.1948.tb01338.x`.

[115] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2017. arXiv: `1412.6980 [cs.LG]`.

[116] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986. DOI: `0.1038/323533a0`.

[117] Y. A. LeCun, L. Bottou, G. B. Orr, *et al.*, "Efficient backprop," in *Neural Networks: Tricks of the Trade: Second Edition.* Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 9–48, ISBN: 978-3-642-35289-8. DOI: `10.1007/978-3-642-35289-8_3`.

[118] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, JMLR Workshop and Conference Proceedings, 2010, pp. 249–256. [Online]. Available: `https://proceedings.mlr.press/v9/glorot10a/glorot10a.pdf`.

[119] K. He, X. Zhang, S. Ren, *et al.*, *Delving deep into rectifiers: Surpassing human-level performance on imagenet classification*, 2015. arXiv: `1502.01852 [cs.CV]`.

[120] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, G. Gordon, D. Dunson, and M. Dudík, Eds., ser. Proceedings of Machine Learning Research, vol. 15, Fort Lauderdale, FL, USA: PMLR, Apr. 2011, pp. 315–323. [Online]. Available: `http://proceedings.mlr.press/v15/glorot11a.html`.

[121] F. J. M. Jr., "The Kolmogorov-Smirnov Test for Goodness of Fit," *Journal of the American Statistical Association*, vol. 46, no. 253, pp. 68–78, 1951. DOI: `10.1080/01621459.1951.10500769`.

[122] D. Wolpert and W. Macready, "No free lunch theorems for optimization," *IEEE Transactions on Evolutionary Computation*, vol. 1, no. 1, pp. 67–82, 1997. DOI: `10.1109/4235.585893`.

[123] N. Srivastava, G. Hinton, A. Krizhevsky, *et al.*, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, no. 56, pp. 1929–1958, 2014. [Online]. Available: `http://jmlr.org/papers/v15/srivastava14a.html`.

[124] G. Hinton, L. Deng, D. Yu, *et al.*, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 82–97, 2012. DOI: `10.1109/MSP.2012.2205597`.

[125] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017, ISSN: 0001-0782. DOI: `10.1145/3065386`.

[126] T. Mikolov, K. Chen, G. Corrado, *et al.*, "Efficient estimation of word representations in vector space," in *1st International Conference on Learning Representations, ICLR 2013, Scottsdale, Arizona, USA, May 2-4, 2013, Workshop Track Proceedings*, Y. Bengio and Y. LeCun, Eds., 2013. arXiv: `1301.3781`.

[127] Z. Lu, H. Pu, F. Wang, *et al.*, *The expressive power of neural networks: A view from the width*, 2017. arXiv: `1709.02540 [cs.LG]`.

[128] L. Breiman, J. Friedman, C. J. Stone, *et al.*, *Classification and regression trees.* CRC press, 1984.

[129] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001. DOI: `10.1023/A:1010933404324`.

[130] Y. Freund and R. E. Schapire, "Experiments with a new boosting algorithm," in *Proceedings of the Thirteenth International Conference on International Conference on Machine Learning*, ser. ICML'96, Bari, Italy: Morgan Kaufmann Publishers Inc., 1996, pp. 148–156, ISBN: 1558604197.

[131] T. Chen and C. Guestrin, "Xgboost," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug. 2016. DOI: `10.1145/2939672.2939785`.

[132] Y. Ganin and V. Lempitsky, "Unsupervised Domain Adaptation by Backpropagation," Proceedings of Machine Learning Research, vol. 37, F. Bach and D. Blei, Eds., pp. 1180–1189, Jul. 2015. arXiv: `1409.7495`.

[133] G. N. Perdue *et al.*, “Reducing model bias in a deep learning classifier using domain adversarial neural networks in the MINERvA experiment,” *JINST*, vol. 13, no. 11, P11020, 2018. DOI: `10.1088/1748-0221/13/11/P11020`. arXiv: `1808.08332 [physics.data-an]`.

[134] J. Neyman and E. S. Pearson, “On the Problem of the Most Efficient Tests of Statistical Hypotheses,” *Phil. Trans. Roy. Soc. Lond. A*, vol. 231, no. 694-706, pp. 289–337, 1933. DOI: `10.1098/rsta.1933.0009`.

[135] L. Demortier, “P values and nuisance parameters,” in *Statistical issues for LHC physics. Proceedings, Workshop, PHYSTAT-LHC, Geneva, Switzerland, June 27-29, 2007*, 2008, p. 23. DOI: `10.5170/CERN-2008-001`. [Online]. Available: `http://cds.cern.ch/record/1099967/files/p23.pdf`.

[136] G. Cowan, K. Cranmer, E. Gross, *et al.*, “Asymptotic formulae for likelihood-based tests of new physics,” *Eur. Phys. J. C*, vol. 71, p. 1554, 2011, [Erratum: Eur.Phys.J.C 73, 2501 (2013)]. DOI: `10.1140/epjc/s10052-011-1554-0`. arXiv: `1007.1727 [physics.data-an]`.

[137] S. Baker and R. D. Cousins, “Clarification of the Use of Chi Square and Likelihood Functions in Fits to Histograms,” *Nucl. Instrum. Meth.*, vol. 221, pp. 437–442, 1984. DOI: `10.1016/0167-5087(84)90016-4`.

[138] “Procedure for the LHC Higgs boson search combination in Summer 2011,” CERN, Geneva, Tech. Rep., Aug. 2011. [Online]. Available: `https://cds.cern.ch/record/1379837`.

[139] A. L. Read, “Presentation of search results: The $\mathrm{CL_s}$ technique,” *J. Phys. G*, vol. 28, p. 2693, 2002. DOI: `10.1088/0954-3899/28/10/313`.

[140] T. Junk, “Confidence level computation for combining searches with small statistics,” *Nucl. Instrum. Meth. A*, vol. 434, p. 435, 1999. DOI: `10.1016/S0168-9002(99)00498-2`. arXiv: `hep-ex/9902006 [hep-ex]`.

[141] S. S. Wilks, “The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses,” *Annals Math. Statist.*, vol. 9, no. 1, pp. 60–62, 1938. DOI: `10.1214/aoms/1177732360`.

[142] A. Wald, “Tests of statistical hypotheses concerning several parameters when the number of observations is large,” *Transactions of the American Mathematical Society*, vol. 54, no. 3, pp. 426–482, 1943, ISSN: 00029947. [Online]. Available: `http://www.jstor.org/stable/1990256`.

[143] K. Albertsson *et al.*, "Machine learning in high energy physics community white paper," in *18th International Workshop on Advanced Computing and Analysis Techniques in Physics Research (ACAT 2017), Proceedings, Seattle, USA, August 21-25, 2017*, [J. Phys. Conf. Ser. 1085 (2018) 022008], 2018, p. 0 220 008. DOI: `10.1088/1742-6596/1085/2/022008`. arXiv: `1807.02876 [physics.comp-ph]`.

[144] E. Bols, J. Kieseler, M. Verzetti, *et al.*, "Jet Flavour Classification Using DeepJet," *JINST*, vol. 15, no. 12, P12012, 2020. DOI: `10.1088/1748-0221/15/12/P12012`. arXiv: `2008.10519 [hep-ex]`.

[145] A. M. Sirunyan *et al.*, "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV," *JINST*, vol. 13, no. 05, P05011, 2018. DOI: `10.1088/1748-0221/13/05/P05011`. arXiv: `1712.07158 [physics.ins-det]`.

[146] G. Aad *et al.*, "ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s}$ = 13 TeV," *Eur. Phys. J. C*, vol. 79, no. 11, p. 970, 2019. DOI: `10.1140/epjc/s10052-019-7450-8`. arXiv: `1907.05120 [hep-ex]`.

[147] J. Alwall, R. Frederix, S. Frixione, *et al.*, "The automated computation of tree-level and next-to-leading order differential cross-sections, and their matching to parton shower simulations," *JHEP*, vol. 07, p. 079, 2014. DOI: `10.1007/JHEP07(2014)079`. arXiv: `1405.0301 [hep-ph]`.

[148] A. C. Kraan, "Interactions of heavy stable hadronizing particles," *Eur. Phys. J. C*, vol. 37, pp. 91–104, 2004. DOI: `10.1140/epjc/s2004-01946-6`. arXiv: `hep-ex/0404001`.

[149] R. Mackeprang and A. Rizzi, "Interactions of Coloured Heavy Stable Particles in Matter," *Eur. Phys. J. C*, vol. 50, pp. 353–362, 2007. DOI: `10.1140/epjc/s10052-007-0252-4`. arXiv: `hep-ph/0612161`.

[150] M. Cacciari and G. P. Salam, "Pileup subtraction using jet areas," *Phys. Lett.*, vol. B659, pp. 119–126, 2008. DOI: `10.1016/j.physletb.2007.09.077`. arXiv: `0707.1378 [hep-ph]`.

[151] D. Bertolini, P. Harris, M. Low, *et al.*, "Pileup per particle identification," *JHEP*, vol. 10, p. 059, 2014. DOI: `10.1007/JHEP10(2014)059`. arXiv: `1407.6013 [hep-ph]`.

[152] P. Baldi, K. Cranmer, T. Faucett, *et al.*, "Parameterized neural networks for high-energy physics," *Eur. Phys. J.*, vol. C76, no. 5, p. 235, 2016. DOI: `10.1140/epjc/s10052-016-4099-4`. arXiv: `1601.07913 [hep-ex]`.

[153] M. Abadi *et al.*, *TensorFlow: Large-scale machine learning on heterogeneous systems*, 2015. arXiv: `1603.04467 [cs.DC]`.

[154] F. Chollet *et al.*, *Keras*, 2015. [Online]. Available: `https://keras.io`.

[155] M. Komm, V. Cepaitis, R. Bainbridge, *et al.*, *ROOT-based preprocessing pipeline for TensorFlow/Keras*, version v1.0, Dec. 2019. DOI: `10.5281/zenodo.3571366`. [Online]. Available: `https://doi.org/10.5281/zenodo.3571366`.

[156] K. Bondarenko, A. Boyarsky, D. Gorbunov, *et al.*, "Phenomenology of GeV-scale Heavy Neutral Leptons," *JHEP*, vol. 11, p. 032, 2018. DOI: `10.1007/JHEP11(2018)032`. arXiv: `1805.08567 [hep-ph]`.

[157] R. Ruiz, "HeavyN," [Online]. Available: `http://feynrules.irmp.ucl.ac.be/wiki/HeavyN`.

[158] J. F. Donoghue, F. E. Low, and S.-Y. Pi, "Tensor Analysis of Hadronic Jets in Quantum Chromodynamics," *Phys. Rev. D*, vol. 20, p. 2759, 1979. DOI: `10.1103/PhysRevD.20.2759`.

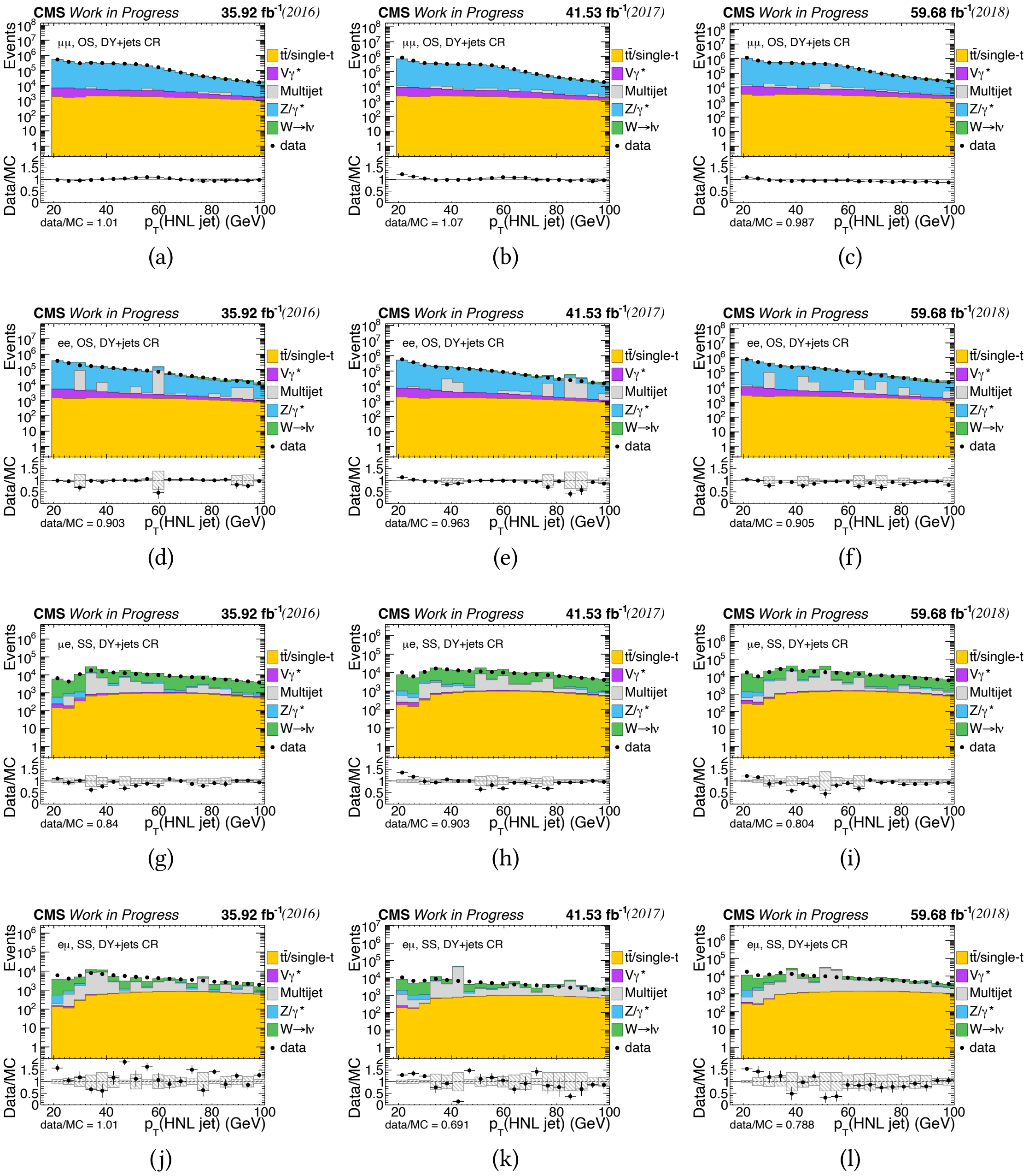
[159] E. W. Weisstein, "Legendre-Gauss Quadrature," *From MathWorld–A Wolfram Web Resource*, [Online]. Available: `https://mathworld.wolfram.com/Legendre-GaussQuadrature.html`.

[160] S. Choi and H. Oh, "Improved Extrapolation Methods of Data-driven Background Estimation in High-Energy Physics," Jun. 2019. arXiv: `1906.10831 [hep-ph]`.

Appendix A

Control region validation plots

The following set of plots shows the modelling in the CR of various key variables used in the search for HNLs, some of which are used in the training of the event-level BDT. The modelling of the BDT and the DNN jet tagger output in the same CR is also shown.



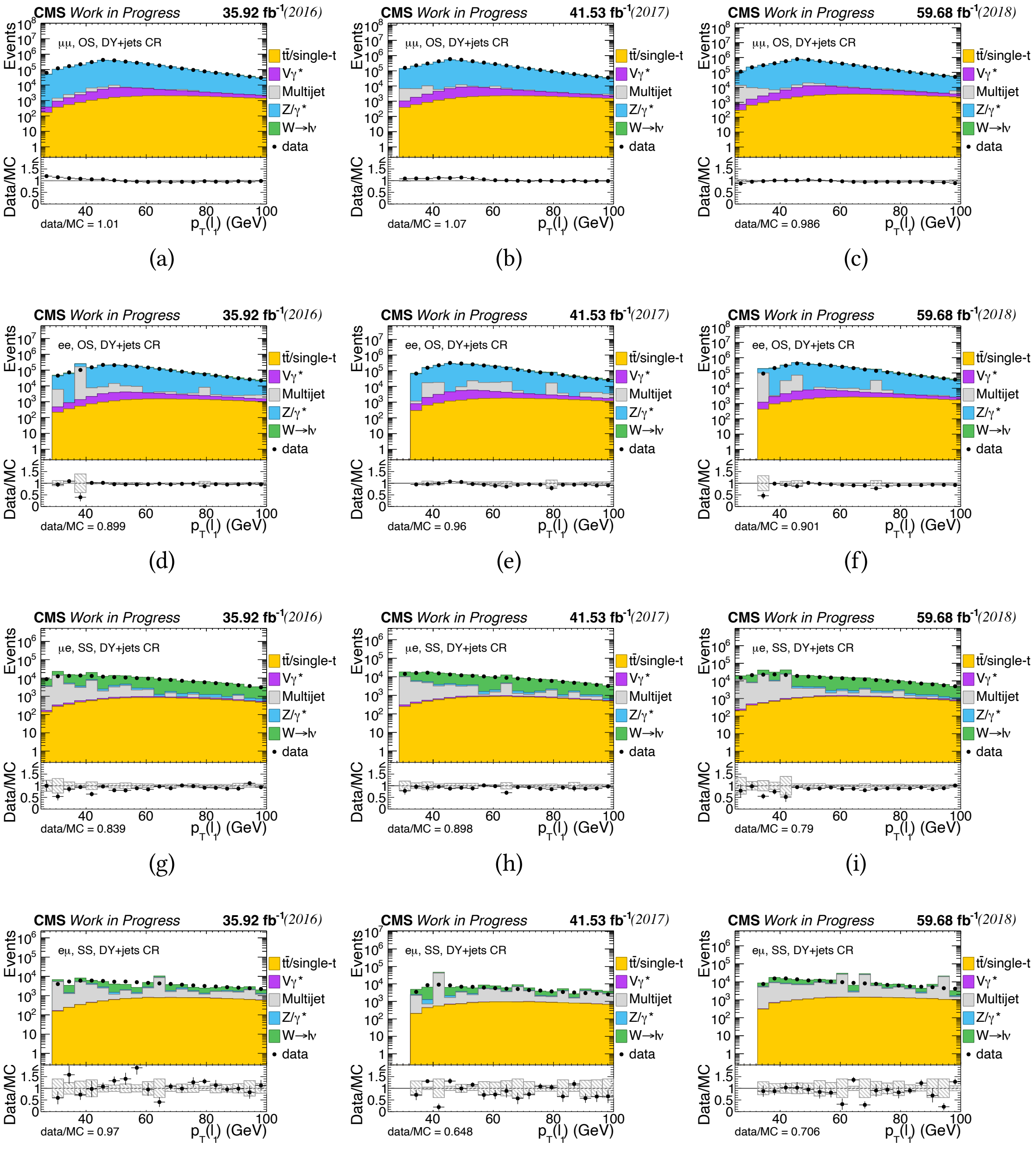
(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

(j) (k) (l)

Figure A.1: Distribution of j* p_T in the CR. The distributions are shown for 2016, 2017, and 2018 scenarios in the left, middle, and right columns, respectively. The distributions are shown for the OSSF categories in the first two rows, and the SS μe and $e\mu$ categories in the bottom two rows.



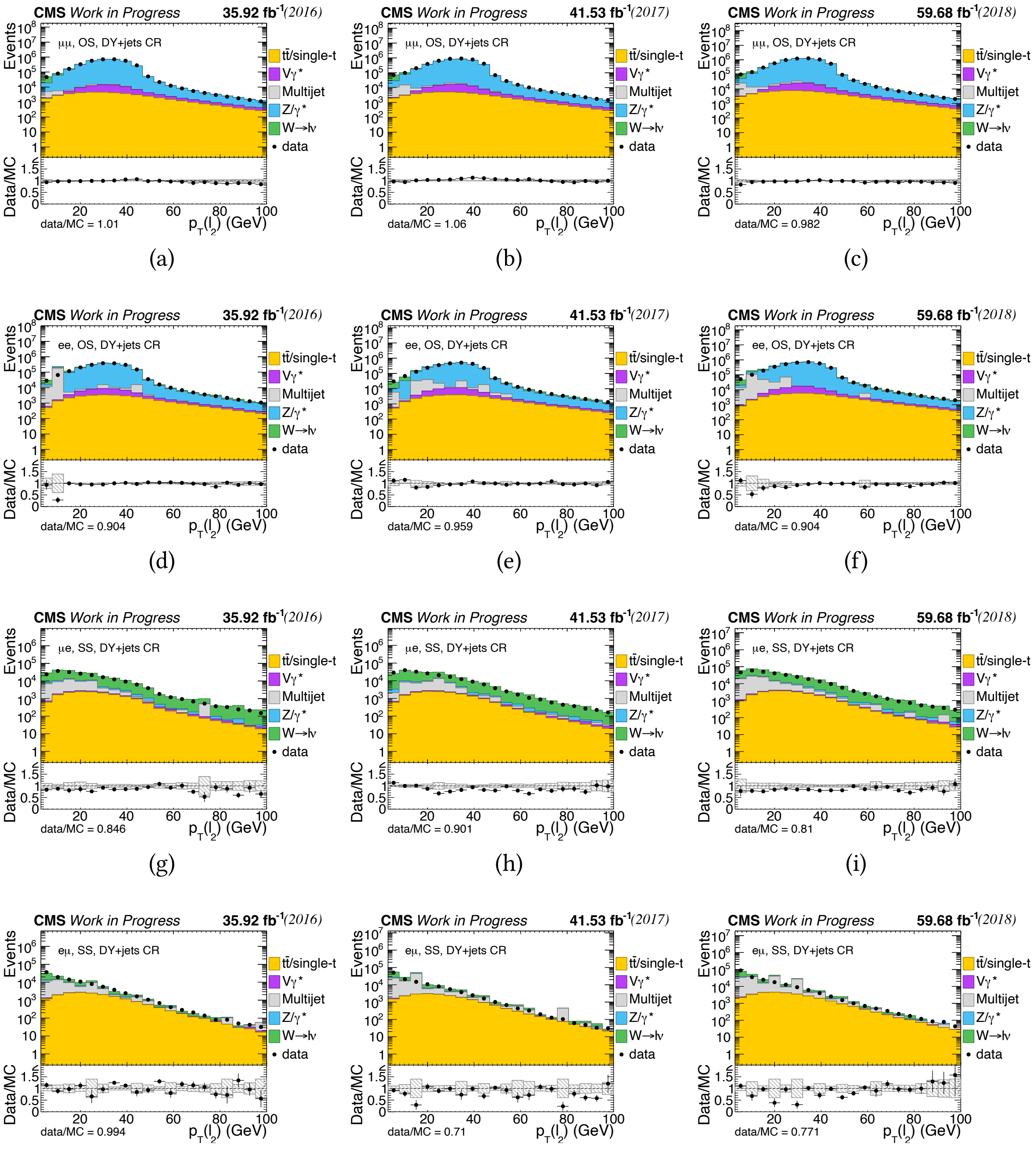
(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

(j) (k) (l)

Figure A.2: Distribution of the ℓ_1 p_T in the CR. The distributions are shown for 2016, 2017, and 2018 scenarios in the left, middle, and right columns, respectively. The distributions are shown for the OSSF categories in the first two rows, and the SS μe and $e\mu$ categories in the bottom two rows.



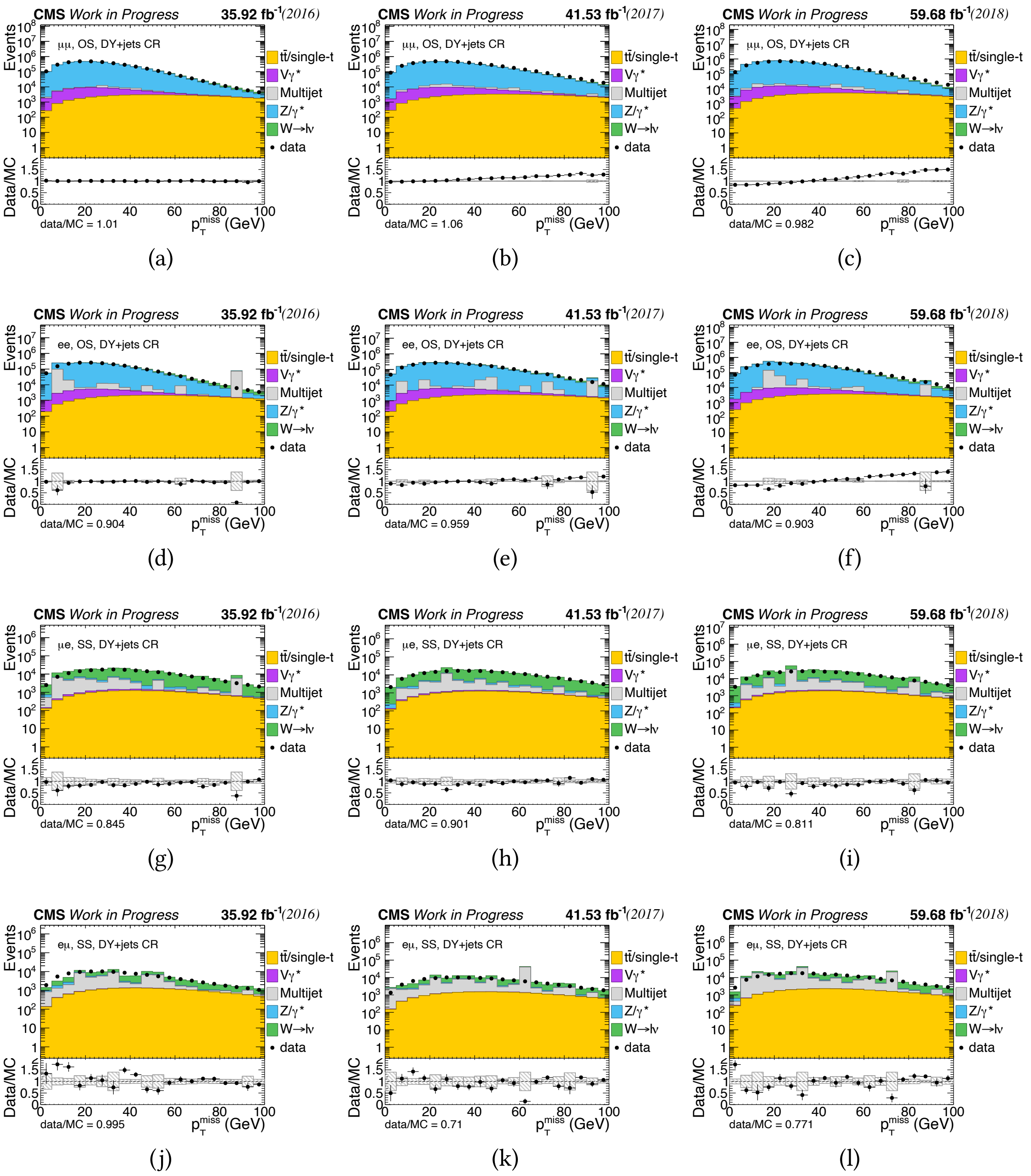
(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

(j) (k) (l)

Figure A.3: Distribution of the ℓ_2 p_T in the CR. The distributions are shown for 2016, 2017, and 2018 scenarios in the left, middle, and right columns, respectively. The distributions are shown for the OSSF categories in the first two rows, and the SS μe and $e\mu$ categories in the bottom two rows.



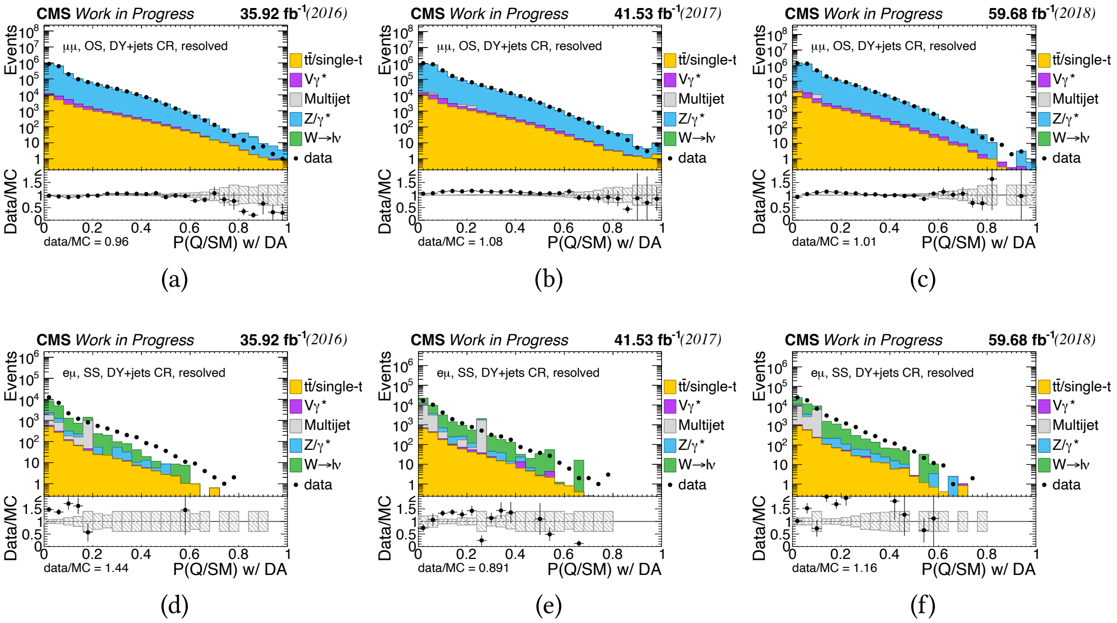
(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

(j) (k) (l)

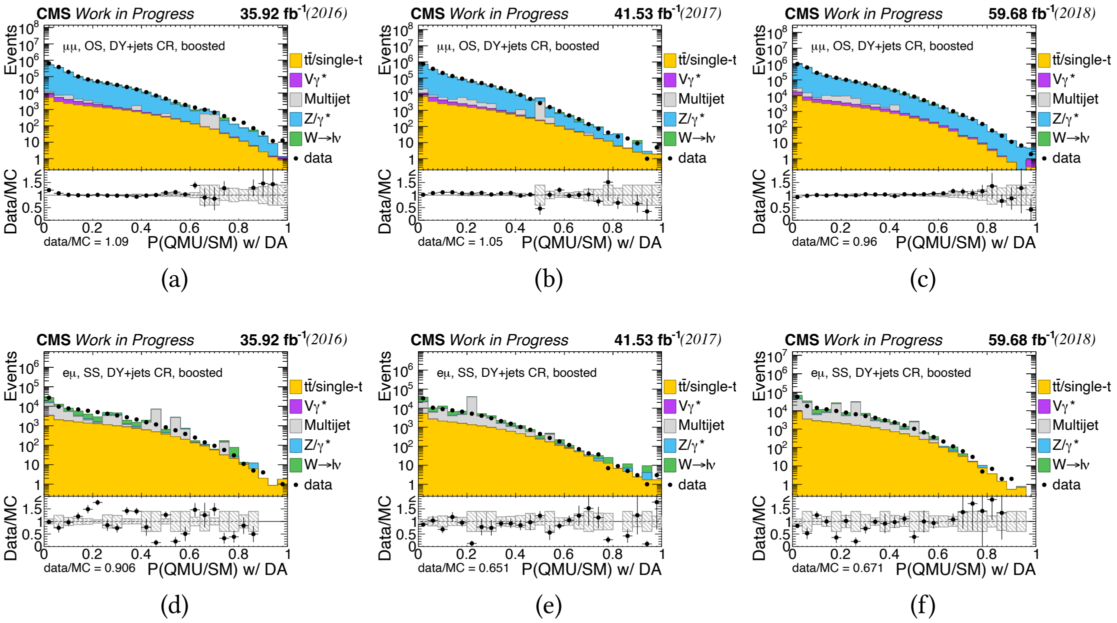
Figure A.4: Distribution of the p_T^{miss} in the CR. The distributions are shown for 2016, 2017, and 2018 scenarios in the left, middle, and right columns, respectively. The distributions are shown for the OSSF categories in the first two rows, and the SS μe and $e\mu$ categories in the bottom two rows.



(a) (b) (c)

(d) (e) (f)

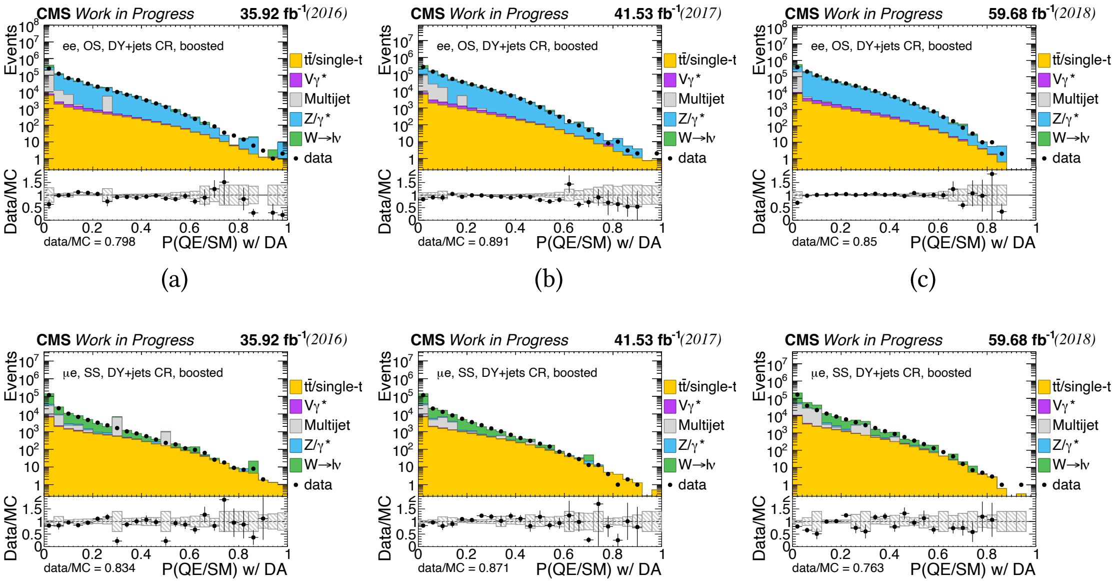
Figure A.5: Modelling of the tagger output in the CR, resolved scenario. The distributions are shown for 2016, 2017, and 2018 scenarios in the left, middle, and right columns, respectively. The distributions are shown for the OS dimuon category in the first row, and the SS *eμ* category in the bottom row.



(a) (b) (c)

(d) (e) (f)

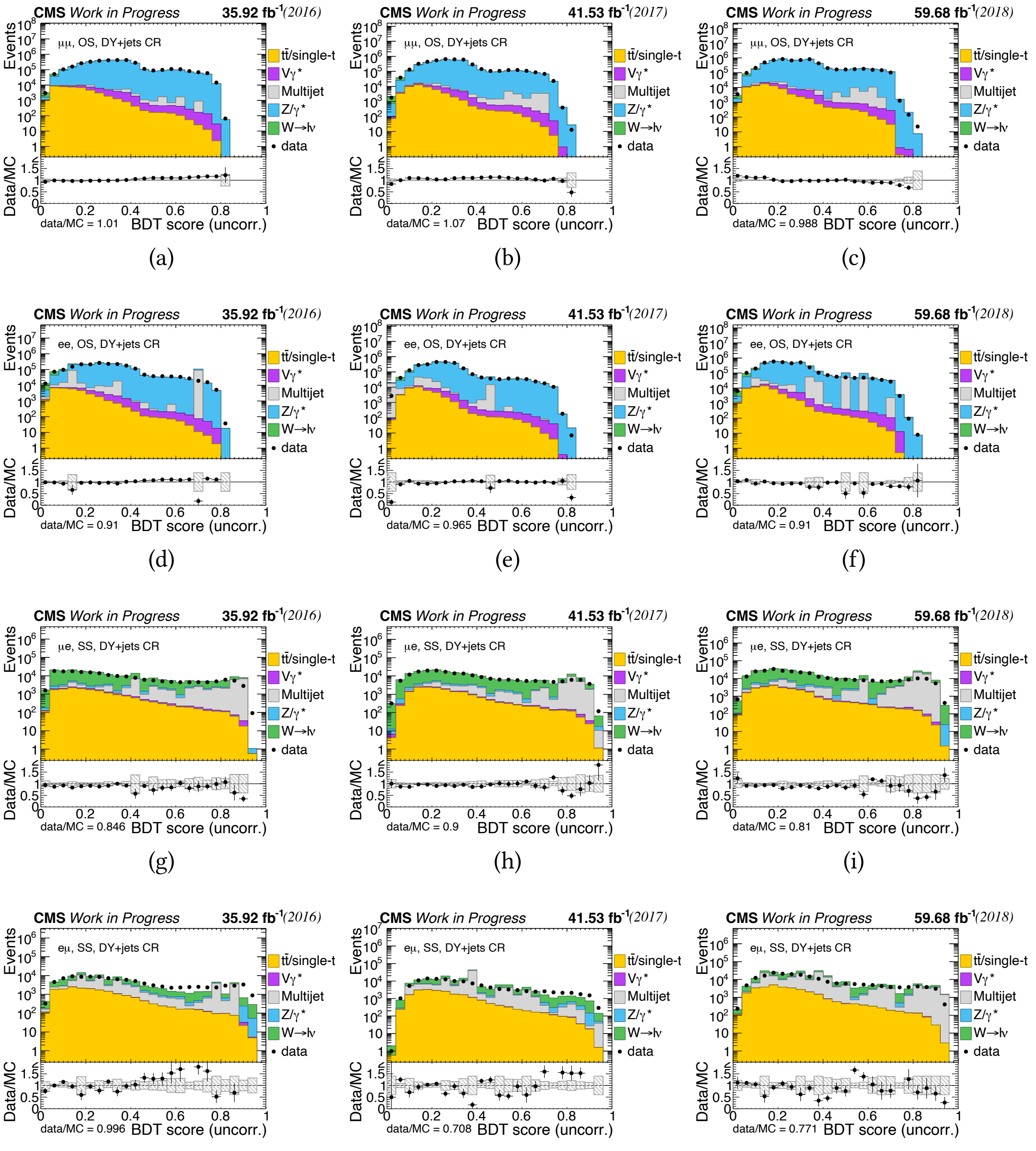
Figure A.6: Modelling of the tagger output in the CR, boosted muon scenario. The distributions are shown for 2016, 2017, and 2018 scenarios in the left, middle, and right columns, respectively. The distributions are shown for the OS dimuon category in the first row, and the SS *eμ* category in the bottom row.



(a) (b) (c)

(d) (e) (f)

Figure A.7: Modelling of the tagger output in the CR, boosted electron scenario. The distributions are shown for 2016, 2017, and 2018 scenarios in the left, middle, and right columns, respectively. The distributions are shown for the OS dielectron category in the first row, and the SS μe category in the bottom row.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

(j) (k) (l)

Figure A.8: Distribution of the BDT score in the CR. The distributions are shown for 2016, 2017, and 2018 scenarios in the left, middle, and right columns, respectively. The distributions are shown for the OSSF categories in the first two rows, and the SS μe and $e\mu$ in the bottom two rows.

Appendix B

Extended displaced jet tagger inputs

The following is the list of input features used for the training of the second iteration of the DNN jet tagging algorithm. The first version inputs are listed in Section 6.4.1.

Table B.1: Global jet features.

Internal name	Description
`global_pt`	Uncorrected p_T
`global_eta`	Pseudorapidity
`global_mass`	Mass
`global_energy`	Energy
`global_area`	Area calculated by anti-k_T algorithm
`global_beta`	Fraction of charged PF candidates from PV
`global_dR2Mean`	p_T-weighed ΔR average of charged PF candidates
`global_frac01`	Fraction of jet p_T within $\Delta R < 0.1$
`global_frac02`	Fraction of jet p_T within $\Delta R < 0.2$
`global_frac03`	Fraction of jet p_T within $\Delta R < 0.3$
`global_frac04`	Fraction of jet p_T within $\Delta R < 0.4$
`global_jetR`	Maximum relative p_T carried by one constituent
`global_jetRchg`	Maximum relative p_T carried by one charged constituent
`global_n60`	Number of constituents carrying 60% of total p_T
`global_n90`	Number of constituents carrying 90% of total p_T
`global_chargedEmEnergyFraction`	Charged electromagnetic energy fraction
`global_chargedHadronEnergyFraction`	Charged hadronic energy fraction
`global_chargedMuEnergyFraction`	Muon energy fraction
`global_electronEnergyFraction`	Electron energy fraction
`global_tau1`	1-subjettiness
`global_tau2`	2-subjettiness
`global_tau3`	3-subjettiness
`global_relMassDropMassAK`	Relative mass drop when clustering with anti-k_T
`global_relMassDropMassCA`	Relative mass drop when clustering with Cambridge-Aachen
`global_relSoftDropMassAK`	Relative soft drop when clustering with anti-k_T
`global_relSoftDropMassCA`	Relative soft drop when clustering with Cambridge-Aachen
`global_thrust`	Relative thrust of constituents in jet CM frame
`global_sphericity`	Sphericity of constituents in jet CM frame
`global_circularity`	Circularity of constituents in jet CM frame
`global_isotropy`	Isotropy of constituents in jet CM frame
`global_eventShapeC`	Event shape C of constituents in jet CM frame
`global_eventShapeD`	Event shape D of constituents in jet CM frame
`global_numberCpf`	Number of charged PF constituents
`global_numberNpf`	Number of neutral PF constituents
`global_numberSv`	Number of SVs
`global_numberMuon`	Number of muons
`global_numberElectron`	Number of electron

Table B.2: Features used to train the CSV b-tagging algorithm. These are added to the global jet features.

Internal name	Description
`csv_trackSumJetEtRatio`	Relative E_T of all tracks
`csv_trackSumJetDeltaR`	ΔR between all tracks and jet axis
`csv_vertexCategory`	Classification of PV
`csv_trackSip2dValAboveCharm`	Impact parameter d_{xy} of first track raising jet mass above 1.5 GeV
`csv_trackSip2dSigAboveCharm`	σ_{xy}/d_{xy} of first track raising jet mass above 1.5 GeV
`csv_trackSip3dValAboveCharm`	Impact parameter d_{xyz} of first track raising jet mass above 1.5 GeV
`csv_trackSip3dSigAboveCharm`	σ_{xyz}/d_{xyz} of first track raising jet mass above 1.5 GeV
`csv_jetNTracksEtaRel`	Number or tracks with trackEtaRel
`csv_jetNSelectedTracks`	Number or selected tracks

Table B.3: Features of neutral PF candidates.

Internal name	Description
`npf_ptrel`	Relative p_T
`npf_deta`	$\Delta\eta$ to jet axis
`npf_dphi`	$\Delta\phi$ to jet axis
`npf_deltaR`	ΔR to jet axis
`npf_isGamma`	Flag for passing loose photon ID
`npf_hcal_fraction`	Fraction of E_T in HCAL
`npf_drminsv`	Closest distance in ΔR to SV
`npf_puppi_weight`	PUPPI weight
`npf_relmassdrop`	Relative mass drop when removed from jet

Table B.4: Features of secondary vertices.

Internal name	Description
`sv_ptrel`	Relative p_T
`sv_deta`	$\Delta\eta$ to jet axis
`sv_dphi`	$\Delta\phi$ to jet axis
`sv_deltaR`	ΔR to jet axis
`sv_mass`	SV mass
`sv_ntracks`	Number of tracks used in fit
`sv_chi2`	χ^2 of fit
`sv_ndof`	NDOF of fit
`sv_dxy`	Impact parameter d_{xy}
`sv_dxysig`	σ_{xy}/d_{xy}
`sv_d3d`	Impact parameter d_{xyz}
`sv_d3dsig`	σ_{xyz}/d_{xyz}
`sv_costhetasvpv`	$\cos\theta$ of SV wrt. PV
`sv_enratio`	Relative E_T

Table B.5: Features of charged PF features.

Internal name	Description
`cpf_ptrel`	Relative p_T
`cpf_deta`	$\Delta\eta$ to jet axis
`cpf_dphi`	$\Delta\phi$ to jet axis
`cpf_deltaR`	ΔR to jet axis
`cpf_trackEtaRel`	η relative to jet axis
`cpf_trackPtRel`	p_T relative to jet axis
`cpf_trackPPar`	Momentum parallel to jet axis
`cpf_trackDeltaR`	ΔR
`cpf_trackPParRatio`	Relative momentum parallel to jet axis
`cpf_trackPtRatio`	Relative p_T parallel to jet p_T
`cpf_trackSip2dVal`	Impact parameter d_{xy}
`cpf_trackSip2dSig`	σ_{xy}/d_{xy}
`cpf_trackSip3dVal`	Impact parameter d_{xyz}
`cpf_trackSip3dSig`	σ_{xyz}/d_{xyz}
`cpf_trackJetDistVal`	Distance between track and jet
`cpf_trackJetDistSig`	Significance of distance between track and jet
`cpf_drminsv`	Closest distance in ΔR to SV
`cpf_vertex_association`	Flag if track is used in PV fit
`cpf_frompv`	Flag if track stems from PV
`cpf_puppi_weight`	PUPPI weight
`cpf_track_chi2`	χ^2 of track fit
`cpf_track_ndof`	NDOF of track fit
`cpf_track_quality`	Track quality flag
`cpf_track_numberOfValidPixelHits`	Number of pixel hits
`cpf_track_pixelLayersWithMeasurement`	Number of crossed pixel layers
`cpf_track_numberOfValidStripHits`	Number of strip hits
`cpf_track_stripLayersWithMeasurement`	Number of crossed strip layers
`cpf_relmassdrop`	Relative mass drop when removed from jet
`cpf_trackSip2dValSV`	Impact parameter d_{xy} wrt. SV
`cpf_trackSip2dSigSV`	σ_{xy}/d_{xy} wrt. SV
`cpf_trackSip3dValSV`	Impact parameter d_{xyz} wrt. SV
`cpf_trackSip3dSigSV`	σ_{xyz}/d_{xyz} wrt. SV
`cpf_matchedMuon`	Flag if candidate is matched to a PF muon
`cpf_matchedElectron`	Flag if candidate is matched to a PF muon
`cpf_matchedSV`	Flag if candidate is used in SV fit
`cpf_dZmin`	Minimum distance in z to another PV

Table B.6: Features of PF muons.

Internal name	Description
`muon_ptrel`	Relative p_T
`muon_deta`	$\Delta\eta$ to jet axis
`muon_dphi`	$\Delta\phi$ to jet axis
`muon_deltaR`	ΔR to jet axis
`muon_energy`	Relative energy
`muon_et`	E_T
`muon_numberOfMatchedStations`	Number of muon stations
`muon_IP2d`	Impact parameter d_{xy}
`muon_IP2dSig`	σ_{xy}/d_{xy}
`muon_IP3d`	Impact parameter d_{xyz}
`muon_IP3dSig`	σ_{xyz}/d_{xyz}
`muon_EtaRel`	Relative pseudorapidity
`muon_dxy`	Impact parameter d_{xy} of track
`muon_dxyError`	σ_{xy} of track
`muon_dxySig`	σ_{xy}/d_{xy} of track
`muon_dz`	Impact parameter d_z of track
`muon_dzError`	σ_z of track
`muon_dzSig`	σ_z/d_z of track
`muon_numberOfValidPixelHits`	Number of valid pixel hits
`muon_numberOfpixelLayersWithMeasurement`	Number of crossed pixel layers
`muon_numberOfstripLayersWithMeasurement`	Number of crossed strip layers
`muon_chi2`	χ^2 of track fit
`muon_ndof`	NDOF of track fit
`muon_caloIso`	Calorimeter isolation
`muon_ecalIso`	ECAL isolation
`muon_hcalIso`	HCAL isolation
`muon_sumPfChHadronPt`	Summed PF charged hadron p_T within $\Delta R < 0.4$
`muon_sumPfNeuHadronEt`	Summed PF neutral hadron E_T within $\Delta R < 0.4$
`muon_Pfpileup`	Summed pileup p_T within $\Delta R < 0.4$
`muon_sumPfPhotonEt`	Summed PF photon E_T within $\Delta R < 0.4$
`muon_sumPfChHadronPt03`	Summed PF charged hadron p_T within $\Delta R < 0.3$
`muon_sumPfNeuHadronEt03`	Summed PF neutral hadron E_T within $\Delta R < 0.3$
`muon_Pfpileup03`	Summed pileup p_T within $\Delta R < 0.3$
`muon_sumPfPhotonEt03`	Summed PF photon E_T within $\Delta R < 0.3$
`muon_timeAtIpInOut`	Timing at IP (in/out)
`muon_timeAtIpInOutErr`	Timing uncertainty at IP
`muon_timeAtIpOutIn`	Timing at IP (out/in)

Table B.7: Features of PF electrons (part 1).

Internal name	Description
`electron_ptrel`	Relative p_{T}
`electron_deta`	$\Delta\eta$ to jet axis
`electron_dphi`	$\Delta\phi$ to jet axis
`electron_deltaR`	ΔR to jet axis
`electron_energy`	Relative energy
`electron_EtFromCaloEn`	Calorimeter energy
`electron_isEB`	Flag if electron is in EB
`electron_isEE`	Flag if electron is in EE
`electron_ecalEnergy`	ECAL energy
`electron_isPassConversionVeto`	Flag if electron passes photon conversion veto
`electron_convDist`	Conversion distance
`electron_convFlags`	Conversion flags
`electron_convRadius`	Conversion radius
`electron_hadronicOverEm`	HCAL over ECAL energy
`electron_ecalDrivenSeed`	Flag if electron is seed from ECAL
`electron_IP2d`	Impact parameter d_{xy}
`electron_IP2dSig`	σ_{xy}/d_{xy}
`electron_IP3d`	Impact parameter d_{xyz}
`electron_IP3dSig`	σ_{xyz}/d_{xyz}
`electron_elecSC_energy`	Relative energy in ECAL supercluster
`electron_elecSC_deta`	$\Delta\eta$ between ECAL supercluster and jet axis
`electron_elecSC_dphi`	$\Delta\phi$ between ECAL supercluster and jet axis
`electron_elecSC_et`	E_{T} of ECAL supercluster
`electron_elecSC_eSuperClusterOverP`	E_{T} ECAL supercluster ratio
`electron_superClusterFbrem`	Energy associated to bremsstrahlung
`electron_eSeedClusterOverP`	Electron energy ratio of GSF track over ECAL supercluster seed
`electron_eSeedClusterOverPout`	Electron energy ratio of GSF track over ECAL supercluster seed at exit
`electron_eSuperClusterOverP`	Electron energy ratio of GSF track over ECAL supercluster
`electron_sigmaEtaEta`	$\sigma_{\eta\eta}$ ECAL supercluster shape
`electron_sigmaIetaIeta`	$\sigma_{i\eta i\eta}$ ECAL supercluster shape
`electron_sigmaIphiIphi`	$\sigma_{i\phi i\phi}$ ECAL supercluster shape
`electron_e5x5`	Energy in 5x5 ECAL cells
`electron_e5x5Rel`	Relative energy in 5x5 ECAL cells
`electron_e1x5Overe5x5`	Energy ratio of 1x5 over 5x5 ECAL cells
`electron_e2x5MaxOvere5x5`	Maximum energy ratio of 2x5 over 5x5 ECAL cells
`electron_r9`	ECAL supercluster shape variable
`electron_hcalOverEcal`	Relative HCAL over ECAL energy
`electron_hcalDepth1OverEcal`	Relative HCAL at depth 1 over ECAL energy
`electron_hcalDepth2OverEcal`	Relative HCAL at depth 2 over ECAL energy
`electron_deltaEtaEleClusterTrackAtCalo`	$\Delta\eta$ between cluster and track at calorimeter
`electron_deltaEtaSeedClusterTrackAtCalo`	$\Delta\eta$ between cluster seed and track at calorimeter
`electron_deltaPhiSeedClusterTrackAtCalo`	$\Delta\phi$ between cluster seed and track at calorimeter
`electron_deltaEtaSeedClusterTrackAtVtx`	$\Delta\eta$ between cluster seed and track at PV
`electron_deltaEtaSuperClusterTrackAtVtx`	$\Delta\phi$ between cluster seed and track at PV
`electron_deltaPhiEleClusterTrackAtCalo`	$\Delta\phi$ between cluster and track at calorimeter
`electron_deltaPhiSuperClusterTrackAtVtx`	$\Delta\phi$ between cluster and track at PV

Table B.8: Features of PF electrons (part 2).

Internal name	Description
`electron_sCseedEta`	Pseudorapidity of ECAL supercluster seed
`electron_EtaRel`	Relative η
`electron_dxy`	Impact parameter d_{xy} of track
`electron_dxyError`	σ_{xy} of track
`electron_dxySig`	σ_{xy}/d_{xy} of track
`electron_dz`	Impact parameter d_z of track
`electron_dzError`	σ_z of track
`electron_dzSig`	σ_z/d_z of track
`electron_nbOfMissingHits`	Number of missing hits
`electron_ndof`	NDOF of track
`electron_chi2`	χ^2 of track
`electron_numberOfBrems`	Number of photons from bremsstrahlung
`electron_fbrem`	Fraction of bremsstrahlung energy
`electron_neutralHadronIso`	Neutral hadron isolation within $\Delta R < 0.3$
`electron_particleIso`	Particle isolation within $\Delta R < 0.3$
`electron_photonIso`	Photon isolation within $\Delta R < 0.3$
`electron_puChargedHadronIso`	Pileup isolation from charged hadrons within $\Delta R < 0.3$
`electron_trackIso`	Track isolation within $\Delta R < 0.3$
`electron_ecalPFClusterIso`	ECAL cluster isolation within $\Delta R < 0.3$
`electron_hcalPFClusterIso`	HCAL cluster isolation within $\Delta R < 0.3$
`electron_pfSumPhotonEt`	Summed PF photon E_{T} within $\Delta R < 0.3$
`electron_pfSumChargedHadronPt`	Summed PF charged hadron p_{T} within $\Delta R < 0.3$
`electron_pfSumNeutralHadronEt`	Summed PF neutral hadron E_{T} within $\Delta R < 0.3$
`electron_pfSumPUPt`	Summed pileup p_{T} within $\Delta R < 0.3$
`electron_dr04TkSumPt`	Summed track p_{T} within $\Delta R < 0.4$
`electron_dr04EcalRecHitSumEt`	Summed ECAL E_{T} within $\Delta R < 0.4$
`electron_dr04HcalDepth1TowerSumEt`	Summed HCAL E_{T} from depth 1 within $\Delta R < 0.4$
`electron_dr04HcalDepth1TowerSumEtBc`	Summed HCAL E_{T} from depth 1 behind ECAL cluster
`electron_dr04HcalDepth2TowerSumEt`	Summed HCAL E_{T} from depth 2 within $\Delta R < 0.4$
`electron_dr04HcalDepth2TowerSumEtBc`	Summed HCAL E_{T} from depth 2 behind ECAL cluster
`electron_dr04HcalTowerSumEt`	Summed HCAL E_{T} within $\Delta R < 0.4$
`electron_dr04HcalTowerSumEtBc`	Summed HCAL E_{T} behind ECAL cluster