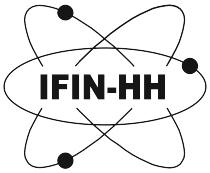
New Small Wheel Muon Detector Trigger Processor Hardware and Beyond Standard Model Searches Using Machine Learning Techniques in the ATLAS Experiment at LHC

Ioan-Mihail Dinu

UNIVERSITY OF BUCHAREST

VIRTUTE ET SAPIENTIA



Ioan-Mihail Dinu

New Small Wheel Muon Detector Trigger Processor Hardware and Beyond Standard Model Searches Using Machine Learning Techniques in the ATLAS Experiment at LHC

Thèse de doctorat soutenue le 16 novembre 2022
University of Bucharest, Faculty of Physics
Université Clermont Auvergne, École Doctorale des Sciences Fondamentales

Directeurs de thèse:

Dr. Călin Alexa	*University of Bucharest, IFIN-HH*
Dr. Julien Donini	*Université Clermont Auvergne, LPC*

Président du jury:

Dr. Theodoros Alexopoulos	*National Technical University of Athens, Greece*

Rapporteurs:

Dr. Theodoros Alexopoulos	*National Technical University of Athens, Greece*
Dr. Tommaso Dorigo	*INFN, Dipartimento di Fisica e Astronomia, Italy*

Examinateurs:

Dr. Virgil Băran	*University of Bucharest, Romania*
Dr. Anne-Catherine Le Bihan	*Institut pluridisciplinaire Hubert Curien, France*

Contents

Synthèse en Français

Le modèle standard de la physique des particules est l'aboutissement de la compréhension humaine des éléments fondamentaux de l'univers et de leur interaction. Cependant, il manque encore quelques pièces au modèle, comme les solutions au problème de la hiérarchie, et les mécanismes qui sous-tendent les masses des neutrinos ou la matière noire.

Tout cela nécessite une nouvelle physique et de nouvelles théories au-delà du modèle standard (BSM). Pour repousser ces frontières, il faut constamment améliorer à la fois les capacités expérimentales et les capacités de recherche. Depuis la découverte du boson de Higgs, les dix dernières années ont vu une croissance explosive des technologies matérielles et logicielles. Les méthodes d'apprentissage automatique (ML) ont été largement adoptées et ont donné de bons résultats dans de nombreux domaines, avec l'aide de processeurs graphiques (GPU) de plus en plus puissants. Cette recherche doctorale vise à étudier l'efficacité potentielle de l'application des méthodes de réseaux neuronaux (NN) pour apporter des améliorations dans plusieurs domaines de l' HEPP.

Du côté du détecteur, je me suis concentré sur l'évaluation de la faisabilité du système de déclenchement de l'acquisition de données au niveau du matériel pour le détecteur de muons New Small Wheel (NSW) récemment installé dans ATLAS. Le processeur de déclenchement (TP) fonctionne in-situ sur le matériel Field-Programmable Gate Array (FPGA) dans un délai de 200 ns prévu pour prendre une décision de déclenchement de niveau 0 (L0). Ces exigences sont très strictes et, étant donné que l'inférence des réseaux neuronaux est gourmande en ressources, le modèle a été spécifiquement conçu pour atténuer ces limitations inhérentes. Enfin, un modèle convolutif (CNN) a été choisi pour interpréter les réponses reçues des huit départs de détecteurs MicroMegas (MM) en étant entraîné à reproduire les variabilités liées à la probabilité d'un tir de muons provenant du point d'interaction (IP).

Le deuxième thème principal de cette thèse consiste à explorer les méthodes d'apprentissage non supervisé comme moyen de recherche agnostique des signaux BSM. Ayant accès à des données générées par des processus bien compris du modèle standard, un réseau neuronal est entraîné à comprimer les événements dans un format plus petit et à l'utiliser pour reconstruire l'événement original. Ce type de modèle est appelé un autoencodeur (AE), et une grande erreur de reconstruction d'un événement a est un indicateur potentiel que l'événement ne fait pas partie de la même distribution que les données d'apprentissage, c'est-à-dire les données SM. La combinaison d'un tel modèle avec une estimation de la

densité de probabilité basée sur un réseau de neurones donne lieu à un ensemble appelé autoencodeur probabiliste (PAE), qui est la technique centrale utilisée dans cette recherche. Le PAE a été testé en tant qu'outil potentiel pour identifier une nouvelle physique de manière indépendante du modèle dans le contexte de la physique des jets et a donné des résultats prometteurs.

Le modèle standard et le détecteur ATLAS

Le modèle standard (SM) a été établi dans le but de créer un cadre unifié pour la physique des particules. Il comprend 17 particules élémentaires dans un formalisme mathématique de théorie quantique des champs (QFT) invariant sous des transformations d'étalonnage locales. Chacune de ces symétries est associée à l'une des trois interactions fondamentales. La gravité est notamment absente de ce cadre théorique.

Malgré l'énorme succès et la précision des théories QED, QCD et *Electroweak*, le modèle standard a encore besoin de certains éléments importants pour devenir une théorie complète. Il existe des phénomènes observés et encore inexpliqués que le modèle standard n'a pas encore élucidés:

- L'une des quatre forces fondamentales connues, la gravité, n'est pas incluse dans le modèle standard. La gravité est mieux décrite par la relativité générale, qui suppose que l'espace-temps est courbe. Cependant, toutes les tentatives de modélisation de la gravité en tant que théorie quantique des champs génèrent des divergences. Pour inclure la gravité, il faudrait également trouver une particule qui sert de médiateur aux interactions gravitationnelles, ce qui n'a été identifié jusqu'à présent par aucune expérience de pointe.

- Les neutrinos sont considérés comme sans masse par le modèle standard, ce qui a été réfuté expérimentalement par des études sur les oscillations des neutrinos [1, 2]. Même si le Modèle standard est suffisamment flexible pour permettre des expansions qui fixeraient les masses des neutrinos, le mécanisme spécifique par lequel les neutrinos acquièrent une masse s'avère toujours difficile à identifier.

- En ce qui concerne l'ajustement des paramètres du modèle standard, il existe un problème de hiérarchie. Lors du calcul de la masse du boson de Higgs, les termes correctifs sont proportionnels aux échelles d'énergie respectives, il serait donc raisonnable que la masse du boson de Higgs ne soit pas éloignée de la masse de Planck. Cependant, les mesures expérimentales montrent que la masse de Higgs est inférieure à cette attente d'un ordre de 1017. L'alignement de ces deux nombres implique des annulations qui ne sont souvent pas considérées comme "naturelles".

- Il existe de fortes preuves indirectes suggérant que la matière noire fait partie de notre univers, mais le modèle standard est incapable d'en fournir une explication. La difficulté d'observer les candidats à la matière noire provient du fait qu'ils ne peuvent pas interagir de manière électromagnétique. Cela devrait encore permettre d'autres voies expérimentales de détection, mais le modèle standard n'offre aucun aperçu de la matière noire et de son comportement.

Les théories Beyond Standard Model (BSM) visent à introduire des modifications permettant de résoudre certains des problèmes SM ou à fournir des cadres entièrement nouveaux et plus généraux dans une tentative de généralisation de sa forme actuelle.

L'un de ces développements théoriques prometteurs est le Supersymmetry (SUSY) [3]. Il introduit la symétrie boson-fermion qui implique de doubler effectivement le nombre de particules dans le SM en ajoutant un *superpartener* à chaque particule existante. Les tentatives de modélisation de SUSY dans le SM tout en limitant autant que possible le nombre de particules et d'interactions se sont manifestées par les extensions suivantes : Minimal Supersymmetric Standard Model (MSSM) [4] et Next-to- Minimal Supersymmetric Standard Model (NMSSM) [5]. SUSY aborde le problème de la hiérarchie et fournit des candidats à la matière noire, mais, à l'heure où nous écrivons ces lignes, aucune preuve expérimentale définitive n'a été trouvée pour la confirmer.

De la même manière que les interactions électromagnétiques et faibles peuvent être modélisées par un cadre unifié, les modèles de type Grand Unified Theory (GUT), tentent de rassembler toutes les interactions fondamentales dans un cadre *unifié*. Ils représentent une classe entière de modèles BSM, chaque variante étant conçue pour aborder des phénomènes légèrement différents. Bien qu'on puisse dire qu'ils sont tombés en désuétude ces derniers temps, l'intérêt et le potentiel de prédictions expérimentalement testables demeurent.

Une tentative de résoudre l'incompatibilité apparente entre la SM et la relativité générale, vient de Loop Quantum Gravity (LQG). [6]. Elle tente d'intégrer la gravité par un traitement géométrique plutôt que de l'introduire comme une force fondamentale. Pour ce faire, elle décrit l'espace et le temps au niveau le plus fondamental en termes de boucles finies. LQG emprunte des idées à **théorie des cordes** mais ne nécessite pas autant de dimensions, étant généralement formulée en termes de trois ou quatre dimensions. Certaines formulations tiennent également compte de SUSY, d'autres non.

Le détecteur ATLAS

La collaboration ATLAS mène de nombreuses études couvrant un large éventail de la physique. Le programme scientifique comprend la vérification des prédictions du modèle standard (SM), ainsi que des mesures de précision de ses paramètres physiques, tels que les masses et les durées de vie moyennes des particules, et les constantes de couplage des interactions fondamentales. Les hautes énergies offertes par le LHC permettent au détecteur

ATLAS d'étudier les propriétés des particules les plus massives, telles que le quark top et le boson de Higgs. Le détecteur ATLAS a été conçu comme un dispositif polyvalent, avec pour mission d'étudier les collisions proton-proton (pp) et noyau-nucléus (A + A).

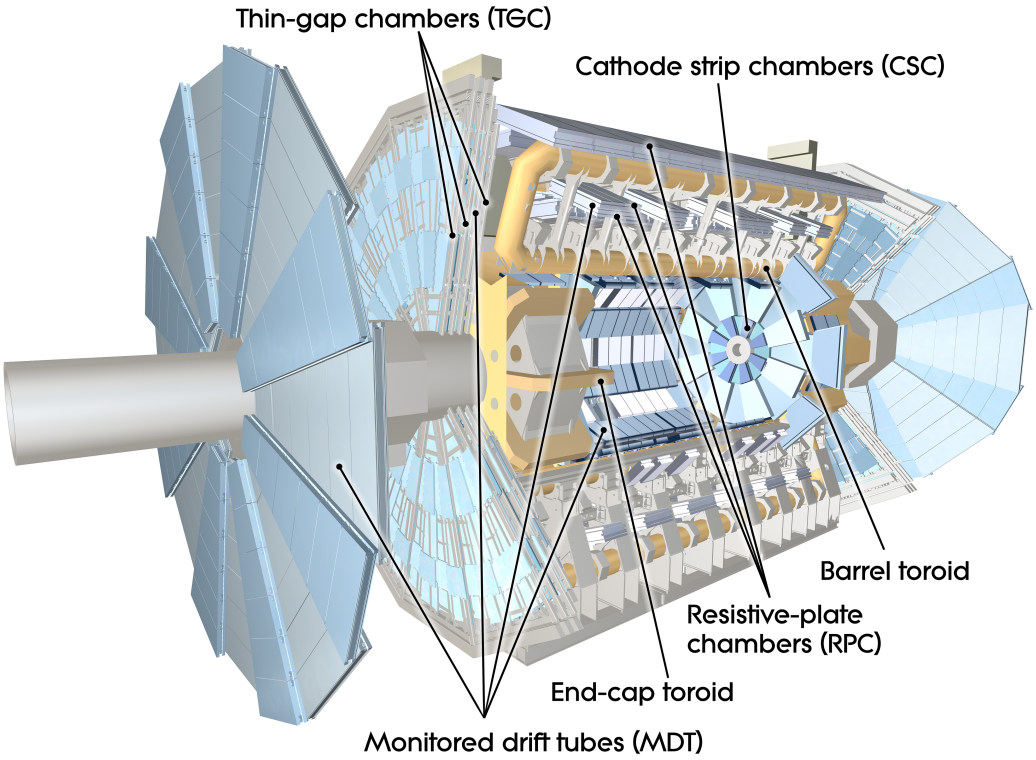


Figure 1.: Représentation schématique des sous-systèmes du spectromètre à muons ATLAS [7].

Par rapport au point d'interaction (IP), le détecteur ATLAS est symétrique dans la direction avant-arrière du faisceau de particules, couvrant approximativement un angle solide complet. Il présente une granularité complexe, impliquant de nombreux sous-systèmes effectuant différentes tâches de détection. Pour une représentation plus facile, ces différents sous-systèmes peuvent être considérés comme des couches, puisque leur disposition par rapport au point central est associée à leur rôle fonctionnel. Au plus près du IP, nous avons les systèmes de détection à l'intérieur de celui-ci, impliqués dans le traçage des particules résultant des collisions. Ici, des détecteurs de haute précision avec des pixels et des bandes sont mis en œuvre pour mesurer le momentum et les sommets. Plus loin de l' IP, le sous-système suivant est le calorimètre électromagnétique (EM) à argon liquide (LAr), dont la tâche principale consiste à mesurer à haute résolution les énergies et les positions. Plus loin encore, on trouve le calorimètre hadronique Tile, qui couvre principalement la région centrale du détecteur ATLAS. Les deux calorimètres sont entourés par le spectromètre à muons, composé de trois couches de bacs de poursuite de haute précision, qui permettent d'obtenir une précision exceptionnelle dans les mesures de résolution des impulsions de

muons.

Étant donné l'importance d'identifier correctement les muons d'état final à haute impulsion au LHC, la collaboration ATLAS a conçu le spectromètre à muons avec des capacités de déclenchement et de mesure d'impulsion à haute résolution. Le spectromètre utilise la déviation magnétique des traînées de muons à l'aide d'un arrangement de trois aimants toroïdaux supraconducteurs pour effectuer les mesures de déclenchement et de p_T. Un toroïde de baril (BT) entourant le calorimètre hadronique est responsable de la déviation des trajectoires de muons dans la plage de pseudorapidité $|\eta| \leq 1.0$, tandis que pour la plage $1.4 \leq |\eta| \leq 2.7$, la même fonction est assurée par deux toroïdes d'extrémité (ECTs) plus petits placés à chaque extrémité du BT. Les BT et les ECTs synergisent leur effet dans la région de transition $1.0 \leq |\eta| \leq 1.4$, générant une déflection magnétique de leur influence combinée [8].

Au cours du Run 1 et du Run 2, le spectromètre à muons utilisait déjà plusieurs types différents de technologies de détection. Les tubes à dérive surveillée (MDT) étaient les principaux instruments pour les mesures de précision dans la majeure partie de la plage de pseudo-apidité, sauf pour les valeurs η élevées, où des chambres à bande cathodique spéciales (CSC) ont été utilisées en raison de leur granularité plus élevée. En ce qui concerne la gâchette, l'histoire est similaire, avec des chambres à plaques résistives desservant la région de $|\eta| \leq 2.4$ et des chambres à gap mince (TGC) couvrant la région aux extrémités [8]. Le schéma général du spectromètre à muons est représenté sur la figure 1.

Comme tous les détecteurs ont beaucoup de canaux de lecture et un taux de collision très élevé, l'expérience ATLAS ne peut pas stocker physiquement chaque événement de collision. En plus des limitations techniques, tous les processus physiques ne présentent pas le même intérêt pour la recherche. Depuis sa conception, le détecteur ATLAS a été conçu pour mettre en œuvre des mécanismes de présélection permettant de décider rapidement si un événement mérite d'être conservé ou non. Ce type de processus est communément appelé "trigger".

Déclencheur d'acquisition de données basé sur l'apprentissage automatique pour la "New Small Wheel"

Compte tenu du caractère récent de ce type de technologie logicielle et du fait que les ressources des FPGA de la génération actuelle sont encore très limitées par rapport au matériel dédié aux ML, la première étape de cette recherche et développement doit être une preuve de concept. Dans cette optique, le prototype initial vise à mettre en œuvre des modèles de réseaux neuronaux au niveau matériel qui, en utilisant uniquement les capacités de calcul du FPGA, parviennent à reproduire le comportement du déclencheur de la phase I actuellement mis en œuvre sans dépasser le budget temps. Dans la mesure

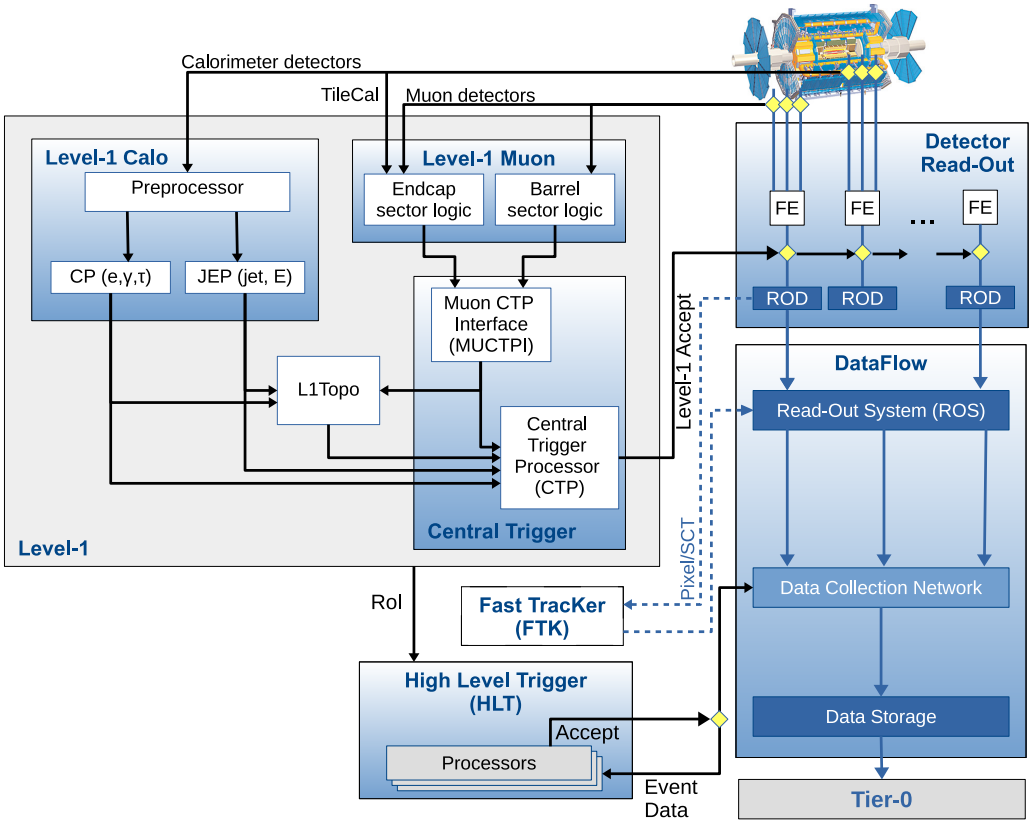


Figure 2.: Diagramme de flux de données du système de déclenchement ATLAS [9].

de la contribution de cette thèse, seuls les détecteurs MicroMegas ont été modélisés jusqu'à présent, car un échec avec cette configuration simplifiée invaliderait la méthode pour le sTGC plus complexe.

Simulation de données

La géométrie NSW est déjà mise en œuvre dans ATHENA [10], ce qui permet de réaliser des simulations Monte Carlo (MC) GEANT4 des trajectoires des muons à travers les détecteurs nouvellement installés. La figure 3 montre le résultat d'une telle simulation en termes d'impacts de MicroMegas, projetés en 2D sur le plan xy. Les petits secteurs sont représentés en bleu, tandis que les grands secteurs sont représentés en rouge. Quel que soit le type de secteur, tous les circuits imprimés MM sont mis en évidence individuellement.

Pour simplifier davantage la mise en place de la preuve de concept, on considère un seul secteur et on donne aux plans des formes rectangulaires, en conservant l'espacement des bandes de 0.4 mm. Le nombre de bandes est arrondi à 8800, tout en préservant la configuration du plan stéréo avec des bandes inclinées de 1.5° et de 1.5° (par rapport à

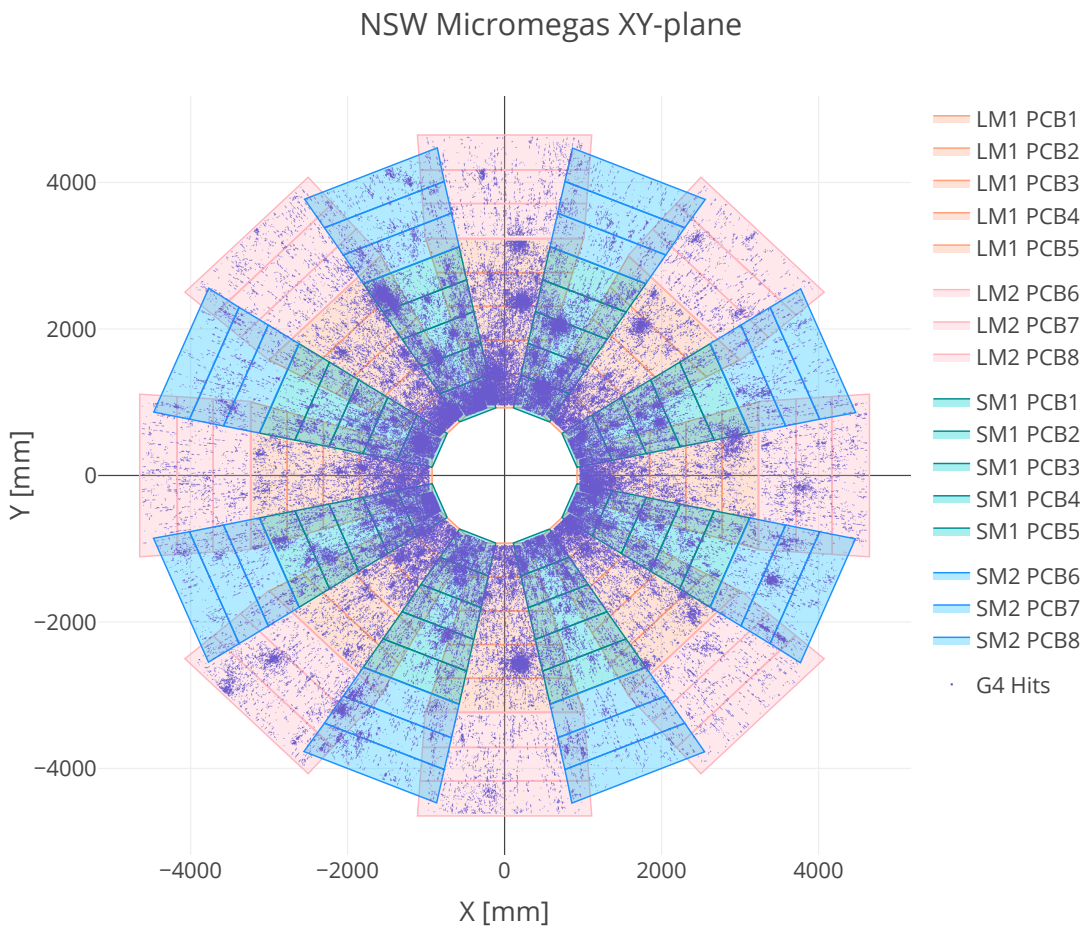


Figure 3.: Position des impacts de muons simulés par GEANT4 dans le plan xy de NSW

l'axe des x). Ces changements sont reflétés dans la figure 5, où le bleu argenté montre les plans X, les plans représentés par le U orange ont des bandes inclinées à 1.5°, et les plans V représentés par le rose ont des bandes inclinées à -1.5°.

L'utilisation de cette configuration a également nécessité moins de ressources informatiques pour la simulation, permettant l'utilisation d'outils logiciels plus simples. La base de code publique GitHub de A. Wang [11] a été utilisée comme point de départ pour développer un code de simulation reproduisant la configuration simplifiée choisie. Un échantillon de données de réponse du détecteur MM a été généré pour un million de trajectoires de muons simulées, avec des représentations ROOT [12] contenant les emplacements d'impact en termes de numéro de bande et de plan, ainsi que les $\Delta\theta$ associés à la trajectoire.

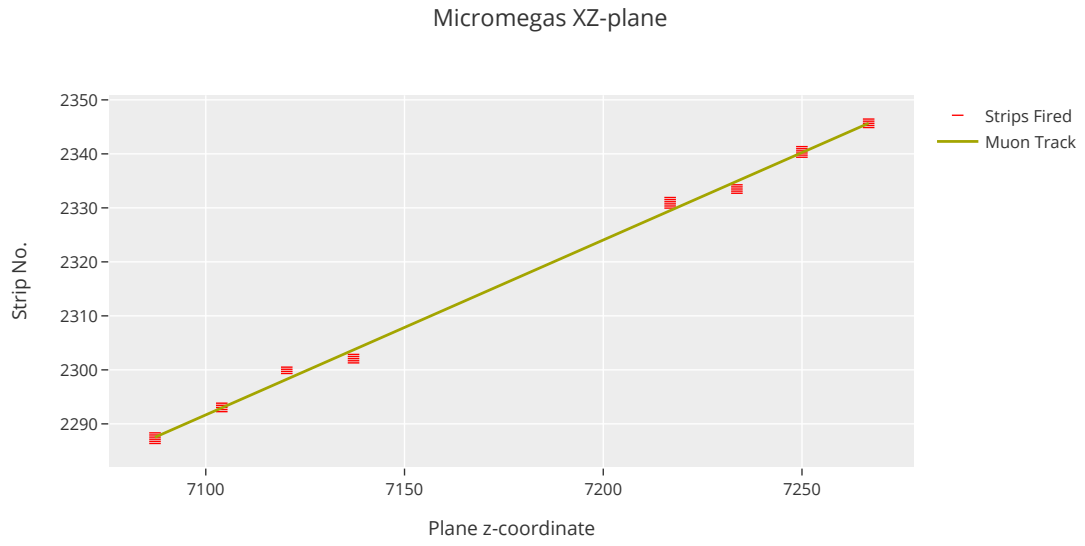


Figure 4.: La trajectoire des muons à travers un secteur MicroMegas est représentée en vert, les bandes déclenchées étant représentées par des lignes horizontales rouges

Algorithme de déclenchement basé sur un réseau neuronal

Les traces de muons à travers un secteur de MM montrent une strucuture bien défini en termes de nombre de bandes et de couches et il peut être représenté visuellement, comme le montre la Figure 4. Cela a incité à explorer les algorithmes de *vision par ordinateur* qui sont spécialisés dans l'identification de motifs géométriques. En termes de Machine Learning, cela peut être réalisé par des modèles Convolutional Neural Network (CNN) qui prennent des images en entrée.

Le *état* du détecteur peut être encodé sous forme d'image si nous tenons compte de toutes les bandes à travers les huit couches à un moment donné. Une matrice binaire éparse $M_{8800\times 8}$ peut englober toutes les informations MM puisqu'elle contient un élément distinct pour chaque bande. Ainsi, si la bande i dans le plan j a enregistré un succès, $M_{i,j} = 1$ et sinon la valeur serait 0. En traitant tous ces éléments de matrice comme des pixels, on obtient une image très étroite contenant l'emplacement de chaque hit enregistré.

Avec la matrice $M_{8800\times 8}$ comme entrée du modèle, l'étape suivante consiste à définir une cible. Les premières itérations ont testé cette approche dans un scénario de classification, le réseau étant chargé de différencier les punaises cosmiques des muons rapides. Cette tâche s'est avérée triviale pour un CNN et l'étape suivante a consisté à tenter une **régression** de la mesure de la qualité des traces $\Delta\theta$.

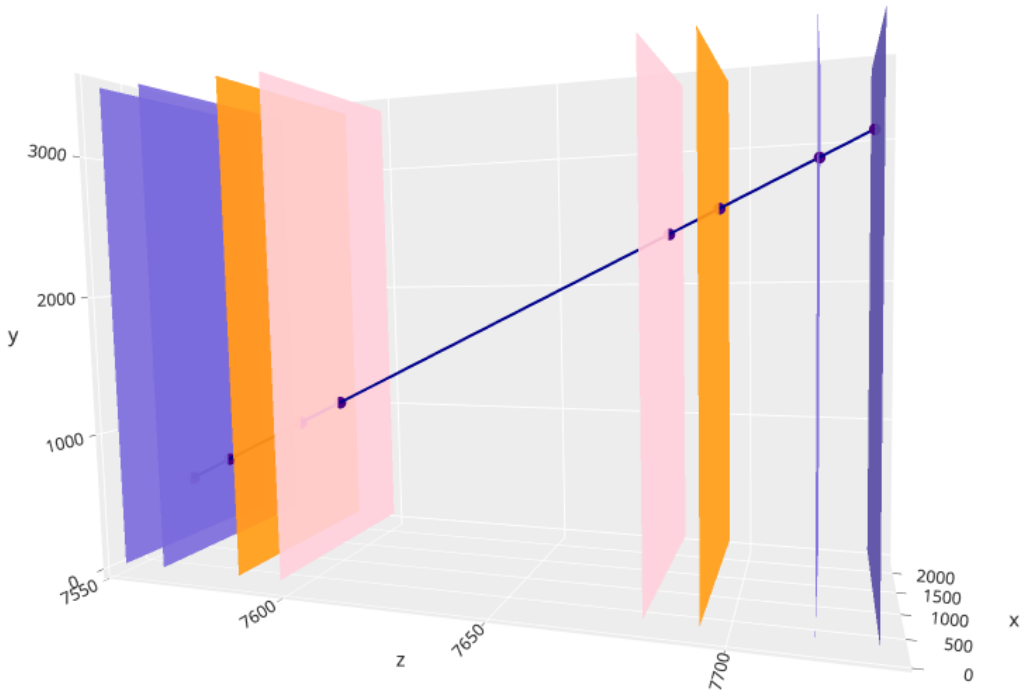


Figure 5.: Représentation 3D d'une trajectoire de muons dans la configuration simplifiée de l'assemblage MM

S'adapter aux limites des ressources du FPGA implique de trouver un équilibre entre le nombre de paramètres du modèle et ses performances. La matrice d'occurrences étant assez grande, puisqu'elle contient 8800 de bandes, il a fallu recourir à des techniques créatives de réduction de la dimensionnalité. Ainsi, les blocs de construction du modèle étaient des *blocs de convolution* qui contiennent une couche de convolution, une couche de mise en commun maximale Rectified Linear Unit (ReLU) d'activation. La figure 6 montre le premier de ces *block*, appliqué à la matrice d'entrée.

Lors du prototypage, un réseau comportant trois *blocs de convolution* a semblé être le meilleur compromis entre les performances de régression et la complexité du modèle. Le *Max Pooling* a été fortement appliqué afin de réduire la dimensionnalité verticale de l'entrée d'un facteur quatre, après chaque bloc. Le modèle complet est représenté dans la figure 7, contenant une chaîne de trois blocs convolutifs dont la sortie finale est aplatie et utilisée pour calculer la cible prédite : $\Delta\theta$.

La conception de cette architecture a donné la priorité à un petit nombre de paramètres, afin de faciliter le déploiement de FPGA. Le tableau 1 montre chaque couche du modèle ainsi que le nombre de paramètres entraînables qui lui sont associés. *Batch Normalisation* [13] et *dropout* [14] ont été utilisés à des fins de régularisation.

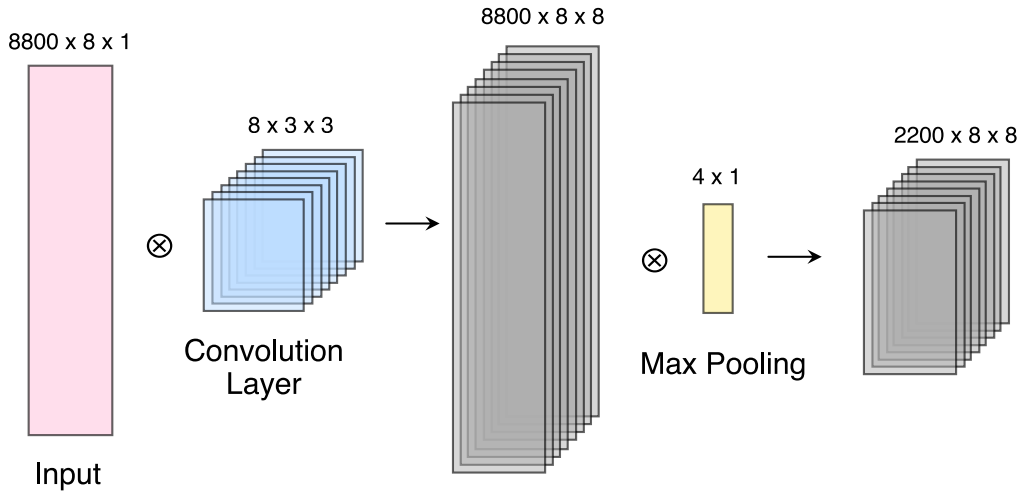


Figure 6.: Bloc convolutif montrant l'entrée et la sortie de l'opération de convolution et de la mise en commun max.

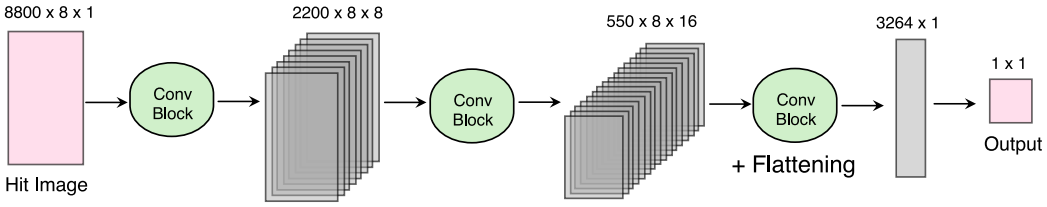


Figure 7.: Architecture de réseau neuronal déployée dans la régression $\Delta\Theta$.

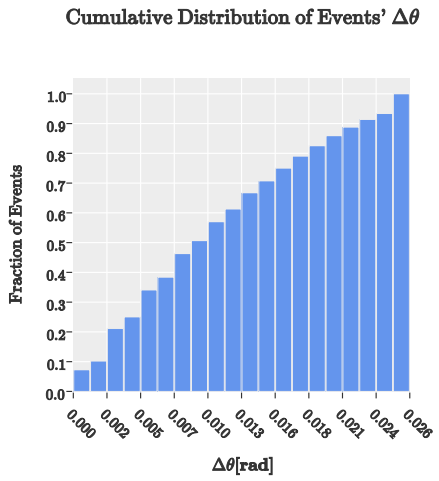
L'apprentissage du modèle a utilisé l'optimiseur Adam [15] avec les paramètres par défaut et a été facilité par l'accélération Graphics Processing Unit (GPU) d'une carte *Nvidia Tesla V100S*. La taille des lots a été fixée à 500 et l'ensemble de données a été réparti à 80% entre la formation, la validation et le test. La fonction de perte était l'erreur quadratique moyenne et le processus de formation n'a montré aucun signe de surajustement en termes de différence entre la perte de formation et de validation.

Résultats

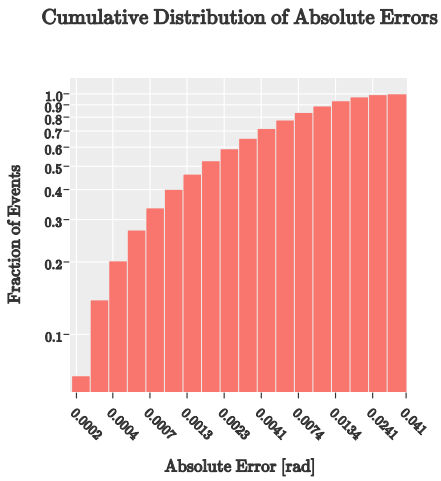
Le modèle entraîné a été évalué sur les données réservées aux tests, en mesurant le Erreur Absolue Moyenne (MAE) des prédictions. L'erreur absolue moyenne en pourcentage MAPE n'aurait pas pu être utilisé de manière fiable compte tenu de la petitesse des valeurs cibles. La figure 8 résume les performances de la régression.

Type	Shape	Parameters
InputLayer	[(None, 8800, 8, 1)]	0
BatchNormalization	(None, 8800, 8, 1)	4
Conv2D	(None, 8800, 8, 8)	80
MaxPooling2D	(None, 2200, 8, 8)	0
ReLU	(None, 2200, 8, 8)	0
Conv2D	(None, 2200, 8, 16)	1168
MaxPooling2D	(None, 550, 8, 16)	0
ReLU	(None, 550, 8, 16)	0
Conv2D	(None, 548, 6, 32)	4640
ReLU	(None, 548, 6, 32)	0
MaxPooling2D	(None, 34, 3, 32)	0
Flatten	(None, 3264)	0
Dropout	(None, 3264)	0
Dense	(None, 1)	3265
Rescaling	(None, 1)	0

Table 1.: Couches et nombre de paramètres de la régression CNN $\Delta\theta$.



(a) Distribution cumulée de la valeur absolue du test $\Delta\theta$.



(b) Distribution cumulative en log-binning de l'erreur absolue de prédiction du test $\Delta\theta$.

Figure 8.: La distribution des valeurs de $\Delta\theta$ de l'ensemble de données de test à côté de la distribution des erreurs absolues de prédiction de $\Delta\theta$ du modèle

En termes d'erreurs absolues, neuf pistes de muons sur dix présentent des erreurs de prédiction absolues inférieures à $0,01$ rad, comme le montre la figure 8(b). Cette distribution cumulative d'erreurs présente un binning logarithmique et exprime le nombre de bin comme une fraction cumulative du nombre total de points de données. Un examen plus approfondi de la figure 8(a) montre que les valeurs de $\Delta\theta$ pour 50% des pistes sont inférieures à $0,01$ rad. Ici encore, la distribution est un comptage cumulatif de la fraction des événements dans le log-binning.

Lorsqu'on les interprète ensemble, il apparaît que pour les trajectoires de $\Delta\theta \leq 0,01$ rad, la prédiction du modèle peut comporter des erreurs allant jusqu'à 100%. Cela ne devrait cependant pas être aussi inquiétant qu'il n'y paraît, puisque les plus petites pistes $\Delta\theta$ sont par définition les plus susceptibles d'être des prompt muons. Une erreur de régression plus élevée sur celles-ci est très peu susceptible d'introduire des déclenchements erronés.

Ce modèle montre que les modèles CNN sont une solution potentielle pour le processeur de déclenchement NSW, au moins dans le cas des détecteurs MicroMegas. La prochaine étape logique est le déploiement sur du matériel FPGA simulé, puis physique. Les premières tentatives de migration FPGA ont été réussies sur des cartes accélératrices *Xillinx Alveo U220*. Ce processus a nécessité la quantification du modèle et l'exclusion des couches de régularisation (dropout et normalisation par lots), mais il n'a pas eu de coût significatif en termes de précision de régression. Au moment de la rédaction de cet article, des études sur la synchronisation de l'inférence sont en cours, les goulots d'étranglement du pipeline de données du logiciel Alveo posant de légers défis techniques.

Recherche de BSM (physique au-delà du modèle standard) à l'aide de méthodes de détection des anomalies

La recherche de phénomènes physiques au-delà du modèle standard (BSM) est l'un des domaines de recherche les plus actifs en physique des hautes énergies (HEPP). La manière établie d'effectuer ce type de recherche dépend principalement de l'utilisation d'analyses spécifiques complexes qui tentent de vérifier les prédictions de certains modèles théoriques à l'aide de données expérimentales réelles. Avec une multitude de modèles parmi lesquels choisir et des quantités de données toujours plus importantes, l'éventail des recherches possibles en matière de BSM s'élargit et il devient de plus en plus difficile de s'y retrouver. Bien qu'elle soit parfois très fructueuse, comme dans le cas de la découverte du boson de Higgs, cette approche exige un travail très spécialisé qui peut ne pas être réellement rentable, sauf si les prédictions du modèle de base peuvent être démontrées ou réfutées avec succès. Compte tenu des défis mentionnés ci-dessus, une partie de la communauté de la physique des particules se tourne vers d'autres moyens de recherche des signaux BSM.

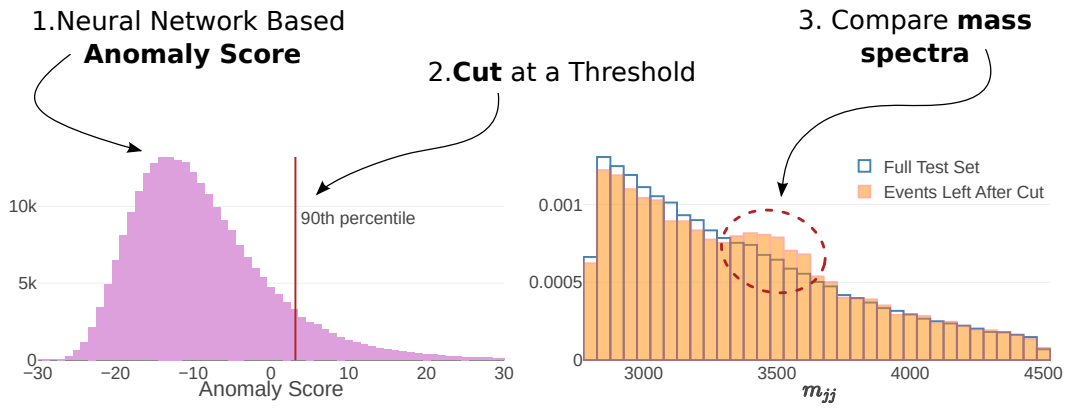


Figure 9.: Stratégie de recherche BSM utilisant des méthodes de détection d'anomalies basées sur les réseaux neuronaux

Les méthodes basées sur l'apprentissage automatique pour détecter les anomalies non supervisées et semi-supervisées se sont avérées extrêmement utiles dans un large éventail d'applications. Cela soulève la question de savoir si une telle approche pourrait être développée pour les recherches indépendantes du modèle des signaux BSM. Compte tenu des énormes quantités de données recueillies par ATLAS au cours de ses années de fonctionnement, la détection d'anomalies basée uniquement sur les données pourrait permettre de trouver des signaux inhabituels indiquant une nouvelle physique.

L'un des principaux objectifs de recherche de cette thèse était d'explorer des méthodes non supervisées basées sur le ML pour identifier les jets présentant des anomalies. La figure 9 donne un aperçu de la stratégie de recherche BSM utilisant ce type de cadre et montre trois étapes principales :

- Un modèle de réseau neuronal (NN) est entraîné pour générer des scores d'anomalie qui représentent une évaluation quantitative du degré de dissimilarité de certains événements par rapport à d'autres.

- L'application d'un seuil défini par l'utilisateur à la distribution des scores d'anomalie permet d'obtenir un sous-échantillon contenant les données les plus anormales.

- Comparez les propriétés de l'ensemble de données original et du sous-échantillon anormal. Dans le scénario de recherche BSM, les distributions de masse invariantes d'objets physiques, lorsqu'elles sont placées les unes au-dessus des autres, peuvent révéler une résonance liée à une nouvelle particule.

Apprentissage non supervisé

L'approche supervisée de l'apprentissage automatique utilise des cibles de sortie connues pour les données de formation, mais ces informations ne sont pas toujours accessibles. Les modèles ML tentent d'identifier les modèles de données qui sont étroitement liés aux cibles de prédiction, mais ces modèles inhérents peuvent exister même en l'absence d'une étiquette de formation bien établie. L'apprentissage non supervisé se réfère exactement à ce scénario, où le ML est utilisé pour rechercher la structure sous-jacente des données et les relations entre les observations, sans accès aux variables dépendantes.

Dans le contexte de l'utilisation de la recherche BSM, certaines recherches suggèrent qu'une combinaison des deux classes de méthodes peut être plus performante que l'utilisation de l'une d'entre elles individuellement [16].

Il pourrait y avoir de nombreuses façons de combiner un autoencodeur (AE) avec l'estimation de densité, mais un ensemble particulièrement intéressant qui y parvient est l'autoencodeur probabiliste (PAE) [17]. Le PAE consiste en un modèle AE dont l'espace latent est mappé dans un espace distribué normal par un modèle de type Normalizing Flow (NF) [18]. La figure **??** montre un diagramme de l'architecture du PAE.

Dans le modèle d'autoencodeur probabiliste, l'autoencodeur et le modèle de Normalizing Flow peuvent travailler ensemble pour tenter de fournir une estimation de la probabilité des entrées $p(\vec{x})$. NF est déjà capable de fournir une fonction de densité traçable pour la représentation latente $p_\gamma(\vec{z})$, mais malheureusement, comme la propriété d'injectivité n'est pas applicable pour le codeur $\mathbf{f}_\phi$, les densités d'entrée et de codage ne sont pas identiques : $p(\vec{x}) \neq p_\gamma(\vec{z})$. Pour parvenir à une approximation de $p(\vec{x})$, nous commençons par écrire la probabilité conjointe $p(\vec{x}, \vec{z})$ comme un produit de probabilités conditionnelles selon l'équation 0.1 [17].

$$p(\vec{x}, \vec{z}) = p_\theta(\vec{x}|\vec{z}) p_\gamma(\vec{z}) \tag{0.1}$$

Ensuite, $p(\vec{x})$ peut être isolé en marginalisant sur z, comme dans l'équation 0.2. Cependant, cette intégrale n'est pas directement traitable et nécessitera une certaine approximation.

$$p(\vec{x}) = \int p_\theta(\vec{x}|\vec{z}) p_\gamma(\vec{z}) d\vec{z} \tag{0.2}$$

La probabilité implicite $p_\theta(\vec{x}|\vec{z})$ est supposée avoir une forme gaussienne, comme l'indique la référence [17], tandis que la matrice de covariance $\mathbf{œ}$ est diagonale, et ses éléments non nuls sont représentés par les erreurs moyennes de reconstruction pour chaque caractéristique. Cette probabilité étant définie, l'intégrale de l'équation 0.2 peut être calculée en utilisant l'approximation de Laplace [19]. Selon les auteurs de l'article original sur l'auto-encodeur

probabiliste, les termes hessiens introduits par la méthode de Laplace réduisent la stabilité numérique, ils seront donc abandonnés à partir de maintenant.

En prenant tout cela en compte, l'équation 0.3 montre la formule résultante pour l'approximation de $p(\vec{x})$.

$$\ln p(\vec{x}) \approx -\frac{1}{2}||\vec{x} - \vec{x}'||^2 \vec{\sigma}^{\circ -2} - \frac{1}{2} b_\gamma(\vec{z})^2 + \ln|\det \mathcal{J}_\gamma| \tag{0.3}$$

En analysant ce score, le premier terme peut être reconnu comme l'erreur quadratique moyenne (MSE) de l'autoencodeur, et $\vec{\sigma}$ est la diagonale de la matrice de covariance de la vraisemblance implicite. L'autre terme du score d'anomalie est la vraisemblance de la représentation de l'espace latent, fournie par le modèle de Normalizing Flow.

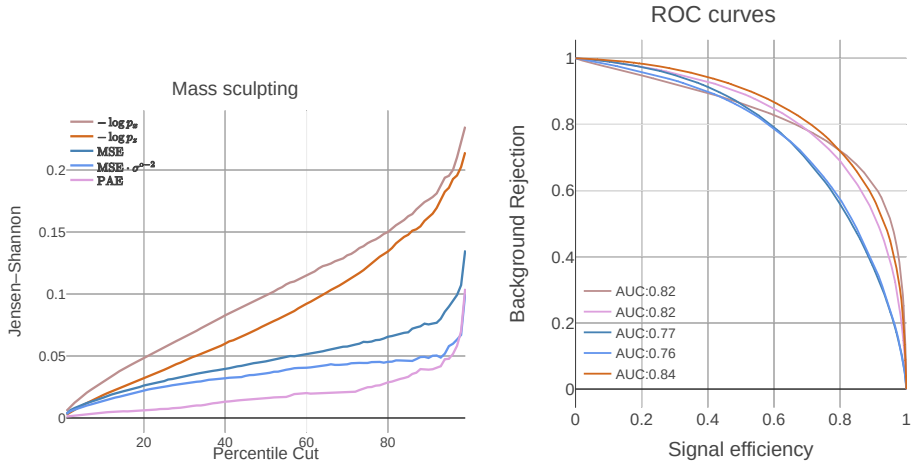
Pour la détection des anomalies, un autoencodeur probabiliste entraîné fournit plusieurs indicateurs potentiels pour la classification. On peut considérer à la fois la probabilité de la représentation latente et l'erreur de reconstruction de l'autoencodeur, ou quelque chose comme $||\vec{x} - \mathbf{g}_\theta(\vec{z}')||^2$, qui utilise la densité NF indirectement par optimisation logarithmique négative de la log-vraisemblance requise pour trouver $\vec{z}'$. Un autre choix évident serait d'estimer la vraisemblance des données d'entrée $\ln p(\vec{x})$. Compte tenu de cette flexibilité en termes de scores d'anomalies, l'autoencodeur probabiliste peut utiliser efficacement les informations relatives à la densité et à l'erreur de reconstruction en même temps, ce qui en fait un bon et précieux candidat pour d'éventuelles recherches BSM non supervisées ou semi-supervisées.

Résultats

Bien que plusieurs approches aient été expérimentées, l'autoencodeur probabiliste (PAE) a fini par être le principal point de mire, étant donné le grand nombre de scores d'anomalie potentiels qui peuvent en être dérivés. En utilisant la stratégie générale décrite au début de ce chapitre, un flux de travail complet de détection d'anomalies a été développé pour rechercher les signaux BSM inclus dans les ensembles de données Olympics [20] du LHC. Puisque la stratégie consiste à rechercher des résonances dans le spectre m_{jj} après avoir sélectionné les événements en fonction du score d'anomalie, il est particulièrement important de s'assurer que le score d'anomalie n'est pas corrélé avec la masse invariante du système à deux jets.

Une fois que le score d'anomalie le plus prometteur est identifié, l'étape suivante consiste à mettre en œuvre la méthode d'analyse complète de l'ensemble de données *Black Box*. En appliquant une division de la distribution des scores d'anomalie pour sélectionner un sous-échantillon des événements les plus anormaux, ce sous-échantillon est ensuite comparé aux données complètes de l'ensemble de données *Black Box*. Si les deux distributions normalisées de m_{jj} sont analysées côte à côte, on s'attend à trouver un excès correspondant au signal BSM

dans le sous-échantillon enrichi en anomalies. Cet excès est recherché à l'aide de l'algorithme Bump Hunter [21], qui, en plus de faciliter la recherche de résonances, est également capable de calculer la signification des résultats. Afin d'intégrer de manière transparente la chasse aux bosses dans la base de code d'analyse de PYTHON [22] et de JUPYTER Notebook [23], la bibliothèque PYBUMPHUNTER a été utilisée pour les fonctionnalités et les performances qu'elle offre.



(a) Biais de masse calculé à chaque seuil de score d'anomalie percentile et exprimé en termes de divergence JS

(b) Performances en matière de détection des anomalies, exprimées par la courbe ROC.

Figure 10.: Mesures de performance pour les scores d'anomalies potentielles accessibles par l'ensemble PAE.

Évaluation quantitative du biais du score d'anomalie par rapport à m_{jj} impliquant la divergence Jensen-Shannon (JS) [24]. Après la formation, les scores d'anomalie ont d'abord été extraits sur l'ensemble de formation conçu pour être exempt de signaux. Ensuite, des seuils successifs couvrant tous les percentiles de la distribution des scores d'anomalie ont été appliqués. Pour chacun de ces seuils, on a calculé la divergence JS entre les spectres m_{jj} de l'échantillon d'entraînement complet et du sous-échantillon au-dessus du seuil. L'objectif ici est de déterminer dans quelle mesure la distribution de m_{jj} est affectée par le score d'anomalie. Comme la divergence JS mesure la distance entre ces deux distributions, un score non biaisé montrerait des valeurs faibles pour la divergence JS dans ce scénario, quel que soit le seuil, puisqu'il n'y a pas de résonance à trouver dans ce scénario.

Les graphiques de la déformation de la distribution de masse exprimée sous forme de divergence JS en fonction du seuil percentile du score d'anomalie sont présentés dans la Figure 10(a). Un comportement commun, quel que soit le score en question, est la forte augmentation de la divergence JS aux derniers percentiles. Cela peut très probablement être attribué aux fluctuations statistiques qui deviennent de plus en plus prononcées à

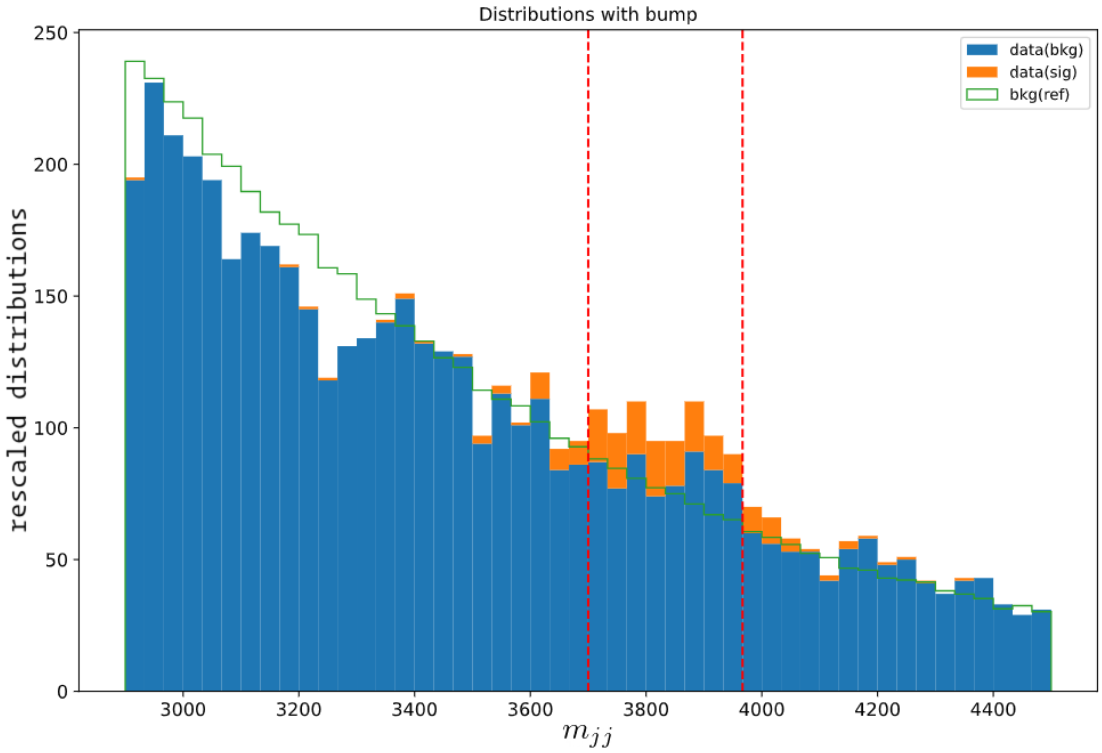


Figure 11.: Résultats de la recherche de bump hunting après comparaison des données de la boîte noire 1 avec le sous-échantillon d'événements anormaux dans le 99e percentile. Des couleurs sur le graphique ont été ajoutées à la fin pour confirmer la présence d'un vrai signal BSM.

mesure que de moins en moins d'événements franchissent le seuil du score d'anomalie. Sans tenir compte de cela, il est clair que les probabilités logarithmiques négatives des deux modèles NF introduisent des quantités significatives de biais m_{jj}. D'autre part, les scores d'anomalie basés sur la reconstruction ajoutent nettement moins de distorsion de masse. Il est intéressant de noter que le score PAE semble être généralement le moins biaisé de tous, moins que l'un des deux scores qu'il incorpore.

Le pouvoir de discrimination du signal du score d'anomalie est aussi important que le fait d'avoir une déformation minimale de la masse. Si l'ensemble de données de test synthétiques est équilibré entre les classes étiquetées, les scores d'anomalie peuvent être évalués en tant que classificateurs à travers le prisme d'indicateurs de performance conventionnels tels que la courbe Receiver Operating Characteristics (ROC) et l'aire sous celle-ci. Si l'on juge les scores d'anomalie par la valeur de l'aire sous la courbe ROC (AUC), leurs performances semblent assez similaires, comme le montre la figure 10(b). Bien que ces valeurs AUC ne soient pas comparables à celles d'un modèle de classification supervisé, elles sont tout de même respectables étant donné qu'il s'agit d'une méthode semi-supervisée. Les scores d'anomalie basés sur la reconstruction sont parmi les plus faibles, tandis que les estimations

de densité sont plus fortes, avec de petites différences entre elles.

Résumée dans la figure 10, la comparaison des performances du score d'anomalie a donné des résultats intéressants. Le score PAE est le moins biaisé, tout en ayant l'un des meilleurs chiffres de mérite pour le pouvoir discriminatoire. Bien que l'estimation PAE de la log-vraisemblance négative des données d'entrée soit basée à la fois sur l'erreur de reconstruction de l'autoencodeur et sur la densité de flux de normalisation, elle n'hérite d'aucun des biais m_{jj} de ces composantes, tout en empruntant les performances de classification des meilleures.

La configuration la plus performante a été testée sur la *Black Box* 1, et les résultats sont présentés dans la figure 11. La déformation de la distribution de masse invariante semble inévitable, quelle que soit la configuration utilisée. Malgré cela, le modèle est capable de trouver le signal dans l'ensemble de données de la boîte noire 1 de manière reproductible. Une fois le contenu de l'ensemble de données révélé, le contenu du signal de la bosse-ulu est immédiatement évident. La sélection des événements basée sur le score d'anomalie donne un échantillon avec $S/B = 1,19\%$ (signal sur fond), ce qui est un ordre de grandeur supérieur à $S/B = 0,08\%$.

Conclusions

Le potentiel de progrès dans le domaine de la High-Energy Particle Physics (HEPP) est encore important. Bien que l' Standard Model (SM) fournisse une description exceptionnellement bonne des aspects les plus fondamentaux de la nature, certains phénomènes doivent encore être expliqués. Des théories concurrentes concernant la physique Beyond Standard Model (BSM) proposent des explications pour plusieurs mystères restants tels que les masses des neutrinos, la matière noire ou la gravité quantique. Cependant, les preuves expérimentales nécessaires ne sont pas encore disponibles pour les prouver ou les réfuter toutes.

Bénéficiant de l'expérience A Toroidal LHC Apparatus (ATLAS) au CERN et y contribuant, cette thèse a couvert deux sujets majeurs relatifs à l'état actuel de High-Energy Particle Physics (HEPP).

En ce qui concerne la prise de données et la détection de particules, j'ai développé, dans le cadre de ma thèse, un prototype d'algorithme de trigger basé sur le Machine Learning et ciblant les détecteurs MicroMegas (MM) déployés avec le New Small Wheel (NSW). L' Research and Development (R&D) sur ce front n'en étant qu'à ses débuts, le prototype actuel sert de preuve de concept attestant de la faisabilité de l'approche. L'itération actuelle déploie avec succès un Convolutional Neural Network (CNN) qui utilise les signaux du détecteur pour estimer la variable la plus pertinente impliquée dans le NSW trigger décision.

Enfin, j'ai conçu et étudié un flux de travail d'analyse de détection d'anomalie indépendant du modèle visant des applications dans la recherche de la physique Beyond Standard Model. La méthodologie proposée repose sur l'utilisation d'un ensemble Neural Network pour générer des *score d'anomalie* conçus pour être sensibles à la présence de BSM signaux qui ne sont pas connus *a priori*. Le modèle Probabilistic Autoencoder (PAE) a été choisi comme principal modèle d'intérêt et plusieurs scores d'anomalie potentiels qui pourraient en être dérivés ont été testés sur des ensembles de données synthétiques. L'accent a été mis sur l'atténuation des biais potentiels du modèle afin de réduire la probabilité de trouver de faux signaux. Les résultats obtenus sur le jeu de données **Black Box 1**, qui fait partie du défi *LHC Olympics*, mettent en évidence les performances de la méthode en matière de détection des anomalies. Dans un scénario semi-supervisé, le score d'anomalie PAE a été utilisé pour obtenir une augmentation d'un ordre de grandeur du rapport entre le signal de l'échantillon et le contenu du fond. Bien qu'un certain biais ait été introduit en ce qui concerne la masse du dijet, le déploiement des algorithmes *bump hunting* sur la distribution m_{jj} de l'échantillon enrichi en anomalies a permis une identification correcte du signal BSM par rapport au fond de référence complet **Black Box 1**.

Bien que j'aie fourni des résultats tangibles pour les sujets abordés, mon intérêt pour ceux-ci est loin d'être épuisé. Il existe plusieurs projets que je considère comme une continuation potentielle du travail effectué jusqu'à présent.

Le prototype de déclencheur de preuve de concept MicroMegas pour le modèle ATLAS New Small Wheel (NSW) permet de poursuivre le développement des algorithmes de déclenchement. Une prochaine étape importante serait l'ajout de données small-strip Thin-Gap Chambers au prototype actuel. Avec le Long Shutdown 3 qui débutera fin 2025 et durera jusqu'en 2029, un déclencheur complet basé sur le Machine Learning pourrait être achevé à temps pour le prochain cycle de prise de données.

Comme pour les recherches BSM indépendantes du modèle, l'approche Probabilistic Autoencoder peut encore être optimisée en termes de réduction du biais introduit par le modèle. Il existe des stratégies potentielles qui n'ont pas encore été explorées, comme les techniques de décorrélation de masse et l'utilisation de caractéristiques jet supplémentaires.

Introduction

The existence of yet unexplained phenomena in particle physics presents a great opportunity for research progress in the High-Energy Particle Physics (HEPP) field. While there are many Beyond Standard Model (BSM) models proposing several different solutions to the SM limitations, experimental evidence is necessary to validate their validity. Pushing those frontiers implies constant improvements both in the experimental data acquisition capabilities and data analysis methodology. Machine Learning (ML) methods have seen widespread adoption and success across many fields, aided by the increasingly better performing Graphics Processing Unit (GPU) hardware acceleration. This Ph.D.research aims to study the potential effectiveness of applying Neural Network (NN) implementations to improve different areas of HEPP.

A general overview of the current state of the Standard Model is presented in Chapter 1, alongside a small discussion on its limitations and a few proposed solutions. Chapter 2 introduces the Large Hadron Collider machine and further focuses on the A Toroidal LHC Apparatus (ATLAS) experiment. All major subdetector systems of the ATLAS Detector are covered in detail. The interplay between those systems leads to the reconstruction of hard scattering pp collision events. Some of the intricacies of the physics object reconstruction process from the detector signals are described in Chapter 3.

As part of the qualification process towards becoming an ATLAS author, I have undertaken a study related to electron identification performance. This part of the work is concerned with isolating the heavy-flavour sources of background electrons and has Chapter 4 dedicated to it.

There are two main areas of personal contributions, both involving the use of Machine Learning models. Thus, Chapter5 introduces the general aspects of Machine Learning and Neural Network (NN) models.

On the detector side, my work focused on evaluating the feasibility of a hardware-level trigger for the newly installed ATLAS New Small Wheel (NSW) end-cap muon detector. The Trigger Processor (TP) operates *in-situ* on Field-Programmable Gate Array (FPGA) hardware. Chapter 6 presents a feasibility study for a Machine Learning based muon trigger algorithm targeting future upgrades of the New Small Wheel. A Convolutional Neural Network (CNN) approach was chosen to interpret hits from the NSW's eight MicroMegas (MM) detector planes and trained to output quantities related to the likelihood for a muon track to originate from the interaction point (IP).

The second main topic of this thesis involves the exploration of **unsupervised learning** as a means to model-agnostic BSM searches. A Neural Network ensemble called Probabilistic Autoencoder (PAE) was the central technique employed in my research. Chapter 7 presents the inner workings of the PAE model as well as how it can be deployed within a full anomaly detection analysis workflow. This approach was tested in the context of jet physics, being applied to challenging anomaly detection benchmark scenarios.

1. Theoretical Basis in High Energy Physics

Even before the conception of the scientific method, humans were interested in understanding the fundamental building blocks of our world. Early models of the universe can be traced as far back as ancient Greece where many philosophers' fascination with ontology spawned theories belonging to the realm of physics and even metaphysics. While Democritus is credited as the first to bring forth the idea of fundamental indivisible *particles* testing such a hypothesis was far out of reach for the people of the time. The whole endeavour of fundamental physics research has historically been limited by experimental capabilities and as such, technological progress often allowed finer aspects of the universe to be probed. Humans needed millennia to develop the necessary tools both on the experimental and theoretical fronts before J.J Thomson discovered in 1897 the first elementary particle: the electron.

Scientists of the 20th century set the stage for the emergence of particle physics as a field with the developments of quantum mechanics and relativity. While the theoretical tools were carefully being assembled, engineering science saw unprecedented levels of growth. Particle accelerators were being built for the first time in history and thus enabled many models to be tested and incrementally improved. All of those endeavours culminated with the establishment of the Standard Model (SM). This chapter was elaborated using *Quark & Leptons: An Introductory Course In Modern Particle Physics* Halzen and Martin [25] and 'Review of Particle Physics' Workman et al. [26] as the main bibliographical sources.

1.1. The Standard Model

During the 20^{th} century the scientific community saw many breakthroughs in fundamental physics, such as Yang-Mills theory [27], violation of parity conservation [28], asymptotic freedom of strong interactions [29], unification of the electromagnetic and weak force [30], the development of the Higgs mechanism [31, 32] and many others. Building on this newfound knowledge, the Standard Model (SM) was established in an attempt to create a unified framework for particle physics. It comprises 17 elementary particles under a Quantum Field Theory (QFT) mathematical formalism that is invariant under local gauge transformations. Each of these gauge symmetries is associated with one of three fundamental interactions. Gravity is notably missing from this theoretical framework.

1.1.1. Elementary Particles

A representative diagram of the Standard Model constituents is displayed in Figure 1.1, showing the 17 SM particles' names and symbols, grouped by their type. The outer layer in this diagram represents all particles that make up *matter*, the *fermions*.

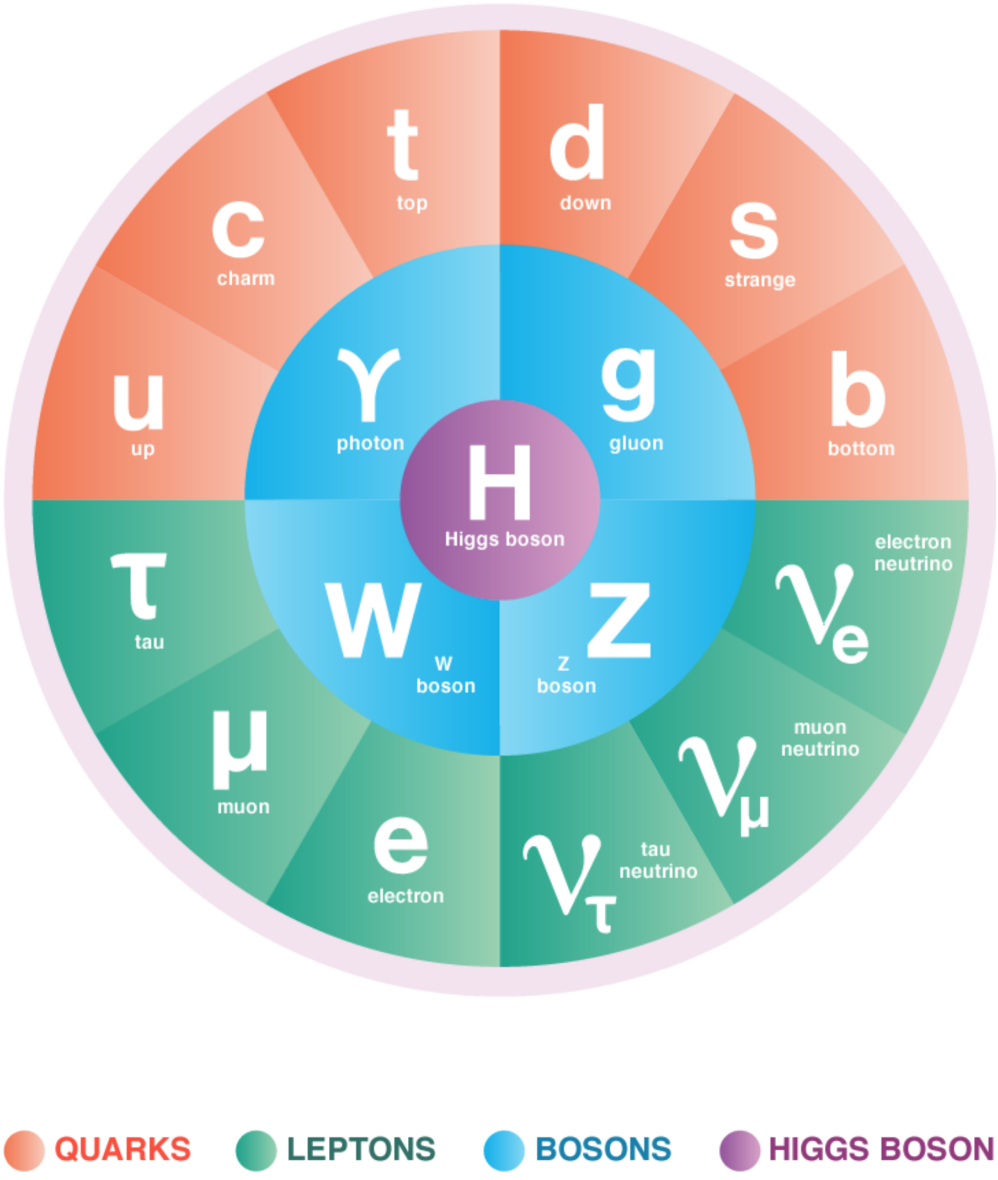


Figure 1.1.: Graphical depiction of the Standard Model (SM) particles colour-coded by type [33].

As the name suggests, fermions follow the Fermi-Dirac statistics, obeying Pauli's exclusion

principle stating that identical particles can not occupy the same quantum state within the same system. All fermions share the same $\frac{1}{2}$ value for their spin and they are further subdivided into additional categories such as quarks and leptons. Those two categories are represented in Figure 1.1 by different colours, but fundamentally the criteria for this division amount to whether or not a particle is subject to the strong interaction. Quarks are strong interacting fermions thus necessarily carrying a colour charge. This makes them subject to *colour confinement*, implying that they can not exist in an isolated state. As a consequence, quarks have never been directly observed as they always group together forming hadrons. One notable exception is the top quark whose half-life is so small that it ends up decaying before being able to undergo hadronisation. Quarks are the only elementary particles subject to all types of fundamental interactions. Most matter in the universe is made up of u and d quarks as they form the most stable baryons: the proton and neutron. All other quarks are significantly more massive and thus the hadrons they form have vastly lower lifetimes before decaying.

Category	Particle Name	Symbol	Mass	Spin	Electric Charge
Quarks	up	u	2.2 MeV	$\frac{1}{2}$	$+\frac{2}{3}$
	down	d	4.7 MeV	$\frac{1}{2}$	$-\frac{1}{3}$
	charm	c	1.3 GeV	$\frac{1}{2}$	$+\frac{2}{3}$
	strange	s	93.4 MeV	$\frac{1}{2}$	$-\frac{1}{3}$
	top	t	173.2 GeV	$\frac{1}{2}$	$+\frac{2}{3}$
	bottom	b	4.2 GeV	$\frac{1}{2}$	$-\frac{1}{3}$
Leptons	electron	e	0.511 MeV	$\frac{1}{2}$	-1
	muon	μ	105.7 MeV	$\frac{1}{2}$	-1
	tau	τ	1776.9 MeV	$\frac{1}{2}$	-1
	electron neutrino	ν_e	0	$\frac{1}{2}$	0
	muon neutrino	ν_μ	0	$\frac{1}{2}$	0
	tau neutrino	ν_τ	0	$\frac{1}{2}$	0
Gauge bosons	gluon	g	0	1	0
	photon	γ	0	1	0
	Z boson	Z	91.2 GeV	1	0
	W boson	$W\pm$	80.4 GeV	1	± 1
Higgs	Higgs boson	H	125.2 GeV	0	0

Table 1.1.: Standard Model particles (excluding antiparticles) and some observables associated with them [26]

Some fermions don't carry colour charges and thus not being subject to strong interactions; they are known as *leptons*. They get further divided corresponding to their electric charge. The charged leptons e, μ and τ interact both electromagnetically and weakly while their associated neutrinos (ν_e, ν_μ and ν_τ) are subject only to the weak interaction. This trait of neutrinos makes them especially hard to detect as opposed to their charged counterparts.

They are also considered massless within the current framework of the SM. As was the case with quarks, only the lighter leptons are stable, such as the electron and neutrinos.

The inner circle from Figure 1.1, depicted in blue, shows the vector bosons. Bosons are the force carriers mediating particle interactions and their name is suggestive of the Bose-Einstein statistics, which allows multiple identical particles to be aggregated in the same quantum state. Electromagnetic interactions are associated with the photon while gluons are responsible for the strong force keeping quarks bundled together. The remaining two massive bosons Z and W are the carriers of the weak force. The Higgs boson could be thought of as belonging to its own category since it's the only scalar boson and is associated with the field that gives the other fundamental particles their mass. In Figure 1.1 the Higgs boson occupies the centre alone since there is no other particle similar to it. Table 1.1 lists all of the Standard Model particles, as well as some observables associated with them, such as mass, elementary electric charge and spin. For every particle in the SM, there is a corresponding antiparticle with the same properties except for its physical charges having the opposite sign. Antiparticles are not directly shown in Table 1.1 since in this representation they would only differ in name and the sign of the electric charge.

1.1.2. Fundamental Interactions

Within the SM framework, elementary particles are formalized as quantum fields appearing in the Lagrangian, while the fundamental interactions are encoded as interaction terms in the definition of Lagrangian density. Each of the SM fundamental forces is modelled by their own QFT although, the electromagnetic and weak forces have been unified under the same definition of **Electroweak** theory.

Electromagnetic Interaction

Quantum Electrodynamics (QED) encompasses the foundational knowledge surrounding electromagnetism. All particles possessing electric charge are interacting via photon exchanges.

The dynamics of free spin-1/2 fermions can be described by the Dirac Lagrangian in Equation 1.1. Notation-wise, ψ is the Dirac spinor, m is the mass and the Dirac matrices are denoted as γ^μ. $\bar{\psi} = \psi^\dagger \gamma^0$ is known as the Dirac adjoint.

$$\mathcal{L}_0 = \bar{\psi}\left(i\gamma^\mu \partial_\mu - m\right)\psi \tag{1.1}$$

In the Lagrangian, the γ^μ Dirac 4×4 matrices from Equation 1.3 can be expressed in terms of the Pauli matrices, shown in Equation 1.2.

$$\sigma^1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma^2 = \begin{bmatrix} 0 & i \\ -i & 0 \end{bmatrix}, \quad \sigma^3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \tag{1.2}$$

$$\gamma^0 = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}, \quad \gamma^i = \begin{bmatrix} 0 & \sigma^i \\ -\sigma^i & 0 \end{bmatrix}, \quad \gamma^5 = \begin{bmatrix} -I & 0 \\ 0 & I \end{bmatrix} \tag{1.3}$$

Under global gauge transformations, the Lagrangian from Equation 1.1 is indeed invariant, but this no longer applies for local $U(1)$ transformations. Enforcing local invariance under transformations, as shown in Equation 1.4, involves the addition of a vectorial gauge field that is massless.

$$\psi(x) \to \psi'(x) = e^{i\alpha(x)}\psi(x) \tag{1.4}$$

The gauge field $A_\mu(x)$ transformation is described in Equation 1.5. On top of that, the covariant derivative $D_\mu(x)$ is also introduced in Equation 1.6.

$$A_\mu(x) \to A'_\mu(x) = A_\mu(x) + \frac{1}{e}\partial_\mu\alpha(x) \tag{1.5}$$

$$D_\mu(x) = \partial_\mu - ieA_\mu(x) \tag{1.6}$$

Using those two, the field strength tensor can be expressed as in Equation 1.7, which is the last piece needed before defining the QED Lagrangian.

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \tag{1.7}$$

With local $U(1)$ symmetry taken care of, the QED Lagrangian is expressed by Equation 1.8

$$\mathcal{L} = \bar{\psi}\left(i\gamma^\mu D_\mu - m\right)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \tag{1.8}$$

A nicer way of representing the QED Lagrangian is shown in Equation 1.9, using the notations introduced so far as well as the electromagnetic charge current density j^μ defined in Equation 1.10.

$$\mathcal{L} = \underbrace{\bar{\psi}\left(i\gamma^\mu\partial_\mu - m\right)\psi}_{\text{free lagrangian}} - \underbrace{ej^\mu A_\mu}_{\text{interaction term}} - \underbrace{\frac{1}{4}F_{\mu\nu}F^{\mu\nu}}_{\text{kinetic term}} \tag{1.9}$$

Symmetry, according to Noether's theorem [34], is associated with a conserved quantity, which in QED is the electromagnetic charge. This charge, represented as q in the QED Lagrangian, can be calculated by integrating j^0 (from Equation 1.10) over the spatial coordinates.

$$j^{\mu} = \bar{\psi}\gamma^{\mu}\psi \tag{1.10}$$

QED, thus describes the interactions between the Dirac fields ψ and the photon field $A_{\mu}(x)$, which emerged as a consequence of $U(1)$ symmetry.

Strong Interaction

Quarks and gluons carry a different type of charge, colour charge, which allows them to interact via the strong force. Modelling for this type of interaction is provided by Quantum Chromodynamics (QCD), founded on Yang-Mills theory [27], which extends the QED formalism to groups that are not abelian. Within this framework, the symmetry group is $SU(3)$, which, unlike in QED, is non-abelian. Since there are eight group generators associated with the $SU(3)$ Lie algebra, just as many gauge fields require being introduced.

$$\psi(x) \rightarrow \psi'(x) = U(x)\psi = e^{i\alpha_a(x)t_a}\psi \tag{1.11}$$

Fermionic fields associated with flavour quarks are represented by **triplets** – three-dimensional vectors. The number of vector components comes from the three colour states: red, green and blue. Under a transformation $U(x)$, the quark field ψ behaves as stated in Equation 1.11. The Gell-Mann matrices t_a represent the group generators, having the commutation relationship: $[t_a, t_b] = if_{abc}t_c$.

Similarly to the QED's case, new fields are introduced: the gluon fields, as well as a covariant derivative, defined in Equation 1.12.

$$D_{\mu} = \partial_{\mu} + ig_s t_a G_{\mu}^a \tag{1.12}$$

The requirements for how the gluon fields behave under a transformation $U(x)$ are summarized by Equation 1.13.

$$G_{\mu}^a \rightarrow G_{\mu}'^a = U(x)G_{\mu}^a t^a U^{\dagger}(x) + \frac{i}{g_s}(d_{\mu}U(x))U^{\dagger}(x) \tag{1.13}$$

Equation 1.14 accounts for the previously introduced requirements and defines the field strength tensor. It depends on f_{abc} which are structure constants, generalizing the asymmetry of the Levi-Civita symbol ϵ_{ijk} to the $SU(3)$ group.

$$G_a^{\mu\nu} = \partial_{\mu}G_a^{\nu} - \partial_{\nu}G_a^{\mu} - g_s f_{abc}G_b^{\mu}G_c^{\nu} \tag{1.14}$$

Lastly, the QCD Lagrangian density can be derived, as in Equation 1.15, using the previously shown elements.

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi - \frac{1}{4}G_a^{\mu\nu}G^a_{\mu\nu} \tag{1.15}$$

A noteworthy aspect of QCD is the presence of self-interaction, founded on the fact that the gluons also carry colour charge themselves.

Electroweak Interaction

The remaining fundamental interaction is called the weak force and manifests itself as nuclear processes such as beta decay. Theoretical physics was, however, able to formalize it under a unified framework that includes electromagnetism as well [27, 30–32]. This is called Electroweak (EW) theory.

In the context of weak interactions, the symmetry group is $SU(2)$ and introduces the conserved quantum number T_3 called **weak isospin**. EW theory is *chiral*, which requires treating separately the two components of the fermionic fields. Left-handed fermions, with $T_3 = 1/2$, are represented by doublets ψ_L while the right-handed ones, $T_3 = 0$, are singlets ψ_R. They behave differently under $SU(2)_L$ and $U(1)_Y$ local transformations. The covariant derivative is consequently acting on the left-handed field according to Equation 1.16 and on the right-handed fields as shown in Equation 1.17

$$D_\mu\psi_L = (\partial_\mu + ig\frac{\sigma_i}{2}W_\mu^i + ig'\frac{Y}{2}B_\mu^i)\psi_L \tag{1.16}$$

$$D_\mu\psi_R = (\partial_\mu + ig'\frac{Y}{2}B_\mu^i)\psi_R \tag{1.17}$$

Pauli matrices σ are used in Equation 1.16, while g and g' are coupling constants for the W^i and B_μ boson fields. The field strength tensors are shown in Equations 1.18 and 1.19. In Equation 1.19 ϵ_{ijk} represents the Levi-Civita symbol.

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu \tag{1.18}$$

$$W^i_{\mu\nu} = \partial_\mu W_i^\nu - \partial_\nu W_i^\mu - \epsilon_{ijk}W_j^\mu G_k^\nu \tag{1.19}$$

Accounting for those, the EW Lagrangian can be put together as Equation 1.20 shows, where the sum over j covers the L doublet and two R singlets.

$$\mathcal{L} = \sum_{j=1}^{3} \bar{\psi}_j [i\gamma^\mu D_\mu] \psi_j - \frac{1}{4} W_i^{\mu\nu} W^i_{\mu\nu} - \frac{1}{4} B^{\mu\nu} B_{\mu\nu} \tag{1.20}$$

Using those ingredients, the massive $W^\pm$ boson arises from linear combinations between W_1 and W_2 as shown in Equation 1.21.

$$W^\pm = \frac{1}{\sqrt{2}} (W_1 \mp i W_2) \tag{1.21}$$

Lastly, by applying a rotation of θ the massless vector field A associated with the photon as well as the massive weak neutral Z can be recovered. Equation 1.22 shows this last step, where θ is known as the Weinberg angle [30].

$$\begin{pmatrix} A \\ Z^0 \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} B \\ W_3 \end{pmatrix} \tag{1.22}$$

1.1.3. The Brout–Englert–Higgs Mechanism

One of the most important additions to the SM is embodied by the Higgs boson. The Electroweak (EW) Lagrangian is notably missing masses for its gauge fields and the reason for it lies within the $SU(2) \times U(1)$ symmetry. Adding boson masses would break this gauge symmetry while adding fermion masses is forbidden by the chiral symmetry requirement. The Higgs mechanism involves *spontaneous symmetry breaking* by separating the local symmetries of the Lagrangian from the symmetries of the quantum vacuum state (lowest possible energy) [35].

At the basis of this mechanism, the Higgs field, which is a complex $SU(2)$ doublet, is defined in Equation 1.23.

$$\begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \tag{1.23}$$

The Lagrangian for such a field can be created as Equation 1.24 shows, with $V(\phi)$ representing the potential energy.

$$\mathcal{L} = (D_\mu \phi)^\dagger (D^\mu \phi) - V(\phi) \tag{1.24}$$

Choosing the most general renormalisable form for the potential, one could represent it as it is shown in Equation 1.25.

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda(\phi^\dagger \phi)^2 \tag{1.25}$$

It is apparent how the sign of μ^2 from Equation 1.25 has a strong impact on the potential's shape. A negative μ^2 allows for an infinite number of minimal energy states, encircling the origin of the complex plane. Figure 1.2 gives a visual representation of the Higgs field in this scenario. In the opposite scenario where $\mu^2 \geq 0$ has a single lowest energy state at $\phi = 0$, which yields a QED-equivalent theory.

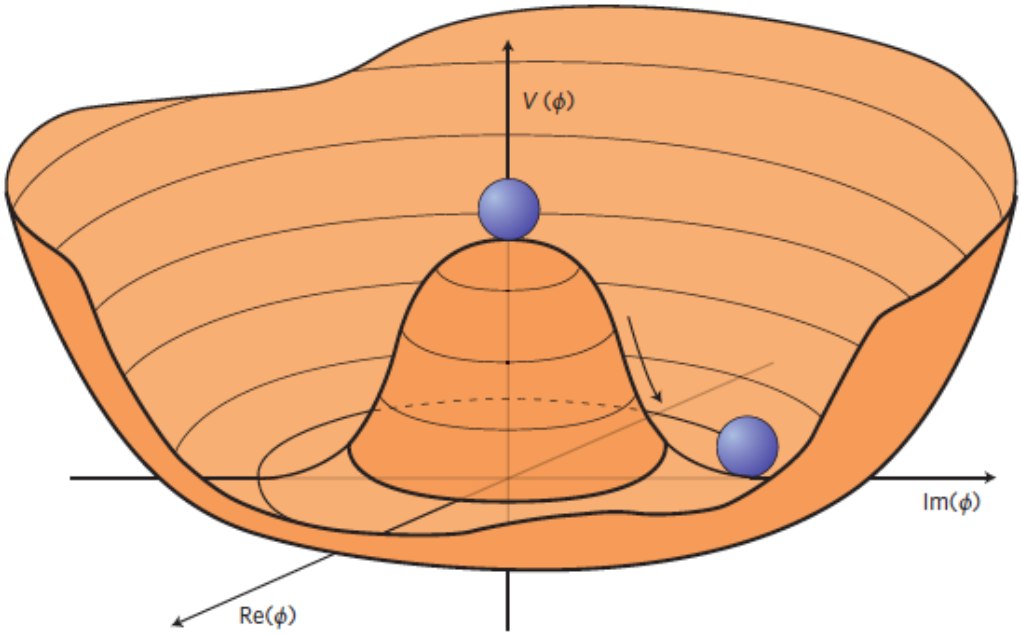


Figure 1.2.: The shape of the Higgs potential relative to the real and imaginary parts of the Higgs field's ϕ_0 component [36]

.

$$\phi_0 \equiv \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \tag{1.26}$$

$$\langle \phi \rangle = \sqrt{\frac{-\mu^2}{\lambda}} \tag{1.27}$$

Equation 1.27 shows the non-zero vacuum expectation value $\langle \phi \rangle$ associated with the minima of the Higgs field, which is shown in Equation 1.26, where $v = \frac{-\mu^2}{\lambda}$. States at this point of minima break the $SU(2) \times U(1)$ gauge symmetry.

The Lagrangian takes the form represented in Equation 1.28 after the chosen form of the potential is substituted in.

$$\mathcal{L} = (\partial_\mu \phi)^\dagger (\partial^\mu \phi) - \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2 \tag{1.28}$$

Introducing the requirement of $SU(2)$ gauge invariance implies the covariant derivative taking the form shown in Equation 1.29.

$$D_\mu = \partial_\mu + ig\tau_a W_\mu^a \tag{1.29}$$

Thus, the three boson fields W_μ^i will transform described in Equation 1.31 under the infinitesimal transformation from Equation 1.30

$$\phi(x) \rightarrow \phi'(x) = (1 + \alpha(x)\frac{\tau}{2})\phi(x) \tag{1.30}$$

$$W_\mu \rightarrow W_\mu - \frac{1}{g}\partial_\mu \alpha - \alpha \times W_\mu \tag{1.31}$$

Thus, Equation 1.32 describes the gauge-invariant Lagrangian, where $W_{\mu\nu} = \partial_\mu W_\nu - \partial_\nu W_\mu - gW_\mu \times W_\nu$ are the kinetic energy terms associated with the gauge fields.

$$\mathcal{L} = (\partial_\mu \phi + \frac{ig}{2}\tau \cdot W_\mu \phi)^\dagger (\partial^\mu \phi + \frac{ig}{2}\tau \cdot W^\mu \phi) - \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2 - \frac{1}{4} W_{\mu\nu} W^{\mu\nu} \tag{1.32}$$

Following through, the field minimum shown in Equation 1.26 is substituted in the Lagrangian expression from Equation 1.32. Using the absolute square notation as shorthand for the product between the object and its adjoint, Equation 1.33 shows the relevant term from the lagrangian after the substitution.

$$\begin{aligned} \left| ig\frac{1}{2}\tau W_\mu \phi \right|^2 &= \frac{g^2}{2} \left| \begin{pmatrix} W_\mu^3 & W_\mu^1 - iW_\mu^2 \\ W_\mu^1 + iW_\mu^2 & W_\mu^3 \end{pmatrix} \begin{pmatrix} 0 \\ v \end{pmatrix} \right|^2 \\ &= \frac{g^2 v^2}{2} [(W_\mu^1)^2 + (W_\mu^2)^2 + (W_\mu^3)^2] \end{aligned} \tag{1.33}$$

Relating Equation 1.33 to how boson mass terms regularly look, the mass can be recovered as $\frac{1}{2}gv$. Through this mechanism of spontaneous symmetry breaking the Standard Model is able to generate masses for gauge fields, explaining the massive bosons: $W^\pm$ and Z^0.

1.2. Beyond the Standard Model

In spite of the huge success and accuracy of QED, QCD and **Electroweak** theories, the SM is still missing some important pieces before becoming a *Theory of Everything*. There are several still unexplained observed phenomena that the Standard Model is unable to account for adequately.

1.2.1. Issues with The Standard Model

- One of the four known fundamental forces, namely gravity, is not included in the SM. Gravity is best described by general relativity, which assumes that spacetime is curved. However, all attempts to model gravity as a QFT are yielding divergences. The inclusion of gravity would also require finding a particle mediating gravitational interactions, which, so far, has not been identified by any state-of-the-art experiments.

- Neutrinos are considered massless by the SM, which was experimentally disproved by studies on neutrino oscillations [1, 2]. Even though the SM is flexible enough to allow expansions that would fix neutrino masses, the specific mechanism through which neutrinos acquire mass is still proving to be elusive.

- In terms of fine-tuning the SM parameters, there is the *hierarchy problem*. When calculating the Higgs mass, contributions from all energy scales ought to be taken into account. The corrective terms are proportional to their respective scales so it would be reasonable for the Higgs mass to not be far off the Planck mass. Experimental measurements, however, show the Higgs mass to be less than this expectation by an order of 10^{17}. Bringing those two numbers in line implies cancellations that are often not considered natural.

- There is strong indirect evidence suggesting dark matter to be part of our universe, but the SM is unable to provide any explanation for it. The difficulty in observing candidates for dark matter comes from the fact they are inherently unable to interact electromagnetically. This should still allow other experimental avenues for detection, however, the SM does not provide any insights into what dark matter is or how it behaves.

- Recent studies in b-physics [37] are casting doubt on *lepton universality*. The SM predicts all charged leptons to behave identically in terms of weak interaction strength. However, experimental measurements of the ratio between $B^+ \to K^+\mu^+\mu^-$ and $B^+ \to K^+e^+e^-$ branching fractions seem to suggest a different behaviour, unaccounted for by the current formulation of the Standard Model.

1.2.2. Possible Extensions

Although the Standard Model provides the best available description of fundamental physics, considering that it is not yet fully complete, there are several theories designed to fill out its few remaining gaps. Those aptly named Beyond Standard Model (BSM) theories aim to introduce alterations that allow solving some of the SM issues or provide entirely new, more general frameworks in an attempt to generalize its current form.

One such promising theoretical development is Supersymmetry (SUSY) [3]. It introduces boson-fermion symmetry which implies effectively doubling the number of particles in the SM by adding a *super-partner* to every existing one. Attempts to model SUSY into the SM while limiting as much as possible the number of particles and interactions manifested as the following extensions: Minimal Supersymmetric Standard Model (MSSM) [4] and Next-to- Minimal Supersymmetric Standard Model (NMSSM) [5]. SUSY addresses the hierarchy problem as well as providing dark matter candidates, but, at the time of writing, no definitive experimental evidence has been found confirming it.

In a similar fashion that the electromagnetic and weak interactions can be modelled by a unified framework, models of the Grand Unified Theory (GUT) type, attempt to bring together all fundamental interactions under a *unified* framework. They represent an entire class of BSM models, each variant tailored to tackle slightly different phenomena. While they have arguably fallen out of favour in recent times there is still interest and potential for experimentally testable predictions [38].

An attempt to fix the apparent incompatibility of the SM and general relativity, comes from Loop Quantum Gravity (LQG) [6]. It attempts to incorporate gravity through geometric treatment rather than introducing it as a fundamental force. This is achieved by describing space and time at the most fundamental level in terms of finite loops. LQG borrows ideas from **string theory** but does not require as many dimensions, usually being formulated in terms of three or four dimensions. Some formulations also account for SUSY while others don't.

2. The ATLAS detector at the LHC

European Organization for Nuclear Research (CERN) is one of the leading organizations doing fundamental research in High-Energy Particle Physics (HEPP). In this field of study, particle accelerators play an important role in finding experimental evidence since the energies required are far beyond what we encounter in our everyday life. For this reason, CERN built the world's biggest particle accelerator to date, the Large Hadron Collider (LHC). Since 2008 it was used with great success towards deepening our understanding of the fundamental laws governing our universe, one of the most notorious discoveries being the Higgs boson. This chapter gives a general overview of the LHC and details the various aspects of the ATLAS Detector with a heightened focus on the features closely related to this thesis' main topics of research. Section 2.1 describes the LHC in broader terms, preparing the ground for the ATLAS Detector description from section 2.2. The New Small Wheel (NSW) is one of the main research topics presented in this work, so section 2.3 is dedicated entirely to its hardware description. All technical documentation is sourced from the official design reports: *LHC Design Report* [39, 40], 'The ATLAS Experiment at the CERN Large Hadron Collider' [41], *ATLAS Muon Spectrometer: Technical Design Report* [8] and *ATLAS New Small Wheel: Technical Design Report* [42].

2.1. The Large Hadron Collider

Spanning 27 kilometres, the LHC complex incorporates numerous subsystems in order to achieve its many functionalities, as illustrated in figure 2.1. Apart from the main ring, there are several other elements to the accelerator complex, providing increasingly larger energies to the protons before they could be injected into the LHC.

There is a set succession of several apparatuses designed to speed up particle beams before entering the LHC. When accelerating protons, hydrogen gas is used as the primary source. Using an electric field, the electrons are stripped away from the nucleus and the resulting protons are accelerated up to an energy of 50 MeV by the linear accelerator: *Linac 2*, the first in this chain. From here the beam ends up in the Proton Synchrotron Booster (BOOSTER), where it is accelerated up to 1.5 GeV. Next up in the acceleration chain is the Proton Synchrotron (PS) that increases the energy furthermore towards 25 GeV. The resulting beam is injected into the Super Proton Synchrotron (SPS) and here the energies obtained are around 450 GeV [40].

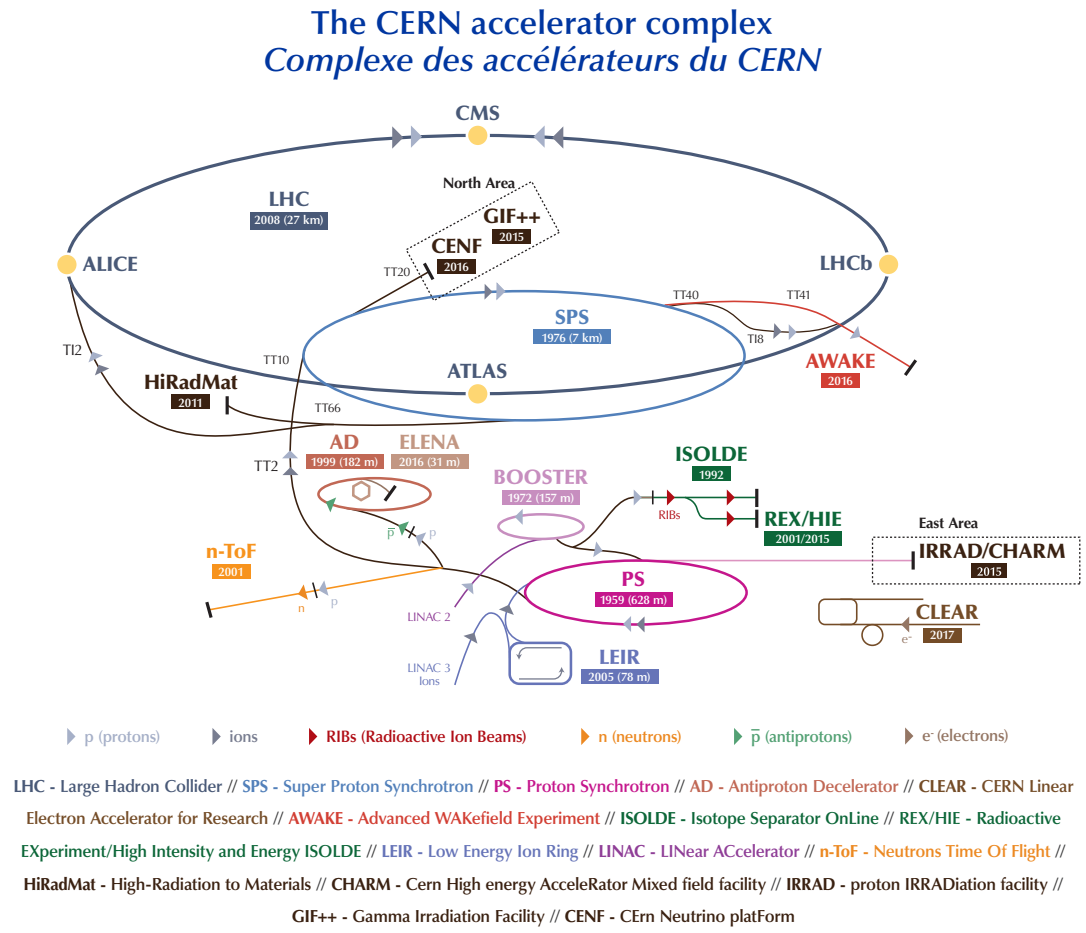


Figure 2.1.: Schematic of the LHC [43]

Besides protons collision, CERN also uses heavy lead ions for other experiments such as ALICE. From a source of vaporised lead, the ions are accelerated by *Linac 3* and then sent to the Low Energy Ion Ring (LEIR). After this point, the lead ions are injected into the PS and from there they follow the same steps in the acceleration chain as protons.

The last element in the acceleration chain is the LHC ring. Two beams of particles from the SPS are injected into the LHC: one travels clockwise and the other anticlockwise. Proton beams needed about 20 minutes to reach their Run 2 maximum energy of 6.5 TeV and after that point, they are brought into collision at one of the main sites where the event is recorded by the available particle detectors.

One of the challenges posed by the centre-of-mass energy target of $\sqrt{s} = 14$ TeV is bending the beams across the ring. In order to achieve this, a lot of engineering effort went into designing the superconducting dipole magnet system shown in Figure 2.3. Those are

so-called *two-in-one* magnets since the system involves two separate dipoles with field orientations in opposite directions. For the LHC's bending radius of $R = 2804\,\text{m}$, the necessary field for bending 7 TeV beams is about $B = 8.33\,\text{T}$ [44]. Those numbers are derived using the following approximation:

$$E = 0.3BR \tag{2.1}$$

Equation 2.1, implies the energy expressed in GeV, the magnetic flux density B in *Tesla* and R in meters. Additionally, to ensure beam focusing and stability, quadrupole magnets are also deployed all along the main ring as part of FODO cells. As illustrated in Figure 2.2, this arrangement of quadrupole magnets alternates between focusing the beam closer to the beamline and defocusing it in order to correct its trajectory.

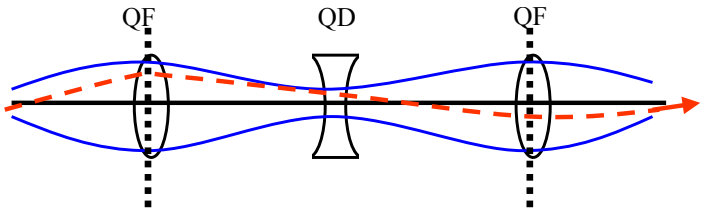


Figure 2.2.: Conceptual drawing of a FODO cell featuring a particle's trajectory with a dashed red line and the beam envelope with a blue solid line. QF are focusing quadrupoles while QD represents a defocusing quadrupole [44]

When adding together all of those magnets alongside the other thousands of multipoles in charge of further improving the beam focusing and all the normal magnets used for squeezing the beams before the collisions, the total count exceeds 9000. Magnets alone represent around 90% of the LHC's cold mass.

Sustaining the required magnetic field necessitates large amounts of electric current, topping up at 11800 A. As such, cooling the superconducting infrastructure to the nominal temperature of 1.5 K is the second major challenge. Pressurized helium II (He_2) is the technical coolant of choice for the LHC's superconducting magnet system. Since some magnets in the straight sections of the ring don't need to operate at their maximum flux density values, they can be operated at 4.5 K. To accommodate those requirements, the cryogenics system has two stages of cooling. The first stage involves liquid helium at 4.5 K, being delivered by eight compressor plants, each one able to maintain the temperature while providing 18 kW of cooling capacity. The magnets are immersed in a pressurized bath filled up with this cooling agent. In tandem, the second stage of the cooling system makes use of the heat exchange pipe (figure 2.3). By exchanging heat between the aforementioned helium bath and this pipe, the temperature drops to 1.5 K. It takes three weeks to completely cool down one of LHC's sectors from room temperature to the nominal temperature for operation.

With all of those elements working together, the LHC is a truly complex machine whose operation has reached increasingly larger collision rates. Considering that the beam is not

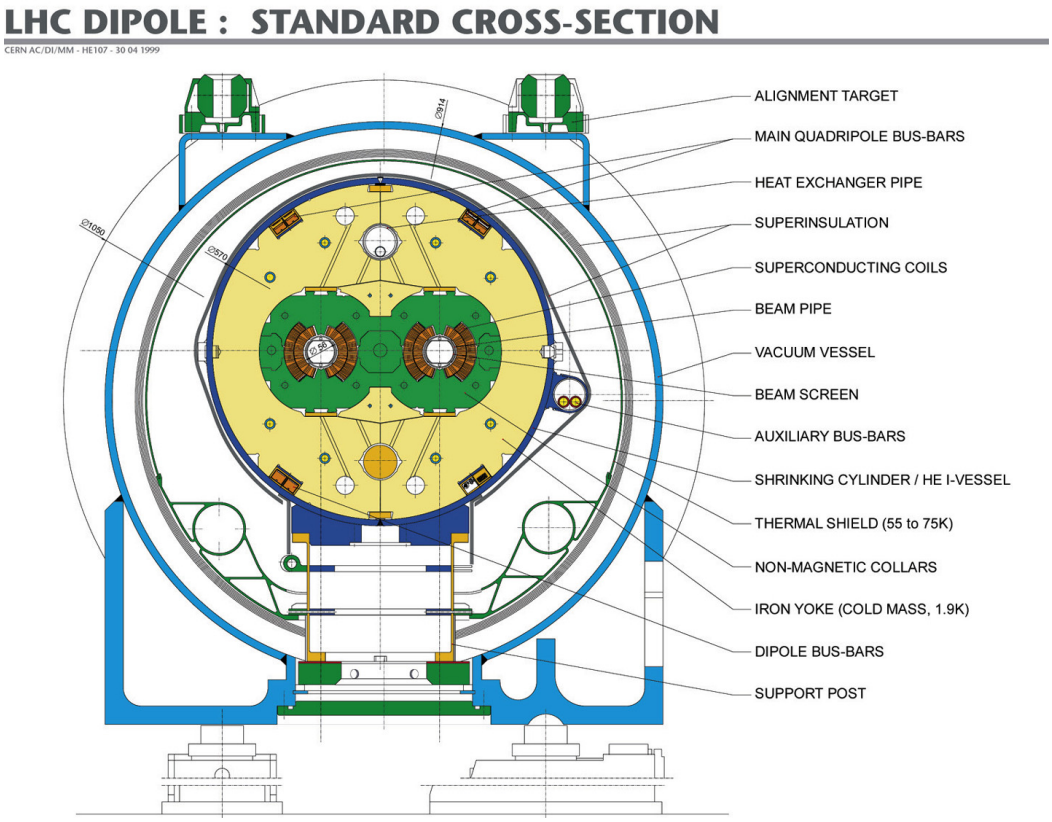


Figure 2.3.: Cross-section of the LHC dipole system [45]

continuous, but rather split into several bunches, the main figure for quantifying those collision rates consists of the **instantaneous luminosity** and is given by the following formula:

$$\mathcal{L}_{\text{inst}} = \frac{n_b^2 N_b f_{\text{rev}} \gamma}{4\pi\epsilon_n \beta^*} F \tag{2.2}$$

In Equation 2.2 n_b is the number of protons per bunch, N_b represents the number of bunches and f_{rev} is the beam's revolution frequency. The other parameters are the Lorentz factor γ, the normalized beam emittance ϵ_n, the value of the beta function (one of the Twiss parameters) at the IP β^* and the geometric factor F. If the collisions are head-on, F will take the value of 1, but it will decrease with higher crossing angles θ_c. For a *Gaussian* beam with the corresponding Root Mean Squared (RMS) of its size in the crossing plane denoted by σ_x and the longitudinal bunch length as σ_z, the expression for F is described by the Equation 2.3.

$$F = \frac{1}{\sqrt{1+\Phi^2}} \tag{2.3}$$

where $\Phi = \theta_c \sigma_z / 2\sigma_x$ [39]. A list of the LHC's design parameters relevant in pp collisions could be found in Table 2.1.

Parameter	Symbol	Value	Units
Beam energy	E	7	TeV
Bunch spacing	τ_b	25	ns
Protons per bunch	n_b	10^{11}	
Bunches per beam	N_b	2808	
Revolution frequency	f_{rev}	11245	s^{-1}
Peak luminosity	$\mathcal{L}_{\text{peak}}$	10^{34}	$\text{cm}^{-2}\,\text{s}^{-}1$
Normalized beam emittance	ϵ_n	3.75	μm rad
Beta function at IP	β^*	0.55	m
Crossing angle at IP	θ_c	300	μrad
Longitudinal bunch length	σ_z	0.075	m
RMS beam size	σ_x	16	μm

Table 2.1.: Design parameters of the LHC's main accelerator ring for pp collisions [39]

The luminosity figure is useful in determining the number of events where a specific process of interaction takes place. Knowing the cross-section of the process σ_p, the resulting number of events N is given by Equation 2.4.

$$N = L\sigma_p \tag{2.4}$$

For those kinds of calculations, we use the integrated luminosity L rather than the instantaneous $\mathcal{L}_{\text{inst}}$, as defined in Equation 2.5.

$$L = \int \mathcal{L}_{\text{inst}} dt \tag{2.5}$$

Finally, another key parameter is the **pile-up**, describing the number of simultaneous processes taking place within the same bunch crossing. Since there are 10^{11} protons in a bunch, it is expected to end up with more than one pp collision in any given event. Equation 2.6 shows how the pile-up μ relates to the instantaneous luminosity (as defined in Equation 2.2) and the inelastic cross section σ_{inel}.

$$\mu = \frac{\mathcal{L}_{\text{inst}}\sigma_{\text{inel}}}{f_{\text{rev}}} \tag{2.6}$$

The LHC's collision sites for the resulting beams of protons (or heavy ions) are represented by the main experiments listed below:

- A Large Ion Collider Experiment (ALICE)

  As the name suggests, in the ALICE experiment heavy-ion collisions are primarily studied. Its main purpose is to study the strongly interacting hadronic matter at the highest energies available. Under these conditions, quark-gluon plasma can be produced and studied.

- A Toroidal LHC Apparatus (ATLAS)

  The ATLAS experiment could be best described as being general-purpose. It is deployed in the study of both *proton-proton* (*pp*) and *Nucleus-Nucleus* (A + A) collisions and facilitates studies for a wide range of modern-day physics problems such as searches for dark matter candidates and BSM physics processes.

- Compcat Muon Solenoid (CMS)

  This is the other general-purpose experiment being held at LHC. Although it has similar goals as the ATLAS experiment, the technical design is entirely different.

- LHCb

  The main focus of LHCb is the study of *b*-physics. Having this sole purpose removes the need for full coverage of the collision point. Thus, LHCb detectors are placed in such a way as to find preponderantly forward-moving particles.

This thesis' main subjects revolve entirely around the ATLAS experiment, which is going to be described in more detail in the following section.

2.2. The ATLAS Detector

The ATLAS Detector has been built as a general-purpose apparatus tasked with probing *proton-proton* (*pp*) and *Nucleus-Nucleus* (A + A) collisions.

Many studies are being conducted by the ATLAS Collaboration covering a wide range of physics. The scientific program includes testing the predictions of the Standard Model (SM) as well as precision measurements for its physical parameters such as particle's masses and mean lifetimes, but also coupling constants for the fundamental interactions. The high energies provided by the LHC allow the ATLAS Detector to investigate the properties of even of the most massive particles like *top* quark and the Higgs boson. Another important area of study is the search for Beyond Standard Model (BSM) processes. Several analyses were conducted, attempting to uncover evidence for particles predicted by BSM models such as Supersymmetry (SUSY). Some other BSM topics of interest are imbalances between the creation of matter and antimatter, unification of gravity with the SM and the search for dark matter candidates.

Relative to the interaction point (IP), the ATLAS Detector is forward-backwards symmetric, coving almost a full solid angle. It has complex granularity, involving many sub-systems carrying out different detection purposes. For easier representation, those various sub-systems could be thought of as layers, since their disposition relative to the centre is related to their functionality. Closest to the IP, we have the inner detector systems involved in tracking the particles resulting from the collision. Here high-precision pixel and strip

detectors are deployed for momentum and vertex measurements. Farther away from the IP, the next sub-system is the Liquid Argon (LAr) Electromagnetic (EM) calorimeter whose main task involves high-resolution measurements of the energy and position. Farther still, there is the hadronic *Tile* calorimeter, covering mainly the central region of the ATLAS Detector. The two calorimeters are surrounded by the muon spectrometer consisting of three layers of high-precision tracking chambers achieving excellent precision in muon momentum resolution measurements.

2.2.1. Coordinate System

Given the overall shape of the ATLAS Detector, a cylindrical symmetry could easily be ascribed to it. In order to avoid any potential confusion a clear and consistent coordinate system is necessary. Within a Cartesian representation of the detector, an obvious choice for the origin would be the interaction point (IP). Regarding the axes' orientation, the beamline is conventionally labelled as the z direction. Perpendicular to it, the x axis intersects the centre of the LHC ring in the positive direction and the y axis is perpendicular to the resulting xz plane, pointing towards the sky. While this Cartesian system may be useful for describing the detector itself with its many constituent elements, when it comes to event analysis, taking advantage of the rotational symmetry along the beamline simplifies the representation a lot. A cylindrical coordinate (r, θ, ϕ) representation of physics objects may be simpler than a Cartesian one, but four-vectors are the main mathematic tools used in the study of relativistic kinematics. The momentum of any physics objects created from collisions can be described by a four-vector $\mathbf{p} = (E, p_x, p_y, p_z)$ and the best system of coordinates for working with such objects would provide Lorentz invariant representations. One such system is a spherical coordinate representation where the four-momenta vectors would take the form of $\mathbf{p} = (E, p_T, \eta, \phi)$, where p_T would be the transverse momentum, η the pseudorapidity and ϕ the azimuthal angle. Equations 2.7-2.10 show the formulas for deriving the components of the Lorentz invariant four-momentum.

$$p_{\mathrm{T}} = \sqrt{p_x^2 + p_y^2} \tag{2.7}$$

$$\theta = \arctan \frac{p_{\mathrm{T}}}{p_z} \tag{2.8}$$

$$\eta = -\ln(\tan \frac{\theta}{2}) \tag{2.9}$$

$$\phi = \arctan \frac{p_x}{p_y} \tag{2.10}$$

Rapidity is often an important variable in particle physics when dealing with relativistic speeds, but it is often harder to compute since it requires precise measurements of the

longitudinal momentum p_z, as shown in Equation 2.11.

$$y = \frac{1}{2} \ln \frac{E + p_z}{E - p_z} \tag{2.11}$$

Fortunately, when a particle is moving very close to the speed of light, in other words when $E \approx |\mathbf{p}|$, the rapidity converges towards the pseudorapidity $y \approx \eta$. Thus, at the energies where the LHC operates, pseudorapidity makes for a good kinematic variable.

$$\Delta R = \sqrt{\Delta\eta^2 - \Delta\phi^2} \tag{2.12}$$

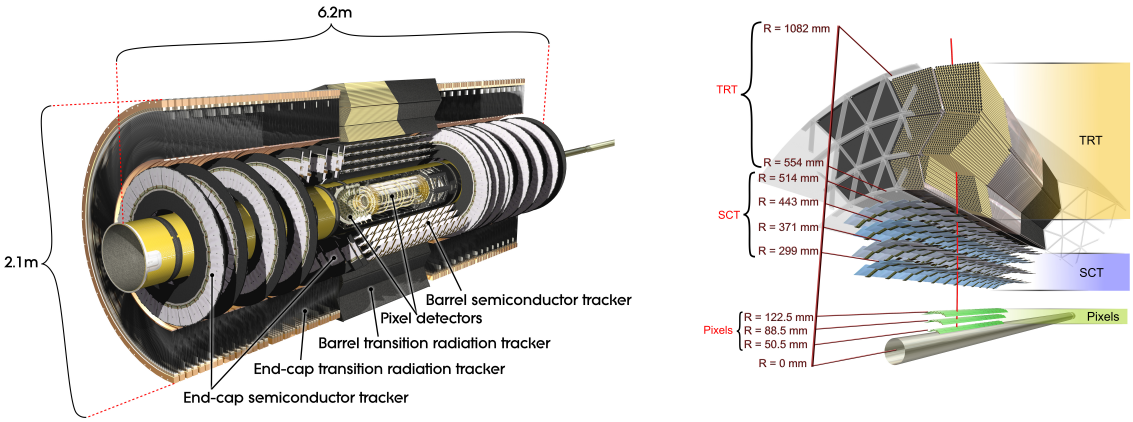
Operating with those variables requires redefining the way of computing spatial distances between four vectors. For this reason, the metric of choice in this coordinate system is the ΔR, showcased in Equation 2.12, where $\Delta\eta$ and $\Delta\phi$ are the differences between the values of η and ϕ, respectively.

2.2.2. Inner Detector

With the original design dating from 1997 [46, 47], the inner detector was always a crucial part of the ATLAS experiment's operation. The main point of this assembly is to perform **tracking**. This means determining the tracks (trajectories) of the **charged** particles resulting from the collision. Magnetic fields ($B = 2T$) are applied in this region to bend the trajectory of those particles, allowing measurements of charge and momenta based on the subsequent curvature. Having those tracks provides vital information in determining the primary interaction point (primary vertex) as well as the subsequent decay vertices, allowing the analysis teams to target specific interactions for their studies. In terms of coverage, tracks are detected across the whole $|\eta| < 2.5$ range. As part of the inner detector, there are three separate sub-systems whose synergistic functionalities achieve the full scope of the inner detector ensemble. From the interaction point outwards the subsumed detectors are the *Pixel* detector, the Semiconductor Tracker (SCT) and the Transition Radiation Tracker (TRT). Layout-wise, Figure 2.4 shows schematic representations of the inner detector, with a full cross-section along the z-direction displayed in Figure 2.4(a) and a cross-section of the $r - \eta$ plane in Figure 2.4(b).

The *Pixel* detector is situated millimetres away from the beamline and originally consisted of three separate layers divided into (50×400) μm pixels. During the first Long Shutdown an additional layer was introduced: the Insertable B-Layer (IBL). This addition was placed closer to the IP than the other three and significantly improved data quality and tracking performance [49].

Next in line, the Semiconductor Tracker (SCT), is made up of four layers all along the barrel region, while at the end-caps it takes the form of parallel discs, nine on each side. The region covered by this detector is where the aforementioned magnetic field is deployed through



(a) Full inner detector layout

(b) $r - \eta$ plane cross section

Figure 2.4.: Computer generated representations of the ATLAS inner detector system [48].

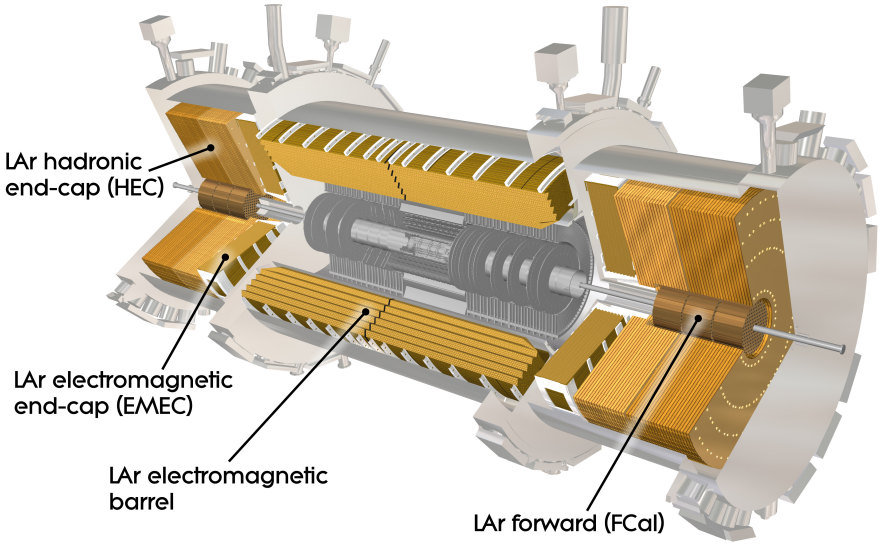
solenoid magnets. As opposed to its *Pixel* counterpart, the detector elements deployed in the SCT are strips rather than pixels. This allows coverage of a much bigger area.

At the outer region of the inner detector, the Transition Radiation Tracker (TRT) comprises gas-filled straw tubes as the active detector volume. This only allows for position measurements in the $r - \phi$ plane but provides a large number of points per track (around 36). While this part of the inner detector provides a smaller spatial resolution, namely $\approx 130\,\mu$m, it is still instrumental in tasks like electron identification, mitigating fakes from charged mesons.

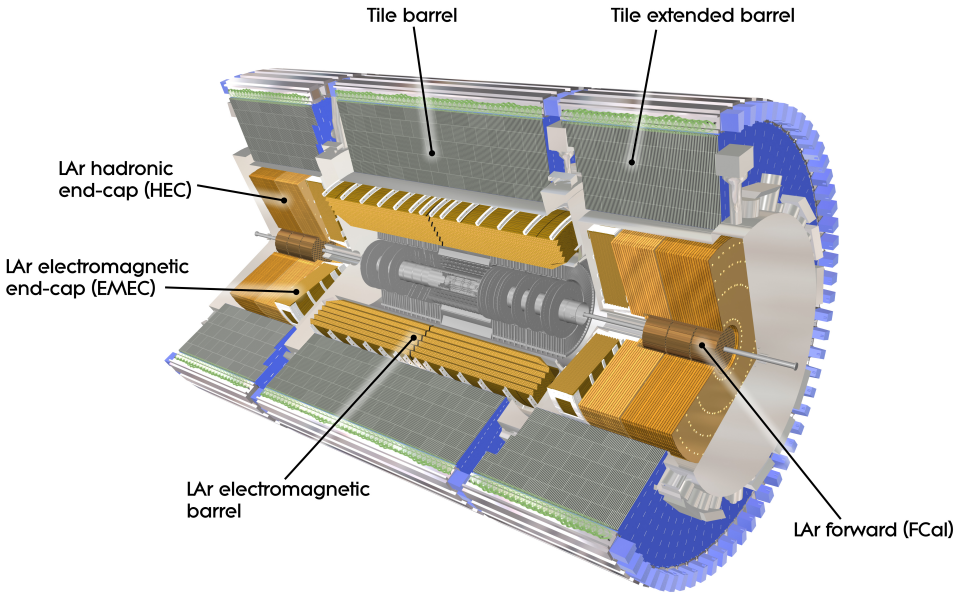
2.2.3. Calorimeters

For precise energy measurements as well as Missing Transverse Energy (MET) estimation, ATLAS commissioned a complex calorimeter system [47]. While there are several types of detector technologies deployed within ATLAS' calorimeters, they all rely on the principles of sampling calorimetry. Coverage in terms of pseudorapidity range extends up to $|\eta| < 4.9$. There are two main calorimeter systems at play, each tasked with a different purpose. For physics objects of interest such as electrons and photons, the Electromagnetic (EM) calorimeter provides adequate granularity for measuring their energy. On the other hand, jets and heavier particles the likes of hadrons are not stopped by the LAr and end up in the *Tile* hadronic calorimeter where they are fully stopped and their energy is recorded. Layouts of the two calorimeter systems are displayed in Figure 2.5

As suggested by its name, the Liquid Argon (LAr) calorimeter uses liquid Argon as the active detection media alongside lead plates for absorbing the energy of highly energetic particles.



(a) Liquid Argon (LAr) calorimeter [50]



(b) The *Tile* calorimeter alongside the LAr EM calorimeter [51]

Figure 2.5.: Computer generated representations of the ATLAS calorimeters.

The liquid Argon gets ionized by incoming particle showers resulting in the interaction between the particles of interest and the lead plates. Structurally, multiple elements are making up the EM calorimeter. In the barrel region, there are two LAr *half-barrels* deployed in the region covering $|\eta| < 1.475$. At the end-caps, there is a separate structure called LAr EM end-cap (EMEC) which covers the pseudorapidity interval $1.375|\eta| < 3.2$. LAr-based elements are also deployed for hadronic calorimetry as the LAr hadronic end-cap (HEC) and for calorimetry in the *forward* region as the LAr forward calorimeter (FCal). Figure 2.5(a) shows the placement of all subsystems using LAr technologies for calorimetry.

The other big part of the calorimeter system of ATLAS is represented by the *Tile* calorimeter. This one wraps around the LAr detectors, being deployed for *hadronic* calorimetry exclusively. It also uses plates for absorption, but the material is steel instead of lead. The detection media consists of scintillating tiles deployed as in the layout highlighted in Figure 2.5(b), where three distinct regions could be distinguished: one central barrel that is fixed and two extended barrels placed on either side of the detector.

2.2.4. Muon Spectrometer

Designing a complex system like the ATLAS Detector requires impressive amounts of resources and planning. Akin to many other of its major parts, the Muon Spectrometer's specifications were meticulously detailed in the initial Technical Design Report (TDR) from 1997 [8].

Provided the importance of correctly identifying high-momentum final-state muons at the LHC, the ATLAS Collaboration has designed the muon spectrometer with high-resolution trigger and momentum measuring capabilities. The spectrometer uses the magnetic deflection of muon tracks in an arrangement of three sizable air-core superconducting toroid magnets, in order to perform both triggering and p_{T} measurements. There is the large Barrel Toroid (BT) encompassing the hadronic calorimeter which is responsible for muon track bending in the pseudorapidity range $|\eta| \leq 1.0$, while for the $1.4 \leq |\eta| \leq 2.7$ interval, the same function is provided by two smaller End-Cap Toroids (ECTs), placed at both ends of the BT. Both the BT and ECTs are synergizing their effect over the transition region $1.0 \leq |\eta| \leq 1.4$, generating magnetic deflection from their combined influence [8].

During Run 1 and Run 2, the Muon Spectrometer was already making use of several different types of detector technologies. Monitored Drift Tubes (MDT) were the primary instruments for precision measurements across most of the pseudorapidity range. However, at $|\eta| \geq 2$, the MDT's maximum count rate of $150\,\mathrm{Hz/cm^2}$ becomes insufficient, prompting the deployment of special Cathode Strip Chambers (CSC) in the forward region which can be operated reliably at rates up to $1\,\mathrm{kHz/cm^2}$. For triggering the story is similar, with Resistive-Plate Chambers (RPC) serving the barrel region of $|\eta| \leq 2.4$ and Thin-Gap Chambers (TGC)

covering the end-cap region [8]. The general layout of the muon spectrometer is shown in 2.6.

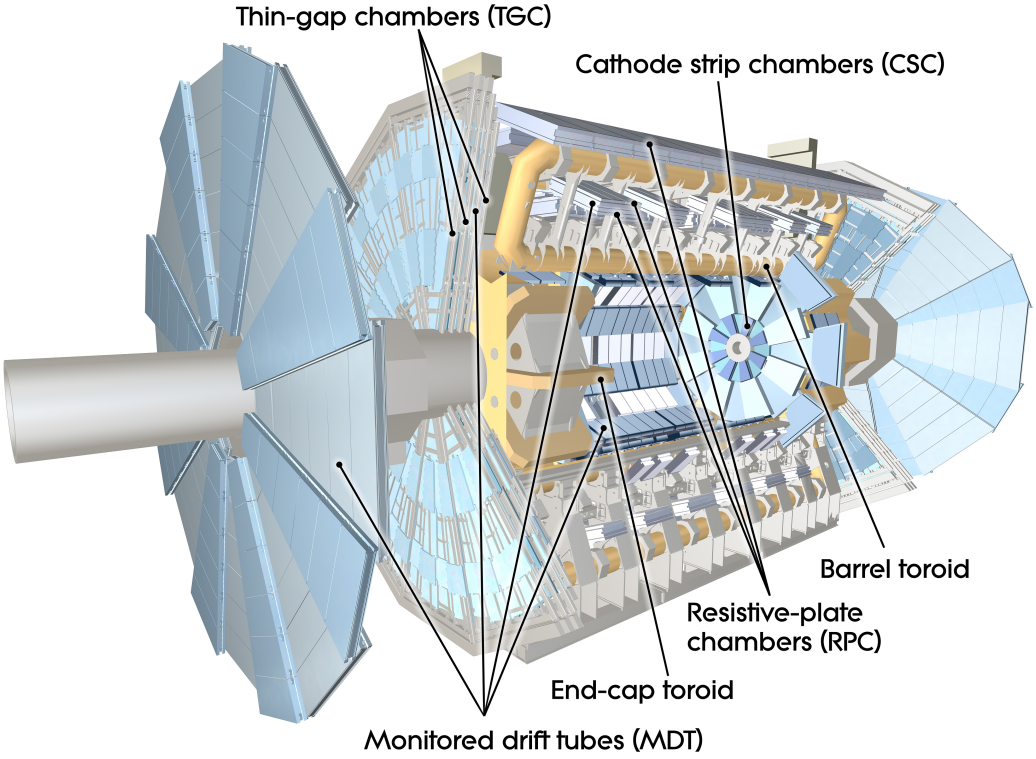


Figure 2.6.: Schematic representation of the ATLAS Muon Spectrometer Subsystems [7] . The barrel and end-cap toroids are labelled alongside some of the main detector technologies used: Thin-Gap Chambers, Cathode Strip Chambers, Resistive-Plate Chambers and Monitored Drift Tubes.

The design decisions shaping the Muon Spectrometer were consequences of the very clear and ambitious set of initial requirements stated in the TDR [8]. In terms of performance, the main goal of this system was to maximize the discovery reach for new physics, while reducing systematic bias as much as possible. This also meant a solid discrimination power against ever-increasing levels of charged and neutral particle backgrounds. From a physics standpoint, such performance would imply:

- Momentum and mass resolutions hovering around 1% and p_{T} resolution as uniform as possible over the entire pseudorapidity spectrum.
- Measurement of a *second coordinate* in the non-bending plane.
- Pseudorapidity coverage targeting $|\eta| \approx 3$ and ample hermeticity
- A selective trigger with multiple operating p_{T} thresholds and strong efficiencies

- Time resolution good enough to identify individual bunch-crossings

And finally, the whole system was foreseen also to be resilient to harsh environmental conditions such as temperature gradients, strong magnetic fields and high levels of radiation.

While the Muon Spectrometer's implementation and operation were successfully meeting the requirements during the first two data-taking runs, some of the systems involved are expected to struggle with the new operating parameters of the LHC throughout future runs. The end-cap region, especially the inner part, necessitated substantial upgrades to be ready for Run 3. Section 2.3 covers the upgrade schedule of the LHC and dives deeper into the changes made to the end-cap regions of ATLAS Detector's Muon Spectrometer.

2.2.5. Trigger and Data Acquisition

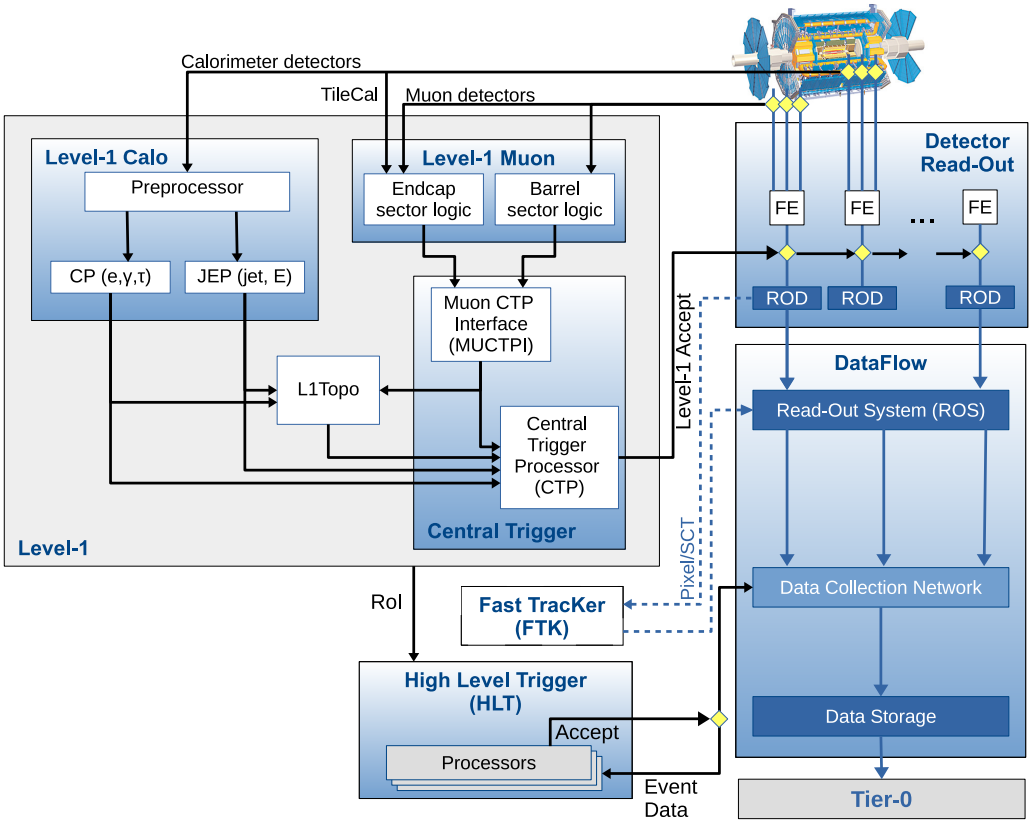


Figure 2.7.: Diagram of the ATLAS trigger system data flow [9].

With all of the different detectors having lots of readout channels and a very high bunch crossing rate, the ATLAS experiment can not physically store every collision event. On top of the technical limitations, there is also the fact that not all physics processes pose the

same amount of research interest. From its very conception [47], the ATLAS Detector was designed to implement preselection mechanisms in order to quickly decide if an event is worth keeping or not. This type of process is commonly known as **triggering**.

The system set up in place for triggering the data acquisition operates at multiple levels with different tradeoffs between the trigger decision's delay and complexity [9]. Figure 2.7 shows a general diagram encompassing the major elements of the Trigger and Data Acquisition (TDAQ) system and the flow of data between them.

At the first level, the Level-1 (L1) trigger system operates on custom hardware (FPGAs and ASICs) which makes it possible to operate at rates as high as the bunch crossing rate of 40MHz. The data sources for the L1 trigger are the *Tile* and EM calorimeters for energy information and the Muon Spectrometer since the presence of muons in the final state is one of the telltale signs of many interesting processes.

After a delay of no more than 2.5 μs, the L1 trigger makes a decision and if it accepts the event, the signal is passed to the Front End (FE) boards that perform broader detector readout based on the Regions of Interest (RoIs) found by the L1. The second stage involves the Level-2 (L2) trigger which is, in turn, part of the High-Level Trigger (HLT) system. With initial preprocessing of the readout data performed by the Readout Drivers (RODs), the L2 performs a more thorough event reconstruction, while using only readouts from the RoIs. This step, akin to anything else falling into the umbrella of the HLT, uses CPUs for processing.

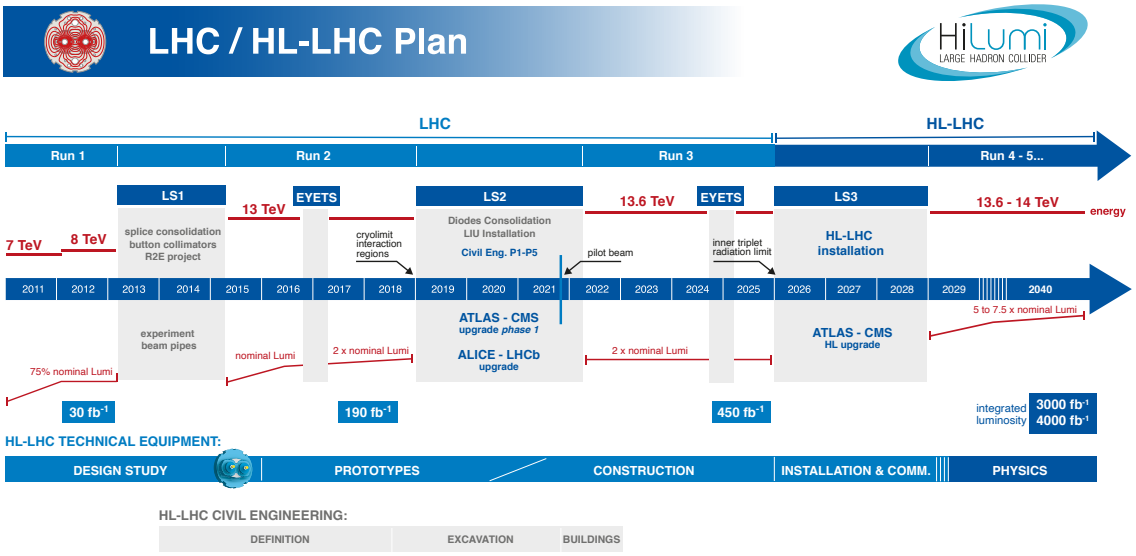


Figure 2.8.: CERN's upgrade path towards the High-Luminosity LHC (HL-LHC) [52].

Finally, after a second positive trigger decision, the HLT deploys its final system: the *Event Filter*. This involves a full event reconstruction with the common analysis framework ATHENA using all available detector data. It is the slowest of all trigger steps but nonetheless important since accepted events would go to mass storage and potentially be used in physics analysis.

2.3. The New Small Wheels

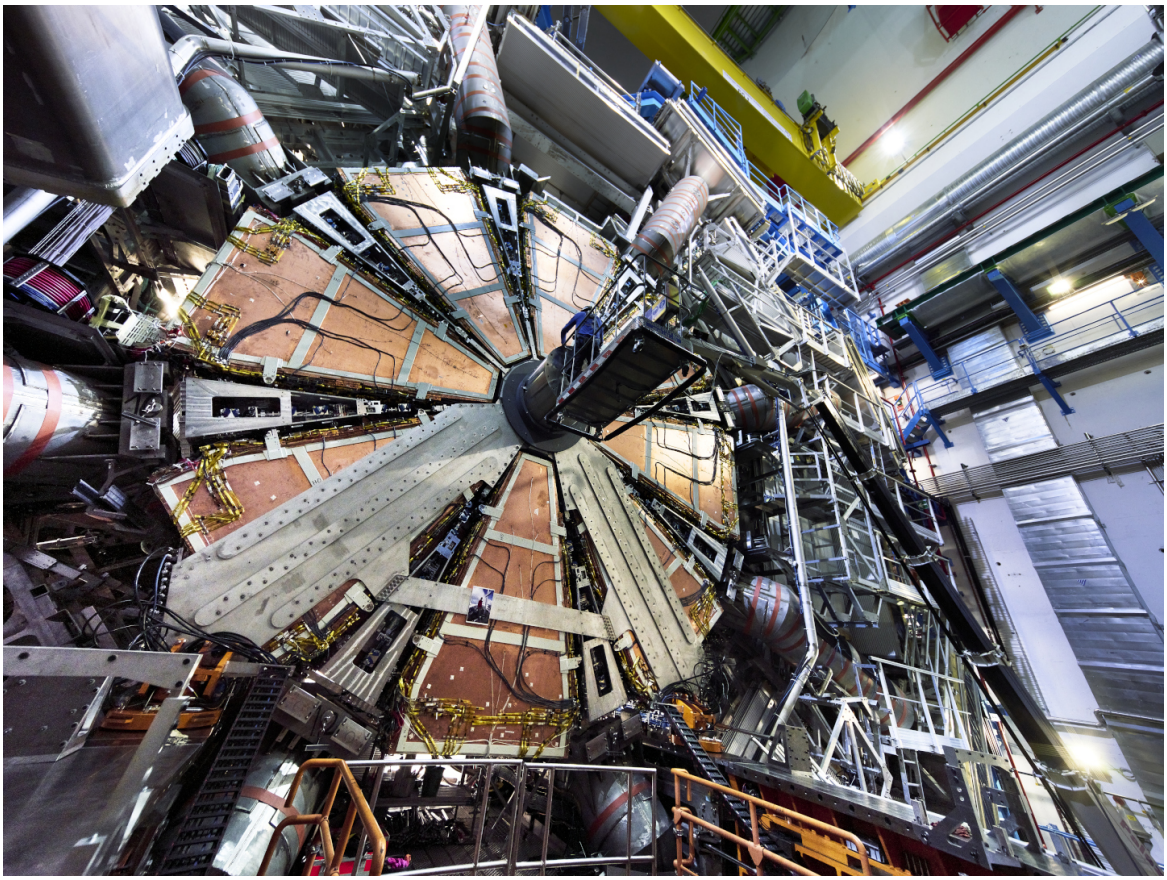


Figure 2.9.: Picture of the **A**- side NSW detector system being installed [53].

The LHC's operation has been consistently and gradually upgraded ever since Run 1. If by the end of 2011, the collider was operating at a centre-of-mass energy of just $\sqrt{s} = 7\,\text{TeV}$, Run 3 collisions are expected to take place consistently at $\sqrt{s} = 13.6\,\text{TeV}$, starting 2022. While the energy almost doubled over this period, the integrated luminosity increased 15-fold. Those massive upgrades required the LHC experiments to undergo improvements as well in order to keep up with the latest data-taking conditions. Figure 2.8 shows the evolution

of the LHC's centre-of-mass energy and luminosity during the last decade, alongside the shutdown periods when the upgrades took place.

The end-cap region of the Muon Spectrometer consists of the so-called *wheels*, which are visually represented in figure 2.10. Muon detection systems placed in those regions carry the responsibility of measuring high-η muons. Unfortunately, this region is particularly vulnerable to muon tracking performance issues generated by the increase in cavern background radiation associated with higher luminosity levels. Both the tracking resolution and efficiency were expected to significantly degrade for Run 3 operating conditions in the *Small Wheel* region of the Muon Spectrometer. On top of that, according to the NSW TDR [42], around 90% of the triggered muons in the end-cap region turned out to be fake when analyzing 2012 data. Those major issues prompted one of the most meaningful upgrades to the ATLAS muon system with the installation of the New Small Wheel (NSW), which took place during the Long Shutdown 2. By replacing the *Small Wheels* with this new detection ensemble, both problems could be addressed at the same time. It comprises both precision tracking and trigger detectors designed to successfully operate at the high data rates expected in Run 3.

2.3.1. Layout of the New Small Wheels

As the name suggests, the NSW is a *wheel-shaped* ensemble of detectors, covering the ATLAS Detector's inner end-cap regions on both sides. Figure 2.9 shows one side of the NSW as it is being installed during the second Long Shutdown of the ATLAS Detector. Measuring about 10 m in diameter, the NSW is a gigantic piece of hardware, and, in order to facilitate its construction and assembly, it is divided into 16 trapezoidal *sectors*. Every sector is designed to be able to operate independently of the others, having its own individual set of electronics for readout and monitoring. As a means of maximizing the coverage area and minimizing the overlap between sectors, there are two different sector layouts differentiated by their size: **Small** sectors and **Large** sectors. The sector types are deployed in alternation such as no two Small or Large sectors are placed next to each other. In the final assembly, any of the two NSW sides contains eight **Large** and eight **Small** sectors.

Regardless of the sector size, there are two different types of detector technologies employed across the New Small Wheel (NSW): MicroMegas (MM) and small-strip Thin-Gap Chambers (sTGC). Figure 2.12 contains a visual representation of the two different sector layouts as part of the full NSW's assembly, alongside the relative shape and size of the detector modules for the two different technology types. The highlights in Figure 2.12 show the area of the detector modules themselves as well as the size constraints for the envelope. Envelopes also contain the electronics for powering, readout and monitoring, necessary for detector operation. Proportionally accurate active area overlap diagrams are provided in Figure 2.11, with 2.11(a) displaying sTGC modules' overlap and 2.11(b) displaying MM modules' overlap.

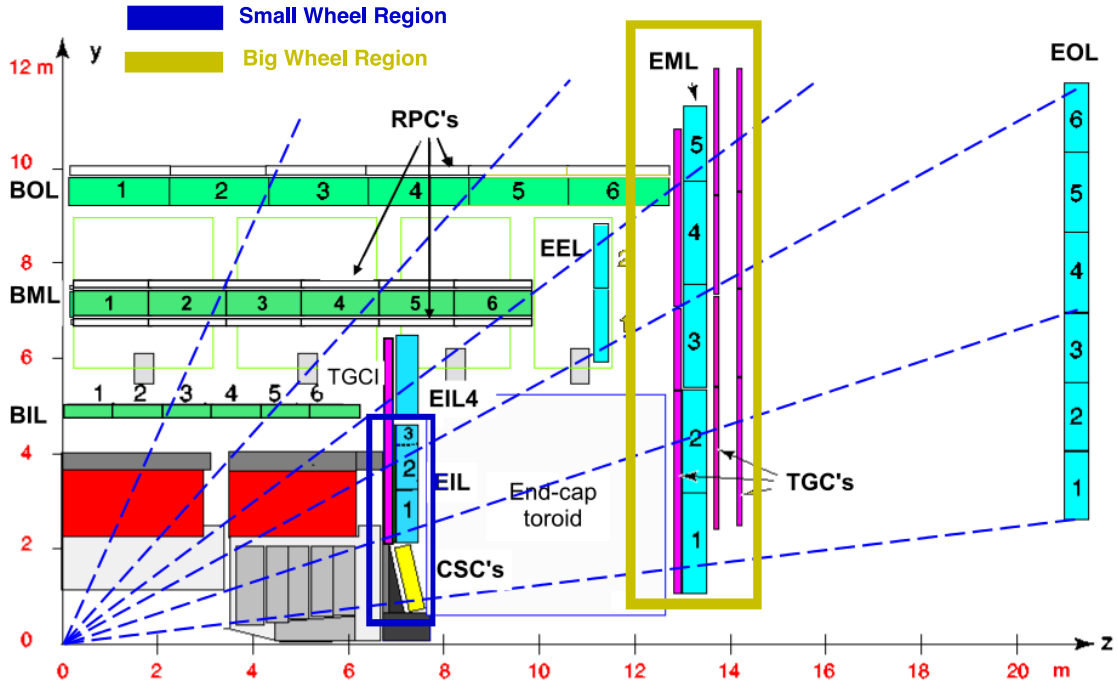
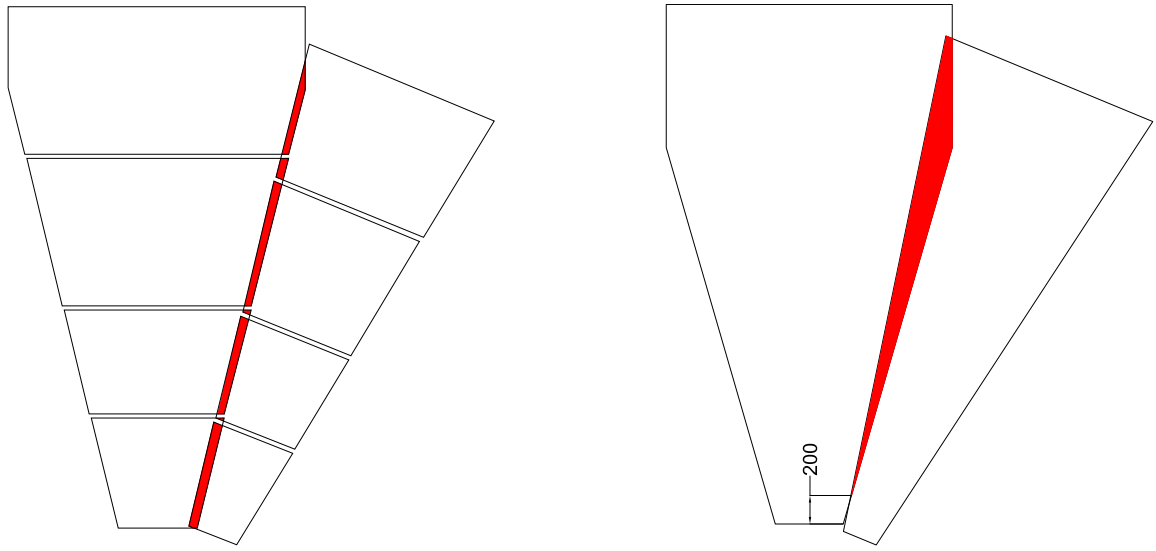


Figure 2.10.: Schematic representing a top view cross-section of one-quarter of the ATLAS detector [42]. The Small Wheel region is outlined in blue while the *Big Wheel* region is in yellow.

In terms of the sectors' structure, the NSW detector planes are grouped in two *multilayers*, each one having four sTGC and four MM detector planes. The arrangement of those layers places the sTGC planes at the beginning and the end of the layer sequence while the MM planes are located in the middle, maximizing the distance between the groups of sTGC layers. Thus, the NSW totals 16 detector planes, eight for each detector technology. Given the complexity of this system, a hierarchical naming scheme was created in order to simplify and clearly define the granularity of the detector. Appendix A dives deeper into the terminology used to describe the NSW's sector components.

With MicroMegas detectors' excellent precision tracking and small-strip Thin-Gap Chambers detectors' ability to trigger within a single bunch crossing, the NSW is expected to be able to provide efficiencies close to 100% for both online and offline track reconstruction during Run 3 and beyond. The choice of eight planes per detector type comes as further future-proofing of the NSW's design, offering additional robustness to accommodate even the expected data rates of the High-Luminosity LHC (HL-LHC). While the two detector types are highly specialized for their task, using them together, in a complementary way, maximizes the performance for both precision measurements and triggering.

The following two subsections are describing the main detector technologies and their



(a) sTGC modules (b) MM modules

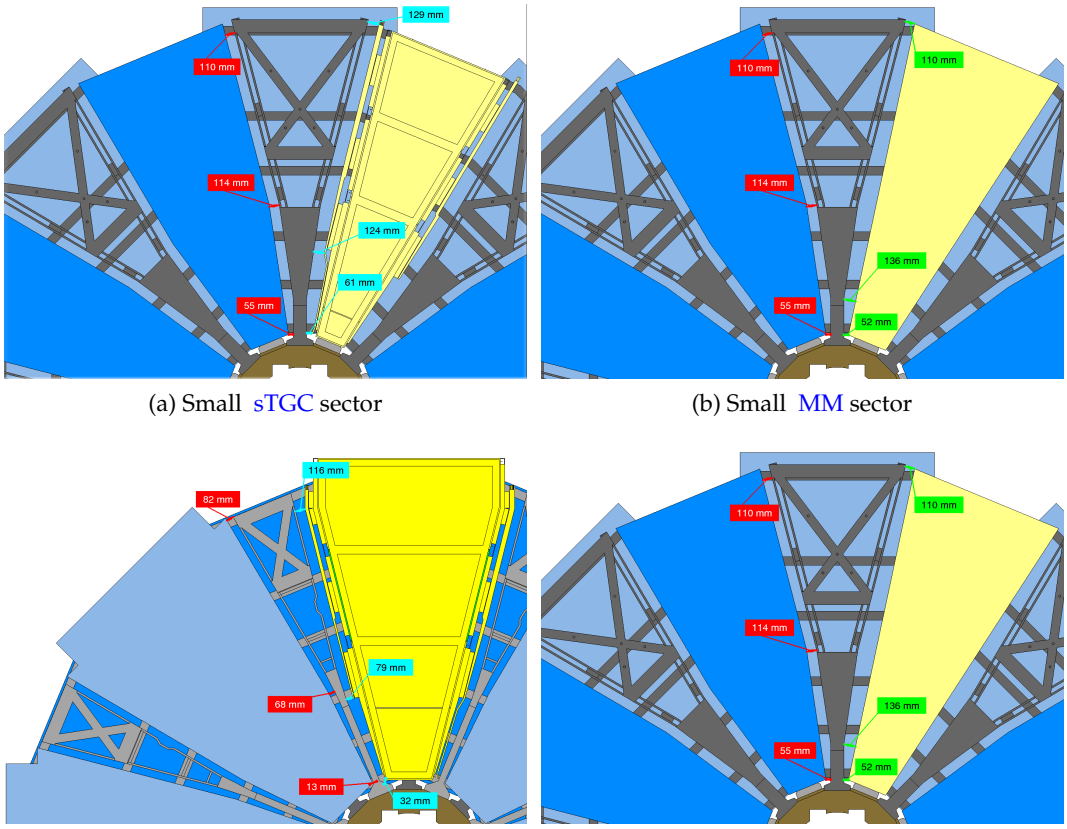
Figure 2.11.: Representation of active area overlap between small and large sectors [42].

implementation within the NSW. Since one of the primary research topics of the thesis revolves around MicroMegas detectors, a lot more details are provided in their corresponding section 2.3.2, but the sTGC overview in 2.3.3 is sure to tackle the most important aspects of its layout and functionality, as well.

2.3.2. MicroMegas Detectors

MicroMegas (MM) detectors first appeared in 1996 as designed and published by I. Giomataris et. al. in [54]. They follow broadly the operating principles of gaseous ionization detectors, featuring a small gas volume that incoming charged particles are ionizing. MM detectors are mainly made out of a planar electrode and a metallic mesh, with a gas gap in between measuring only a few millimetres and acting as a drift region. Primary electrons created through ionization in the gas gap start drifting towards the amplification region which is the space bounded by the readout electrodes and the mesh. Because of the amplification region's high-voltage potential, its electric field is $50 - 100$ times stronger than the one in the drift region. This ensures that more than 95% of electrons end up passing through the mesh. After that point an electron avalanche forms, generating a readout signal in fractions of a nanosecond. Figure 2.13(a) offers an illustration of the entire process.

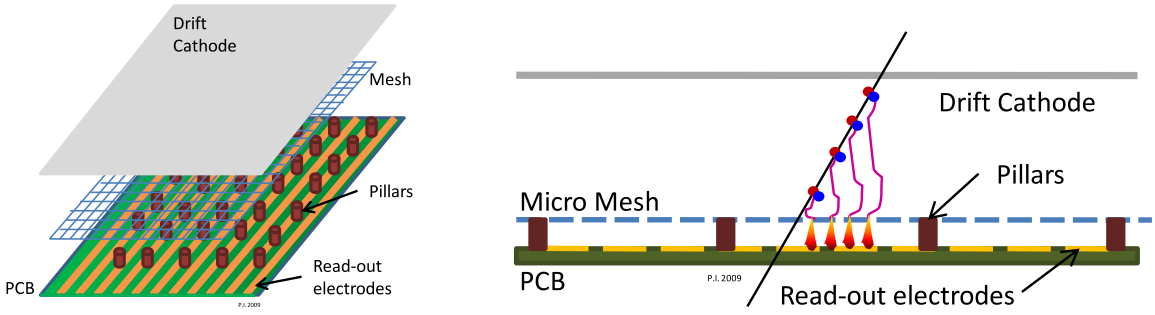
For use within ATLAS's NSW, the general mode of functioning described above still applies, but there have been made some slight alterations. One of the issues with the original MM detectors is their susceptibility to spark emissions caused by the avalanching



(a) Small sTGC sector (b) Small MM sector

(c) Large sTGC sector (d) Large MM sector

Figure 2.12.: Schematic layout of the NSW, highlighting sTGC and MM *modules* with yellow. Blue highlights cover the area allowed for the module's envelope and grey elements of this diagram show the mechanical support structures, as referenced in the TDR [42].



(a) Components of a MM *plane* (b) Operating principle of the MM detector

Figure 2.13.: Schematic depiction of the MM structure and working principles [42].

electrons. Those sparks would expose the detector to an increased risk of damage while also increasing the dead time between readouts. As a prevention measure, the NSW MM design includes between the copper readout strip and the avalanche region an insulating layer with resistive strips on top. This is meant to shield the readout strip from the charges created by the avalanche. For this to work, the readout channels are exposed indirectly to the electrical signal, through a capacitor. Figure 2.14 shows a representation of the readout board's components and assembly.

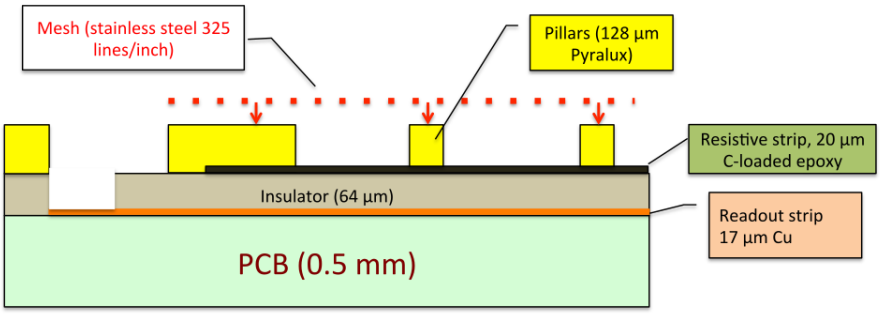


Figure 2.14.: Configuration of the MM readout boards [42].

The other major difference of the NSW MM is related to the applied voltages. As opposed to the original design where the mesh has negative high-voltage applied while the strips are connected to *ground*, the ATLAS design features high positive voltage on the resistive strips and an amplification mesh at *ground* potential. This compounds the other spark-mitigation measures by allowing the mesh to carry out releasing current produced by the spark. Table 2.2 contains the most important parameters of the MM detectors' operation and design. With those two MM design modifications the signal peak may be reduced alongside the spark current intensity, but the NSW is set up to operate for the foreseeable lifetime of the ATLAS Detector without requiring any more significant upgrades.

In terms of the layout of the MM planes, not all of them are identical. Every multiplet of four layers contains two so-called η planes and to stereo planes. Both types are similar in most regards, except for the strip inclination. The stereo layers have their strips angled at a $\pm 1.5^\circ$ relative to the η ones in order to enable the measurement of a second coordinate. Some finer details concerning the terminology used are included in the appendix A.1. The figure A.2 in the appendix shows the configuration of PCBs for small A.2(a) and large A.2(b) sectors, featuring an η layer on the left and a $+1.5^\circ$ stereo layer on the right. A NSW sector with its two MM wedges contains a total of eight planes: four of which are standard η layers, denoted by **X** for simplicity. The other four stereo layers are named according to the angle of the strips: **U** with $+1.5^\circ$ inclination and **V** having -1.5°. Thus, a muon passing through a NSW sector will encounter, in succession, the following layer types: **X X U V** within the first multiplet and **U V X X** within the second one. With such placement, the η-planes end up at the very start and end of the NSW sector, since sTGC chamber is placed between the MM

wedges. On the other hand, the stereo layer groups from the two MM modules are placed as close to each other as the geometry allows.

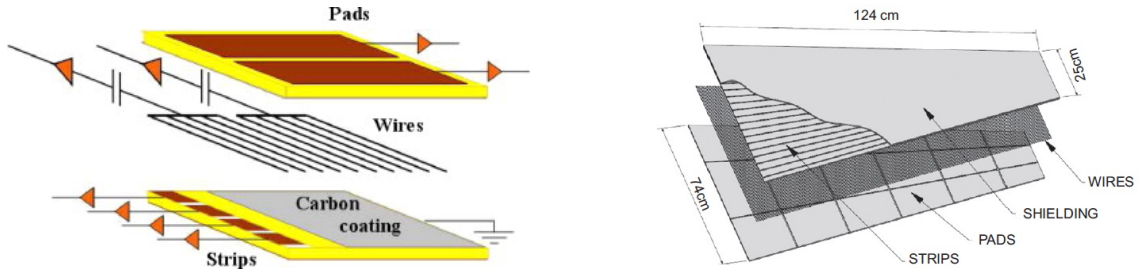
Item/Parameter	Characteristics	Value
Mesh	Stainless steel separate from readout board	325 lines/inch
Amplification gap		128 μm
Drift/conversion gap		5 mm
Resistive strips	Interconnected	R = 10-20 MΩ/cm
Readout strip pitch	4/8 layers	$\pm 1.5^{\circ}$
Total number of strips		2.1 M
Gas	Ar : CO_2	93 : 7
Voltage on resistive strips	positive polarity	550V
Amplification Field		40kV/cm
Drift Field		600V/cm

Table 2.2.: Design and operating parameters of the MM detector [42]

Apart from its ATLAS applications, this technology is being used today in many other particle physics applications, as elaborated by [55].

2.3.3. Small-strip Thin Gap Chamber Detectors

The small-strip Thin-Gap Chambers (sTGC) was designed to carry out the most important trigger functions of NSW and help compound the precision measurements of MM detectors. Similar technology has already been deployed in the ATLAS Detector's with the Thin-Gap Chambers (TGC) being a part of the Muon Spectrometer since its conception. Compared to the TGC, whose uses and localization are restricted to the *Barrel Region* and are touched upon in subsection 2.2.4, the sTGC was engineered to have a much better spatial resolution (100 μm compared to 10 mm).



(a) Electronics of the sTGC (b) Layout of the sTGC

Figure 2.15.: Representations of sTGC's main components [42].

At its core, the sTGC readouts signals from three types of elements: pads, wires and strips. Their placement is presented in Figure 2.15. As Figure 2.15(a) shows, the wires and strips are perpendicular to each other and separated by a carbon-coated plane. On top of everything, there are the pads with a large surface area coverage, whose signals are used to identify the region of interest where wire and strip readout should be performed. An entire plane of sTGC detectors, part of a small sector, can be seen in Figure 2.15(b).

3. Object Reconstruction in ATLAS

The previous chapter introduced the many parts of the ATLAS Detector diving deeper into their structure and purpose, but all of those elements have to work together in order to reproduce in as much detail as possible the specifics of every *triggered* event. Detector signals can not be used directly so they have to be refined, often using complex methods to derive higher level information usable within ATLAS analysis. A suggestive diagram representing how certain particles interact with the detector is shown in 3.1. While there are many physics objects of interest, this chapter will dive deeper into the ones related to the main topics of the thesis, namely: electrons, muons and jets. With previous sections covering the data-taking infrastructure and the data flow up to permanent storage, the final section will introduce the ATLAS data model and data analysis methodologies, providing additional context for the data processing model and infrastructure employed by ATLAS. The content of this chapter is based on the following major bibliographical sources: *ATLAS detector and physics performance: Technical Design Report, 1* [47] and 'The ATLAS Experiment at the CERN Large Hadron Collider' [41].

3.1. Low-level Objects

There are several layers of abstraction in physics object reconstruction. Starting from the lowest level illustrated by raw electrical signals coming from the ATLAS Detector, the data becomes more and more refined as more processing is applied to it. Before tackling the actual physics objects it is important to understand some intermediate representations that are lower on the abstraction hierarchy.

3.1.1. Tracks

Tracking, in general, refers to position measurements and, within ATLAS, the Inner Detector plays the biggest role in this task. Subsection 2.2.2 has already briefly covered the Inner Detector's structure. Track finding relies on signals from the *Pixel* and SCT detectors and starts by clustering adjacent signals. The locations of those clusters are referred to as *space-points* since they represent a location in 3D space [57]. Those go on to become seeds for iterative combinatorics, attempting to form a track from any possible set of 3 points. After scoring the quality of the tracks based on criteria such as the number of points and

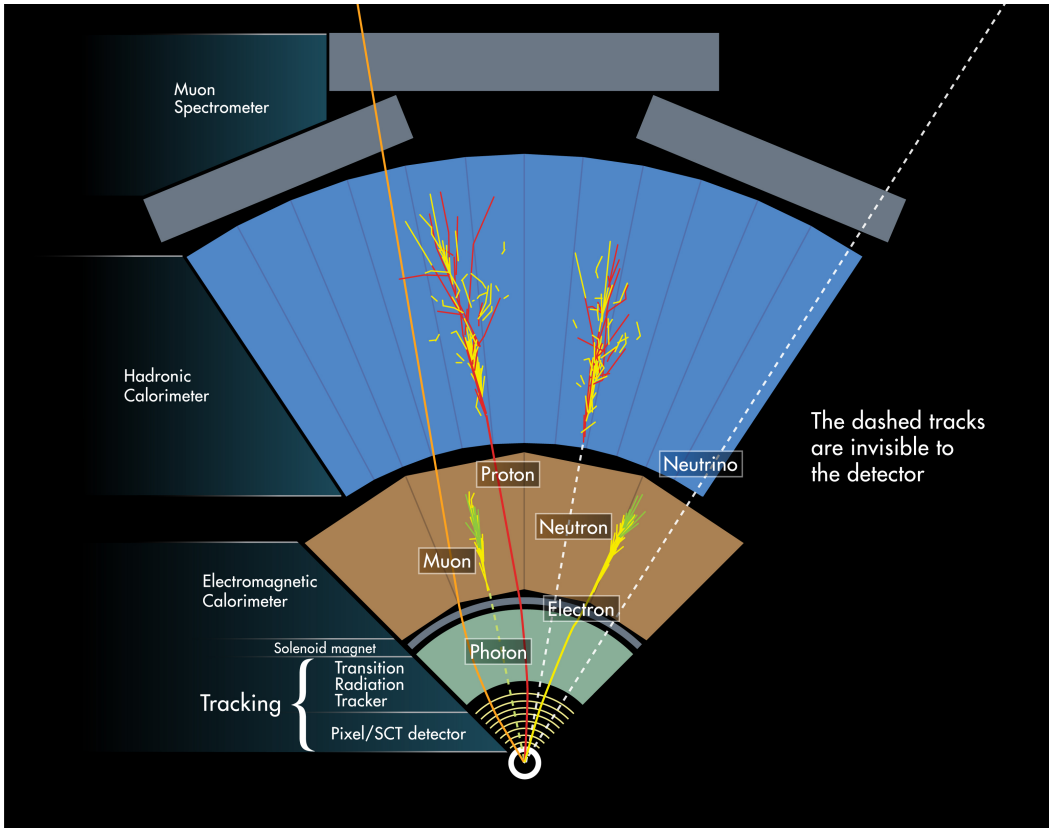


Figure 3.1.: Signatures of several physics objects of importance in ATLAS [56].

impact parameter, track candidates are ranked based on the score in descending order. The combination method is purposely designed to allow the same cluster to be part of several track candidates. To solve the inevitable ambiguities raised by this approach, a Neural Network (NN) is used to predict cluster position, attempting to extrapolate and fit all track candidates. Figure 3.2 shows a flowchart of track candidates through the ambiguity-solving algorithm.

3.1.2. Vertices

Having found the tracks, vertices can be reconstructed as the intersection points between them. The main point of interaction, namely the **primary vertex** is the most important one to identify correctly. Unfortunately, high pile-up conditions make this process increasingly difficult, due to the existence of many hard scattering vertices per event [58]. Thus, primary vertex finding is also an iterative procedure, involving many recalculations through *iterative annealing* fitting [59, 60]. Once the primary vertex is found, the procedure continues, looking for new vertices using the remaining tracks until there are none left.

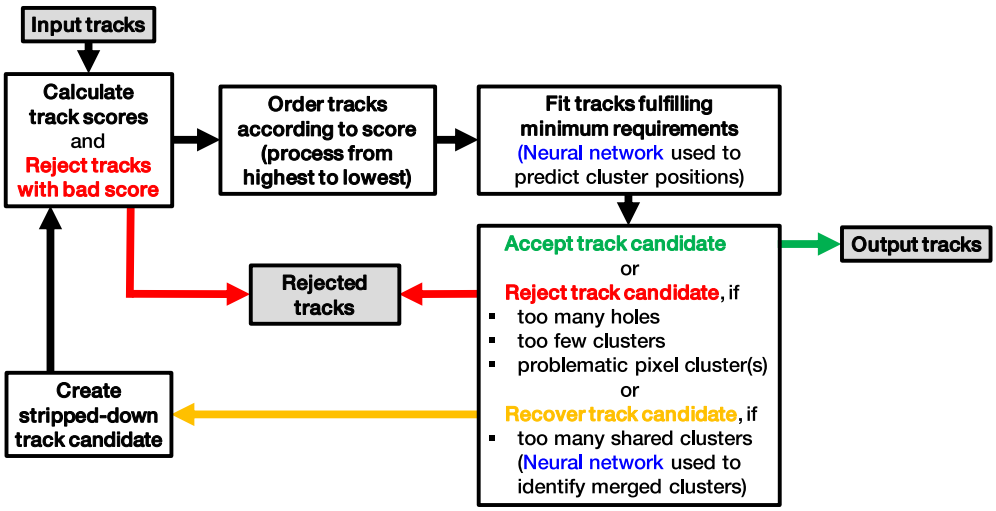


Figure 3.2.: Schematic representation of the track finding process [57].

3.2. Electrons

Electron reconstruction primarily uses information from the EM calorimeters alongside tracks recovered by the inner detector. [61]. There are two major challenges related to the treatment of those particles. First of all, the **reconstruction** process has to be resilient to fakes, since other types of physics object could also produce similar tracks and energy deposits inside the calorimeter. Secondly, even after making sure that what has been reconstructed is indeed an electron, it is still necessary to differentiate between prompt and non-prompt electrons through a process referred to as **identification (ID)**.

Figure 3.3 shows the ATLAS Detector subsystems that should generate readout signals from interacting with an electron. The path of the electron is highlighted with a solid red line but while interacting with the detector's materials there is often *Bremsstrahlung* radiation being emitted. This poses some additional difficulties because the radiated photons carry enough energy to be able to convert into a e^+e^- pair, which will, in turn, bring forth another set of signals in the detector.

3.2.1. Reconstruction

Due to the additional particles created through an electron's interaction with the different detector's materials, what one could expect to be a straightforward trace (albeit bent by the magnetic fields), becomes a cluster of track signals and energy deposits.

The reconstruction procedure begins with a fit of the signals from the inner detector and an attempt to correlate the resulting track candidate with an energy cluster from the LAr EM

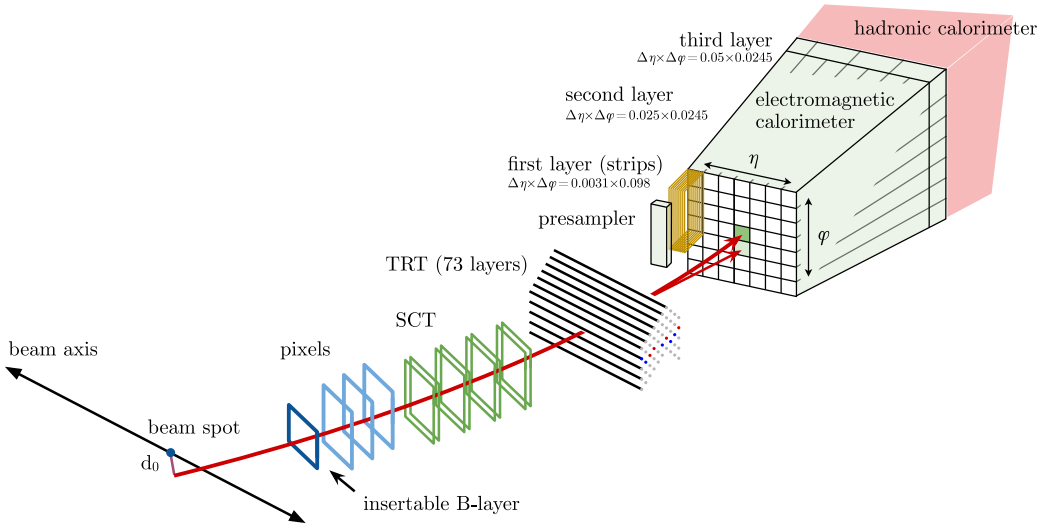


Figure 3.3.: Schematic of ATLAS Detector elements that an electron would interact with, as of the beginning of Run 2 [61].

calorimeter. Energy clusters are sought after first, using sliding windows to successively scan regions of 3×5 *towers*. Each tower covers an area of $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$.

Track candidates that could be associated with a corresponding energy cluster undergo a more rigorous fitting procedure using the ATLAS Global χ^2 Fitter algorithm [62]. Finally, an addition fit is performed based on a multitude of Kalman filters [63] that are meant to increase the reconstruction performance by taking into account the losses in energy due to the interaction with the detector. This algorithm is known as the Gaussian-sum filter (GSF) method [64] and it is the last step in the process of reconstructing an electron candidate.

3.2.2. Identification

Once an electron candidate has been reconstructed, it is necessary to establish whether or not it's a prompt electron or not. All of the detector information previously used in the reconstruction step is once again processed and re-analyzed in order to assess the likelihood of the candidate being part of the signal or the background distribution. For this purpose, a statistical model (likelihood) is the method of choice.

$$L_S(\mathbf{x}) = \prod_{i=1}^{n} P_{S,i}(x_i) \tag{3.1}$$

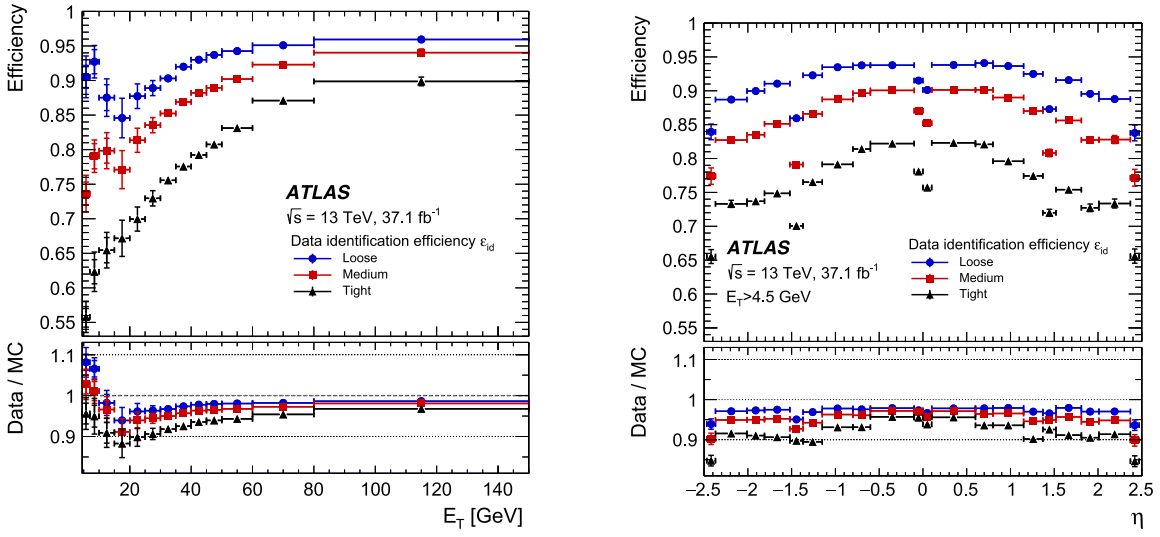
In Equation 3.1 $L_S(\mathbf{x})$ is the likelihood for the electron to be prompt and $P_{S,i}(x_i)$ is the value of the signal Probability Density Function (PDF) for the variable labeled i ($P_{S,i}$) at value x_i. A similar likelihood model $L_B(\mathbf{x})$ is constructed for determining the probability of the

electron to be non-prompt. The n variables in question making up the likelihood model are displayed in Table 3.1; those are the components of the **x** vector. Any statistical correlations between the quantities involved in the likelihood model are neglected.

The two likelihood probabilities L_S and L_B are combined into the main variable used for prompt electron ID: the **likelihood discriminant** d_L, as shown in Equation 3.2.

$$d_L = \frac{L_S}{L_S + L_B} \tag{3.2}$$

Having this discriminant allows any analysis team to develop suitable event selection models for their work by choosing a proper operating point. ATLAS has a dedicated Combined Performance () group that designs and test those ID tools. According to their publication for Run 2 data [61], the efficiencies of the d_L discriminant's main operating points show the values displayed in Figure 3.4. The choice of the operating point involves a



(a) (b)

Figure 3.4.: Electron- ID efficiencies in $Z \to ee$ events for the three main operating points: *Loose* (blue circle), *Medium* (red square), and *Tight* (black triangle) as a function of E_T 3.4(a) and η 3.4(b). The bottom side of both plots displays the event ratios for data vs. simulation [61]

tradeoff between a higher sample purity achievable through a *Tight* selection and a higher sample size resulting from a *Loose* selection criteria.

3.3. Muons

Prompt leptons are generally important physics objects to reconstruct from collision events. If the previous section tackled the reconstruction of the electron, for this thesis's

Type	Description	Notation
Hadronic leakage	Ratio between E_T in the first layer of the hadronic calorimeter and E_T of the EM cluster	R_{had1}
	Ratio between E_T in the whole hadronic calorimeter and E_T of the EM cluster	R_{had}
EM calorimeter's third layer	Ratio between the energy in the third EM layer and the total EM energy	f_3
EM calorimeter's second layer	Lateral shower width	$w_{\eta 2}$
	Ratio of the sum of the energies of the cells contained in a $3 \times 3 \times (\eta \times \phi)$ rectangle to the sum of the cell energies in a 3×7 rectangle, both centred around the most energetic cell	R_ϕ
	Ratio of the sum of the energies of the cells contained in a $3 \times 7 \times (\eta \times \phi)$ rectangle to the sum of the cell energies in a 7×7 rectangle, both centred around the most energetic cell	R_η
EM calorimeter's first layer	Ratio of the energy difference between the maximum energy deposit and the energy deposit in a secondary maximum in the cluster to the sum of these energies	E_{ratio}
	Ratio of the energy measured in the first layer of the EM calorimeter to the total energy of the EM cluster	f_1
Track conditions	Transverse impact parameter relative to the beam-line	d_0
	Significance of transverse impact parameter	$d_0$sig
	Momentum loss between the perigee and the last measurement point divided by the momentum at perigee	$\Delta p/p$
Transition Radiation Tracker (TRT)	Likelihood probability based on transition radiation measured by the TRT	eProbabilityHT
	$\Delta\eta$ between the EM cluster location in the first layer and the extrapolated track	$\Delta\eta_1$
	$\Delta\phi$ between the EM cluster location in the second layer and the momentum-rescaled track; it carries the same sign as the particle's charge	$\Delta\phi_{res}$

Table 3.1.: Variables used in computing the likelihood discriminant for electron ID [61]

purposes, it is important to discuss the muon as well. As opposed to electrons, muons are significantly less interactive with the matter they encounter. In the muon's case, neither the LAr EM calorimeter nor the hadronic calorimeter is enough to completely stop it. This fact, compounded by the existence of a dedicated muon-detecting system in the form of the Muon Spectrometer, makes reconstructing muons somewhat less of a challenge. At the same time, the correct identification (ID) of prompts is still necessary for the reconstructed candidates.

3.3.1. Reconstruction

Tracks in the Muon Spectrometer are the first step of the reconstruction process. They come in the form of straight lines of detector hits and get matched between the different stations of the Muon Spectrometer. The atlas Global χ^2 Fitter algorithm [62] is once again deployed to backtrace the trajectory of the muon through the entire ATLAS Detector. This tool is important because, even though the tracks from individual stations are usually straight, when combining hits from multiple stations it is necessary to account for the magnetic fields in between. After the ill-fitting hits are removed and the interaction point is added to the mix, the quality of a second fit determines if a muon track is consistent or not [65]. This final fit uses calorimeter information to estimate the muon's energy loss and computes a first estimation of its p_{T}.

With a valid muon track candidate originating from the interaction point, the next step involves adding signals from the other ATLAS detectors. Inner Detector tracks give the most relevant information in this case. However, different detector stations have different coverages in terms of η and there could also be tracks with missing traces or hits which means that there are several particular ways to reconstruct the muon:

- **combined muon**: Those represent the best-case scenario where the tracks from the Muon Spectrometer can be matched with tracks from the Inner Detector. For tracks outside the $|\eta| \leq 2.5$ coverage of the *Pixel Detector* hits from the Transition Radiation Tracker (TRT) are used instead. Those cases are considered a subcategory of combined muons and are referred to as sillicon-associdated forward muons (SiFs).

- **inside-out combined muon**: As the name suggests, if the algorithm finds clear tracks within the Inner Detector that may be matched just with a few hits from the Muon Spectrometer, the resulting physics object would be an *inside-out* muon. This scenario mostly applies to low-p_{T} muons that may not generate complete tracks in the Muon Spectrometer.

- **muon-spectrometer extrapolated muon**: As opposed to the previous case, in this instance, the Inner Detector tracks are missing altogether, but there is a clear set of correlated tracks from the Muon Spectrometer stations. If this set of tracks' trajectory

can be extrapolated to the interaction point (IP), a *muon-spectrometer extrapolated* muon will be reconstructed based on them.

- **segment-tagged muon**: The extrapolation could also happen the other way around, starting from Inner Detector tracks and matching at least a segment from the Muon Spectrometer. Muons reconstructed through this procedure are suggestively named *segment-tagged*.

- **calorimeter-tagged muon**: Using calorimeter segments is a viable option considering how much less ionization would a muon generate, compared to an electron. Matching those energy deposits in the calorimeter with Inner Detector tracks, but without any Muon Spectrometer hits results in a muon candidate that is *calorimeter-tagged*.

3.3.2. Identification

Once again, a reconstructed muon may or may not be a prompt muon and there is an entire process of determining its origin, namely, muon identification (ID). Unlike electrons, where an entire statistical model with measured PDFs is necessary for correct ID, muons' identification criteria consist of a more straightforward set of rules. The hurdles here are coming from all of the distinct muon candidate types introduced in the previous section. Each of those types has its own set of requirements and on top of that, the requirements are dependent on p_{T} in most cases.

From the ATLAS analysis team member's perspective, muon- ID and electron- ID tools come similarly packaged as software. They also come with pre-defined working points having similar names such as *Loose*, *Medium* or *Tight*, all carefully selected to correspond to measured efficiencies. A more comprehensive summary of those rules corresponding to certain muon- ID working points could be consulted in [65] but as a general rule, the number of hits and segments is one of the main factors determining muon candidate quality.

Another thing worth noting is that in terms of background versus signal discrimination for muons, as well as for electrons, **isolation** is a key property of prompt leptons. This refers to a quantitative way of measuring detector signals in the vicinity either of tracks or clusters. Lower such activity around the physics object of interest is useful in excluding hadronic decays as the most likely source of the lepton. The most commonly used isolation variable is the ratio between transverse momentum and the sum of pt reconstructed in the calorimeter from all physics objects in a given ΔR radius.

3.4. Jets

Hadronisation happens in all processes where free quarks or gluons are created. A particle detector has no way of directly measuring any strongly interacting particle because they are always confined together. Thus, whenever quarks or gluons undergo Quantum Chromodynamics (QCD) interactions, what the detector *sees* is a cluster of secondary hadrons each further following its separate decay chain. Those clumps of particles created in the process of hadronisation are known as jets.

Given the nature of a jet and knowing the layout of the ATLAS Detector makes it clear that the main systems involved in jet reconstruction are the two calorimeters, especially the hadronic calorimeter. The shower created through hadronisation contains scores of different particles that are very close together which means they can't reliably be individually reconstructed. The jet physics object is thus a composite one featuring the shape of a cone rather than a track.

3.4.1. Particle Flow Algorithm

When it comes to reconstructing the hadronic showers as jets, the particle flow algorithm takes advantage of the full granularity of both calorimeters. The first step involves the so-called topological clusters [66]. Calorimeter *cells* with energy deposits several times higher than the noise floor are considered the seeds of the topological clusters. Adjacent cells signalling significant energy amounts are then progressively added to the seed clusters. Finally, a check is performed on the topological clusters in order to divide them in two if they show multiple local maxima in bins of energy.

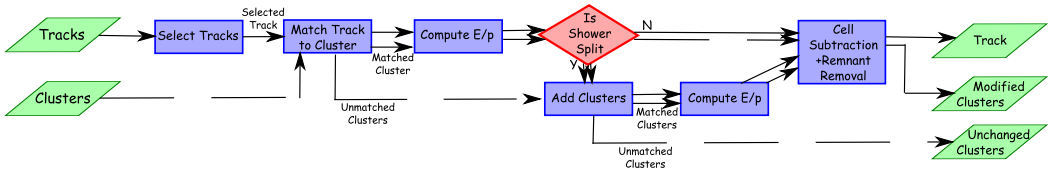


Figure 3.5.: Flow chart diagram showcasing the steps employed by the particle flow algorithm [67].

Topological clusters in the calorimeters are combined with tracks recorded by the Inner Detector before performing an overlap removal procedure. This entire process of subtraction and recombination is conventionally referred to as the *particle flow* algorithm [67]. A schematic representation of the algorithm's logic is shown in Figure 3.5. The tracks are left unchanged in the end, while the algorithm's output features a set of modified clusters alongside the original ones.

Tracks and topological clusters fall under the umbrella of *particle flow objects* and represent the building blocks of jets. Particle flow information is used as the input to the final step in jet reconstruction: jet clustering.

3.4.2. Jet Clustering

Jets are not *fundamental* objects, which implies that there is no universal method of jet finding. The available literature [68] features many different jet clustering algorithms and while some of them may be outdated, one could not find a single best approach for all common applications.

There are two major approaches to jet finding: *cone* algorithms and *sequential* algorithms. The former is revolving around the assumption that jets have inherently inflexible boundaries that always take the shape of a cone, which simplifies computation. The latter allows for more flexibility regarding the jet boundary, placing as much emphasis on momentum as on the location in the $\eta\phi$-plane, thus providing IRC safety, but at a higher computational cost. With the significant increases in processing power seen across the past decade, *cone* algorithms fell out of favour with the HEPP community while their better and more processing-heavy counterparts took their place.

Even when it comes to sequential algorithms, there are still several to choose from, the most used being: k_t, Cambridge–Aachen and anti-k_t. They all work iteratively computing a distance measure d_{ij} between every two particles and identifying the pair with the minimum. If the distance between those two objects is smaller than a given threshold, they are merged into a composite object called *pseudojet* [69]. This process is repeated until there is nothing left to merge and, at that point, all of the resulting *pseudojets* are promoted to the rank of jets. For pp collisions, there is an additional distance variable supposed to evaluate the distance between a *pseudojet* and the beamline.: d_{iB} When this distance is the smallest, the *pseudojet* is recombined with the *beam-jet* instead. The expressions of the two distances could be generalized between the three aforementioned sequential algorithms as follows.

$$d_{ij} = \min(p_{Ti}^{2p}, p_{Tj}^{2p}) \frac{\Delta R_{ij}^2}{R^2} \tag{3.3}$$

$$d_{iB} = p_{Ti}^{2p} \tag{3.4}$$

In Equations 3.3 and 3.4, the parameter p can take the values 1, 0 and -1 depending on the algorithm used: k_t, Cambridge–Aachen and anti-k_t respectively. R is another parameter related to the desired radius of the cone for the jets and ΔR_{ij} is described by Equation 3.5.

$$\Delta R_{ij}^2 = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2 \tag{3.5}$$

The use of d_{iB} determines the variant of the jet finding method which can be either *inclusive* or *exclusive*. Inclusive variants use the cone radius R and as such if the distance between

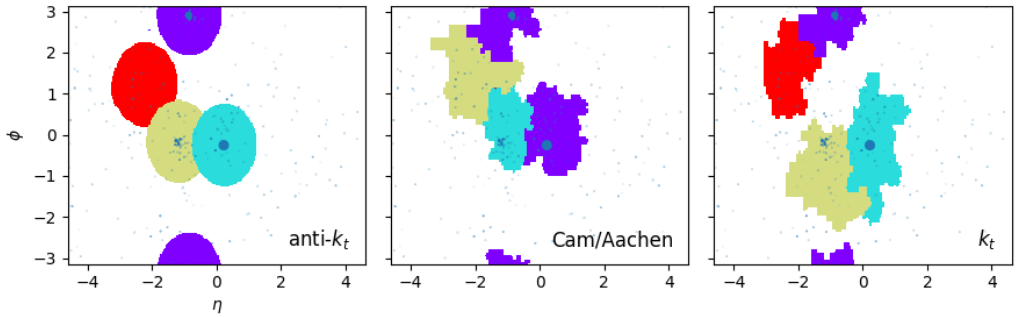


Figure 3.6.: Jet areas of the same event clustered with the three most popular sequential methods. The areas are highlighted with solid colour while the constituent particles are represented as blue dots; the size of the dot is proportional to particles' p_T [70, 71].

two *pseudojets* is greater than R, d_{iB} will always be greater than d_{ij}. In this scenario, instead of merging the i constituent with a beam-jet, it is labelled as a jet itself. Since any stray *pseudojet* is included in the final output it is often useful to apply p_T cuts to select only what could meaningfully be considered as a jet. For *exclusive* variants, either a jet-resolution d_{cut} or a number of jets n_{jets} parameter needs to be specified. Those two parameters signal the stop of the recombination sequence: either when both d_{iB} and d_{ij} are below d_{cut} or when exactly n_{jets} have been found. Every constituent particle is either part of a final state jet or the beam jet which is excluded from the output. A visual representation of the clustering results while using the inclusive variant of each of the three methods on the same event is shown in Figure 3.6.

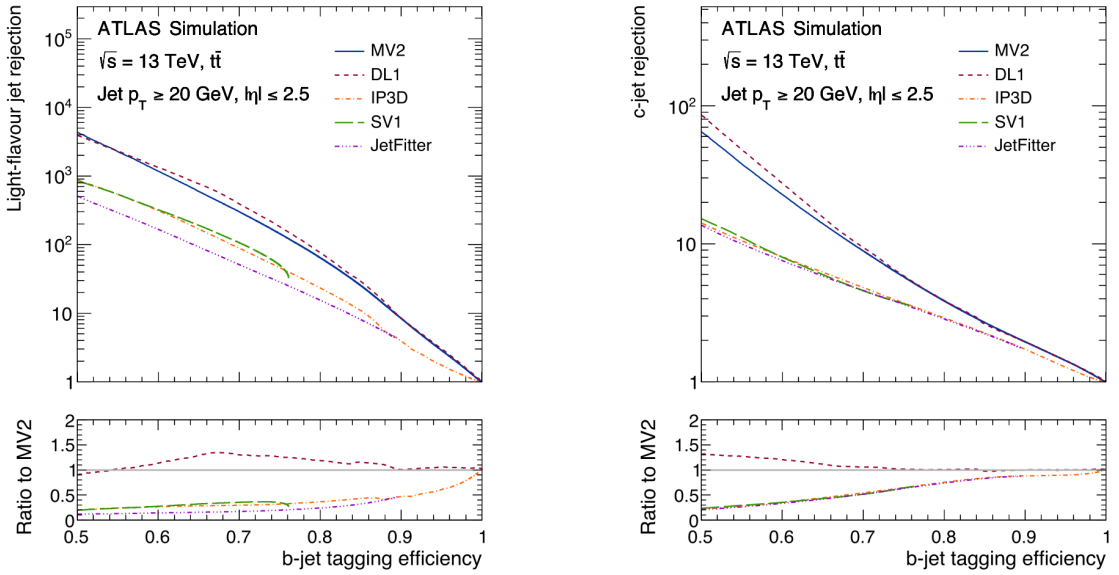
3.4.3. Flavour tagging

Jets are the primary means of probing QCD interactions and may vary vastly in structure, topology and kinematics. Unsurprisingly, all of those jet features are inherently related to the interaction or decay that produced it. The process of identifying the physics process behind a certain jet is known as **tagging**.

Tagging is meant to put a label on the jet based on the ancestral particle's properties. Depending on the mass of a decaying quark, the jet may be tagged as *heavy* or *light*. Tagging could also occur based on *flavour* considering how important correctly pinning down this information is to the study of many processes.

Considering the multitude of possible tags, correctly tagging a jet becomes a multi-class classification problem. During most of ATLAS' operation, there was an entirely different method/tool for each tagging label. Nowadays, Neural Network (NN) powered models, such as DL1 [72], are quickly improving, in a race to provide a unified tagging method for multiple classes.

By far the most common jet tagging task is *b*-tagging. The signature of the *b*-jet is meaningfully distinct from the lighter counterparts and the *b* quark is frequently sought after by many analyses, including the one in this thesis' Chapter 4. Being so ubiquitous, *b*-tagging, has a long history within ATLAS with many tools devoted to it [73, 74]. One of the first approaches focused on tracks' impact parameter information, creating a likelihood model from MC-derived PDFs. The two algorithms using this method were IP2D and IP3D [72]. Another option involved a search for a secondary vertex within the jet and was packaged as the SV1 algorithm [75]. More advanced vertex-finding procedures have also been employed. The JETFITTER reconstruction tool has the fitting capabilities to model the full *b* quark decay chain [76].

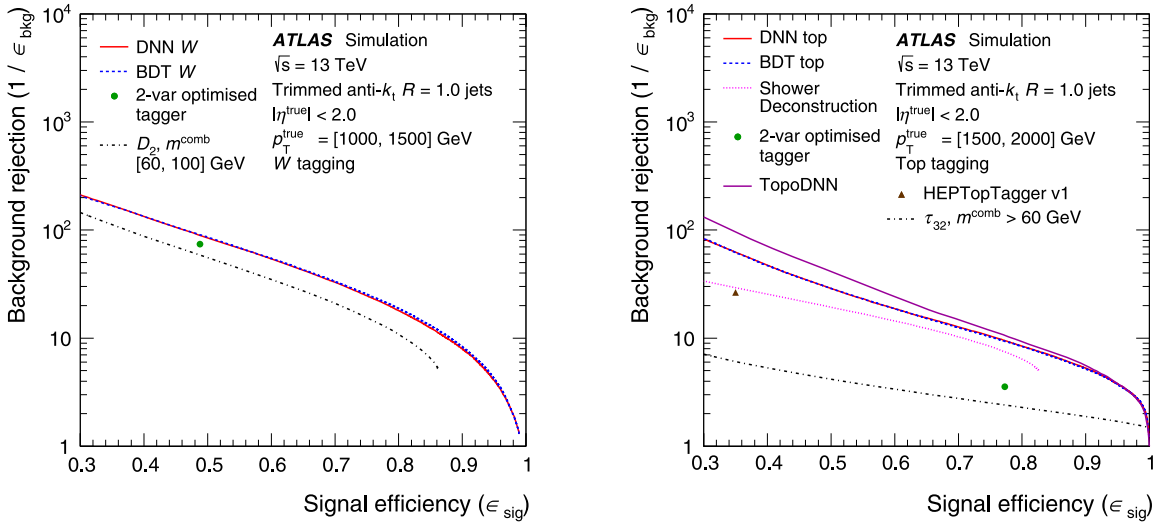


(a) (b)

Figure 3.7.: Background rejection versus the *b*-tagging efficiency for the several methods in $t\bar{t}$ events [74]. In 3.7(a) the background is comprised by light-flavour jets while 3.7(b) shows rejection power for a *c*-jet background.

While having multiple ways of *b*-tagging allows for greater flexibility, in order to make sure no performance is left on the table, they could also be used together as part of an **ensemble method**. The ATLAS Collaboration did that through the use of ROOT Toolkit for Multivariate Data Analysis (TMVA) [77] to train a Boosted Decision Tree (BDT). This algorithm is called internally MV2 [72] and was used extensively during Run 1 and Run 2. There were several releases of this tool with names like `MV2c00`, `MV2c10` and `MV2c20` based on the fraction of *c*-jet events used in the background training sample. Background rejection plots for Run 2 data are displayed in Figure 3.7 where the rejection-efficiency curves are computed for two different backgrounds.

Jets are also often found in the final states of events containing massive particles such as the W boson or the *top* quark, as part of their hadronic decay chains. Those objects are of particular interest since they may interfere, as a background source, with searches for potential BSM boosted massive particles. For those reasons, t and W tagging is an active area of research within the *Jet and* E_T^{miss} CP group. Several methods are being used for solving this type of tagging, ranging from multivariate analysis (similar to b-tagging) to Machine Learning (ML) models such as BDT and Deep Neural Network (DNN). A summary of comparative performances is shown in Figure 3.8, as detailed in Reference [78].



(a) (b)

Figure 3.8.: Performance measured as the background rejection versus the signal efficiency for several popular methods of W-tagging 3.8(a) and t-tagging 3.8(b), at the higher end of the p_T spectrum [78].

For W-tagging, the BDT and DNN methods have similar performance, decisively above cut-based approach involving the best discriminating variables: combined mass and D_2 energy correlation variable. The green dot in Figure 3.8(a) shows the cut-based performance optimized for a 50% efficiency working point.

Performance measurements differ when evaluating t-taggers, according to Figure 3.8(b). In this case, the best two discriminating variables are the ratio between the 3-*subjettines* and 2-*subjettines* alongside the combined mass, but their discrimination power in the cut-based approach is significantly worse than the BDT and DNN approaches. Additional methods were tested for this task, such as *Shower Deconstruction* [79] which uses likelihood ratios, HEPTopTagger [80] and a DNN models trained directly on the jet constituents [81], referred as *TopoDNN* in Figure 3.8(b).

All methods presented so far for tagging W and t jets make use of the **substructure** of the jet. Inputs for both BDT and DNN models contain some high-level jet substructure variables,

with the exception of *TopoDNN*, which still uses jet substructure, but accesses it directly by using the jet constituents.

3.4.4. Jet Substructure

Jets are high-level physics objects carrying the same kinematic properties as particles. However, since they are made up of multiple constituent particles, there is a lot of additional information that can be extracted based on the constituents' characteristics. Properties based on this kind of lower-level information are commonly referred to as jet substructure variables.

The constituents of a jet, may themselves be distributed between smaller-radius clusters, also known as *prongs*. This type of property is reflected by quantities such as the N-subjettiness τ_N [82] which can be interpreted as the extent to which a jet can be characterized as having N subjets. In order to calculate τ_N, one starts by clustering the jet constituents into N smaller subjets. For a jet with n_c constituent particles, N-subjettiness is defined by Equation 3.6.

$$\tau_N = \frac{1}{d_0} \sum_{k}^{n_c} \min\{\Delta R_{i,k} \; ; \; i = \overline{1, N}\} \tag{3.6}$$

ΔR is defined in Equation 3.5 and the d_0 factor is introduced for the purpose of normalization and calculated as the sum of the transverse momentum of the constituents multiplied by the radius of jet whose subjettiness is being evaluated. Using the ratio $\tau_{i,i-1} := \tau_i / \tau_{i-1}$ one could effectively test jets for how likely they are to have exactly i prongs. The usefulness of N-subjettiness is further compounded by the fact that, provided good subjet candidate selection, it is Infrared and Collinear (IRC) safe.

Literature on jet substructure defines many such useful variables such as **eccentricity** [83], **planar flow** [84], **angularity** [84] and many others. There are as well more modern approaches attempting to provide complete descriptions of the substructure, such as Energy Flow Polynomials [85].

Another way of representing substructure takes the form of **jet images** [86]. Jet reconstruction depends on calorimeter data which, as was previously discussed, is recorded in terms of energy deposits measured by the calorimeter's cells. This binning of energy deposits can be projected as a flat image in $\phi\eta$-space. Starting from just a list of constituents and summing their energy values over $0.1\phi \times 0.1\eta$ bins, the resulting 2D histogram becomes a **jet image**.

Figure 3.9 shows an example of how a two-pronged jet may be represented as an image. Before binning, the constituents were translated and rotated in $\phi\eta$-space in such a way that jet's primary subcluster is located at the origin and the second largest subcluster in terms of total energy is underneath it on the vertical axis. Transformations such as those make it

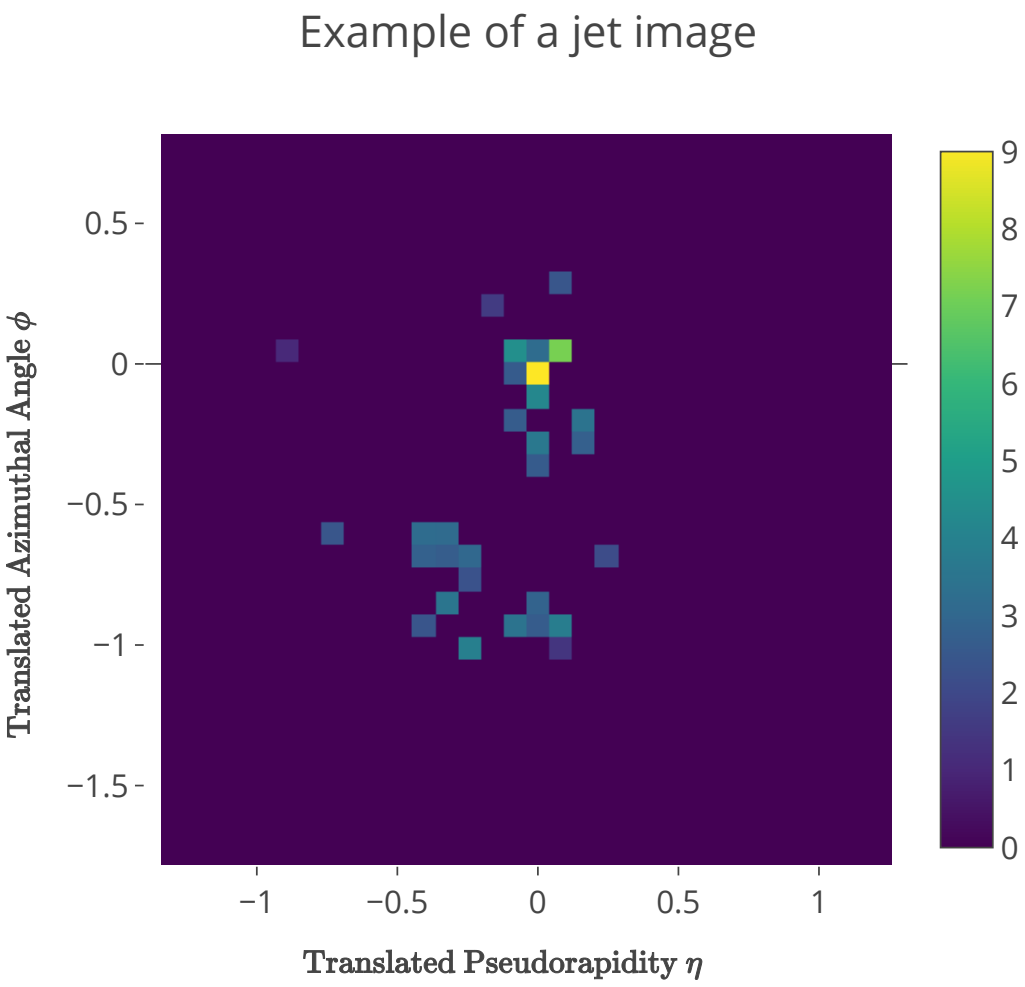


Figure 3.9.: Image of a two-pronged jet

easier for Machine Learning models such as Convolutional Neural Network (CNN) to use **jet images** as input.

3.5. The ATLAS Data Model

Considering the scale at which ATLAS operates, the data processing model has to keep up with the ever-increasing amount of data while streamlining the processing as much as possible through automation. The ATLAS Collaboration had to develop its own computing infrastructure for both hardware and software in order to solve the uniquely challenging problems that arise due to the scope of the experiment.

First of all, data is organized based on the conditions whereupon it was recorded. Successive events with (near) constant LHC operating parameters are part of a Luminosity Block, in groupings of around 10^5. All of the collision events part of the same LHC fill (beam) comprise a *Run*. Higher on the ladder there are *sub-periods* containing multiple *Runs* and being packed together into *Periods*. A LHC run, not to be confused with the previously mentioned *Run*, contains all events within all periods between two Long Shutdowns. In this data organization scheme, every event can be uniquely identified by the *Run* number and the event number.

Not all of the data encapsulated within an event is always necessary and not all events part of the same organizational unit are useful. The data volume available is so large that it is prohibitively difficult to make it available to all members of the ATLAS Collaboration across the world. In order to optimize both availability and storage space, the ATLAS Event Data Model (EDM) [87] implements a data hierarchy comprising several stages of processing. However, this introduces significant computational requirements for data processing and distribution. All of those computing services are provided by the Worldwide LHC Computing Grid (WLCG) [88], a global network of computing centres managing the hundreds of petabytes of data collected by CERN's experiments every year. The grid encompasses a multi-tier infrastructure using both on-site and remote computing resources. Raw data coming from the ATLAS Trigger and Data Acquisition (TDAQ) is sent to the **Tier-0** facility where it gets recorded to tape storage before being sent downstream for further processing. This happens entirely on-site, at CERN, Tier-0 being the entry point to the WLCG. The next step down the line is **Tier-1**, which facilitates both tape and disk storage while acting as a distribution hub for **Tier-2** centres as well as offering options for reprocessing raw data based on the analysis needs. Finally, Tier-2 facilities of the gird take on Monte Carlo (MC) simulations and additional data processing required for analysis [89]. Collaboration members and institutions can access the grid through their own computing infrastructure whether that amounts to a personal computer or a university cluster. Those are called **Tier-3** even though technically they are not part of the WLCG.

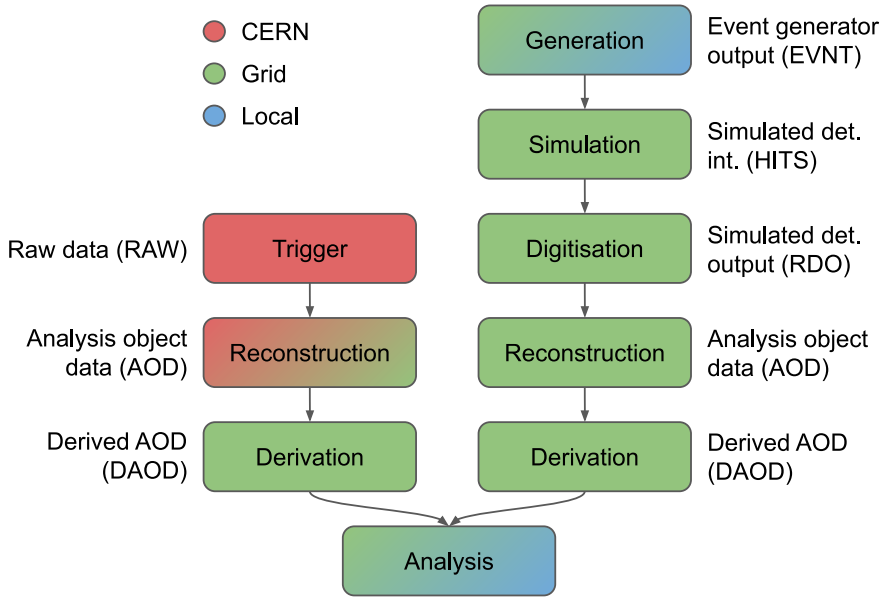


Figure 3.10.: Diagram of data model hierarchy featuring recorded detector data on the left and simulated data on the right [90]. The colour coding relates to where the processing steps take place within the WLCG hierarchy. Red bubbles are representative of ATLAS TDAQ and Tier-0 processing, green is inclusive of both Tier-1 and Tier-2 while blue amounts to Tier-3 or any other local resources.

In terms of the EDM itself, the data format is different depending on the processing stage. Figure 3.10 shows a visual representation of the data processing model featuring most of the intermediate stages between raw data or Monte Carlo (MC) and analysis-ready representations. At Tier-0 of the CERN grid, raw data is used to produce Event Summary Data (ESD) files, which are C++ [91] objects that don't make it to storage but are moreover processed to produce slimmer Analysis Object Data (AOD) files. Tier-2 centres use the AOD files to produce derivations, which are usually what the analysis teams are utilizing in the form of Derived Analysis Object Data (DAOD) files. Simulated data follows most of the same flow, with Raw Data Object (RDO) outputs of the digitization being the equivalent of raw recorded data. However, in order to arrive at the digitization inputs, hard scattering processes need to be simulated with MC event generators like PYTHIA or HERWIG, followed by a GEANT4 simulation of the ATLAS Detector. All those computations are commonly executed on Tier-2 grid hardware.

Hardware infrastructure is only one facet of the ATLAS data model. Software development and services are the other big part of ATLAS computing. In order to ensure consistent methodologies for interpreting and analyzing ATLAS data, a common analysis framework is used, namely ATHENA [10]. Developed and maintained continuously since 2009, ATHENA is a unified set of software tools covering most analysis tasks from basic reading and writing ATLAS-specific file formats to complex use cases such as entire MC simulation chains, event selection and filtering or jet tagging. Being constantly improved and maintained through

an internal Continuous Integration (CI) pipeline, distributing it to the entire collaboration reliably poses a real challenge. CERN has developed a custom read-only file system accessible through a virtual application called CERN Virtual Machine File System (CVMFS) [92, 93] in order to facilitate access to compiled common-use software and services such as ATHENA, ROOT [12, 94] and RUCIO [95].

4. Electron Identification Background

While being part of the ATLAS Collaboration is both an honour and a privilege, obtaining the *Qualified Author* status requires the completion of a Qualification Task (QT). This chapter is meant to document the activities carried out within the e/γ Combined Performance (CP) group that concluded my qualification as an ATLAS author. The results are also available as an ATLAS internal note [96].

4.1. Context and Motivation

The bulk of this QT concerned studying the different sources of non-prompt electrons and developing selection criteria for isolating the ones that originate from heavy-flavour decays. In ATLAS, electron identification is generally performed using a **likelihood discriminant** which is formed using variables constructed from measurements across the detector subsystems that the electrons interact with the most. An in-depth description of this process was presented and discussed in the e/γ group's electron and photon performance paper [97], published by the ATLAS Collaboration. Although many of the operating points of this discriminant offer good performance in terms of background rejection, there is still room for improvements, especially when it comes to rejecting heavy-flavour non-prompt electrons.

For the development of better electron identification (ID) methods, a good understanding of the different types of non-prompts is highly desirable. Therefore, there are numerous efforts in defining control regions for the several common possible sources.

My contribution focused on producing high-purity samples of heavy-flavour electron non-prompts which are invaluable in the pursuit of superior electron- ID tools. With the growing rate of adoption of NN-based methods across ATLAS, those samples could become instrumental in training and testing such models to improve certain physics objects' identification [98]. Most studies targeting electron- ID (with or without Machine Learning) should benefit from the outcomes of this work. And, outside this scope, any ATLAS analysis where heavy-flavour electrons are involved may borrow from the event selection model studied here to generate their own samples.

4.2. Analysis Strategy for Heavy-flavour Electron Background

The analysis explored the rather abundant $t\bar{t}$ process in an attempt to extract electrons resulting from the hadronisation of heavy-flavour quarks. The implementation of the selection algorithm was aided by the e/γ group's software framework: the TAGANDPROBE-FRAME [99]. This is the CP group's internal implementation of the *Tag and Probe Method*, which is more thoroughly described in section 4.2.1. A fork of this repository was created and the necessary changes have been made in order to integrate this heavy-flavour electron selection method.

Foremost, the selection process involved the application of event-level exclusion criteria, habitually referred to as cuts, meant to ensure good data quality and efficient targeting of the $t\bar{t}$ events. Those requirements entailed the following set of cuts:

- exclusion of events with no vertices containing more than two tracks [1]
- matching the data to a Good Run List (GRL)
- skipping events where errors were generated in the process of physics objects' reconstruction

In order to select for $t\bar{t}$ events, appropriate triggers needed to be chosen. In general, the *semileptonic* and *dilepton* decay modes of the $t\bar{t}$ pair are the easiest to target since they include at least one prompt electron and/or muon. As a consequence, the event-level selection criteria incorporated the requirement of at least one electron or muon to be triggered. Further $t\bar{t}$ content enrichment was obtained by implementing a cut on missing transverse energy of $E_{\mathrm{T}}^{\mathrm{miss}} \geq 50\,\mathrm{GeV}$, to account for the expected secondary $W \to \ell\nu$ decays.

Heavy-flavour electrons from $t\bar{t}$ events are in large part the result of the $b \to \ell$ and $b \to c \to \ell$ processes, which means that jet b-tagging is an important part of any analysis attempting to isolate them. For this purpose, the `MV2c10` tagger was used and configured at its 77% efficiency working point. This tagging algorithm was developed by the Flavour Tagging Working Group and its inner workings were presented in the ATLAS Public Note covering b-tagging performance [100]. Alongside b-tagging, overlap removal techniques are applied as well. At least two jets with $p_{\mathrm{T}} \geq 20\,\mathrm{GeV}$ are required to pass this procedure for the event to be further considered.

4.2.1. The Tag and Probe Method

One of the most commonly used algorithms in physics performance studies is the so-called *Tag and Probe method*. The main idea of this method is to use one set of physics objects called

[1] All decay modes of the $t\bar{t}$ pair result in b-jets, which are often characterized by secondary vertices with high track multiplicities.

tags, whose selection narrows down the collection of possible interactions that may have generated the event. Once the *tags* are found, a different selection is applied to find the *probes*, which are the objects that we are interested in studying. This method was used in a plethora of publications; a detailed description of it and its applications can be found in the conference lecture [101], covering the ATLAS *Common Analysis Framework*.

In this heavy-flavour study, once the event-level cuts have been applied, the *Tag and Probe* method was used to isolate the objects of interest. Firstly, the tags were chosen to be prompt electrons or muons, whose presence confirms the event includes one of the targeted $t\bar{t}$ decay modes. Table 4.1 shows the cuts that electrons and muons need to pass in order to be considered tag objects. The **fiducial cuts** mean that the electron has to be *central* ($|\eta| \leq 2.4$), but also requires it to not be detected in the crack region. As for the muon's case, the **combined** label, indicates that there were multiple different detector sub-systems involved in its reconstruction. Isolation cuts are expressed in terms of the ratio between transverse momentum and the $p_{T\text{cone20}}$ variable. This variable quantifies the sum of p_T reconstructed in the calorimeter from all physics objects in a radius of 0.2 (in units of ΔR). However, the likelihood discriminant working point can not be easily described in terms of efficiency, since it is not uniform with respect to the electron's kinematics. Thus, for a better understanding of what they mean, the Reference [97] would be the best source of information.

Electron	**Muon**
Passes overlap removal	Passes overlap removal
Fiducial Cuts	Is Central ($\|\eta\| \leq 2.4$)
Object Quality	Combined Muon
$p_T \geq 27GeV$	$p_T \geq 27GeV$
Track Isolation ($p_{T\text{cone20}}/p_T < 0.1$)	Muon Isolation ($p_{T\text{cone20}}/p_T < 0.1$)
Tight **likelihood discriminant** cut	

Table 4.1.: Selection criteria for tag objects

If at least one *tag* object is found, the algorithm loops through all of the non-tag container electrons searching for heavy-flavour candidates. Given that the goal was to maximize the purity of the heavy-flavour sample, the following cuts have been chosen:

- Very Loose **Likelihood Discriminant**
  Most reconstructed electrons are originating from light-flavour decays. The very loose likelihood operating point was designed specifically to provide a good exclusion of light flavour, while still keeping a considerable amount of heavy flavour.

- Loose Isolation
  This has been implemented using the IsolationSelectionTool developed by ATLAS' *Isolation and Fake Forum* and was described in Reference [102]. The *Loose Fixed*

Cut working point was used to amplify the effects of the *Very Loose Likelihood* cut, significantly increasing the purity of the resulting sample.

- Closeness to a b-tagged jet
  The ΔR between the candidate probe and all of the existing b-jets was calculated, requiring that the minimum ΔR found is ≤ 0.3 to confirm that the electron is inside of a b-jet.

- Significance of transverse impact parameter d_0 ($d_0$sig)
  For prompt electron tagging the value of $d_0$sig plays an important role. In this analysis, we use this parameter to achieve the opposite effects. Applying the cut $d_0\text{sig} \geq 5$ mitigates the prompt electron contamination of the output, furthermore increasing the heavy-flavour purity.

The *probes* were also characterized based on the $t\bar{t}$ decay mode associated with the event. This label was inferred based on the amount and type of *tag* objects found. A complete overview of the naming scheme is shown in Table 4.2. When there is a single *tag* object found, the *probe* label depends on its charge relative to the *tag*'s charge. *Probes* whose charge has the same sign (SS) as the prompt lepton *tag* are much less likely to be prompt electrons due to the nature of $t\bar{t}$ decays. Thus for some given data, if signal contamination is an issue, one could exclude the opposite sign (OS) *probes* in an attempt to ameliorate the problem. Decay modes of $t\bar{t}$ that involve τ leptons were not considered due to the additional difficulties posed by their reconstruction.

Decay Mode	Probe Label	Details
Semileptonic	elOS	Only one electron tag with opposite charge to the probe
	elSS	Only one electron tag with the same charge as the probe
	muOS	Only one muon tag with the same charge as the probe
	muSS	Only one muon tag with the same charge as the probe
Dilepton	el+el	Two tag electrons
	el+mu	One tag electron and one tag muon
	mu+mu	Two tag muons

Table 4.2.: Probe labeling scheme

4.3. Heavy-flavour Selection Results

This section presents the outcome of the heavy-flavour event selection in both in ATLAS real data and Monte Carlo (MC) simulation. Simulated events are by design perfectly reconstructed, but accessing the relevant ancestral particle in the decay chain in order to correctly match an electron to its background source is not a trivial task. In order to evaluate the heavy-flavour content of the MC samples the IFFTruthClassifier tool

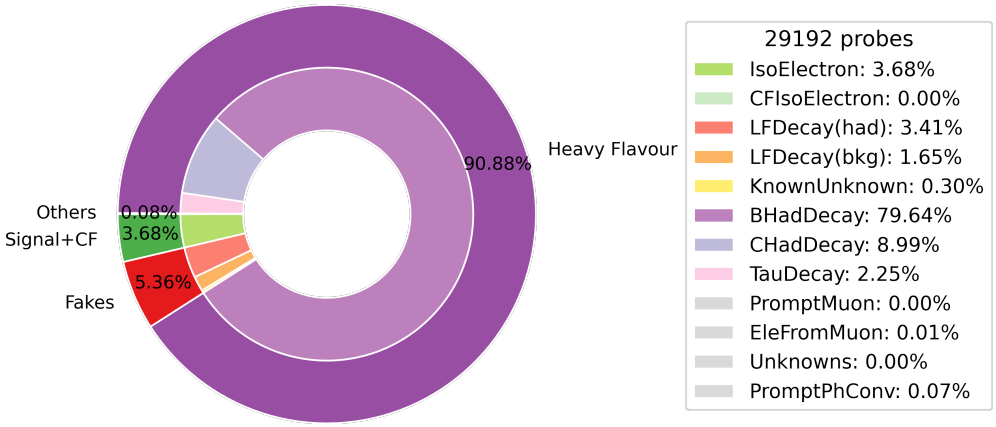


Figure 4.1.: Pie chart of probe electrons IFF Type after the selection cuts applied on **EGAM1** MC

(Reference [102]) was used, which labels electrons based on their source. As a part of the e/γ group, the Isolation and Fake Forum (IFF) specializes in addressing issues of misidentification of leptons and photons. The IFFTRUTHCLASSIFIER tool targets many background categories, but the ones relevant to this study are the following[2]:

- **IsoElectron**: prompt electrons that are properly isolated
- **CFIsoElectron**: prompt electrons that are properly isolated but their charge was flipped during reconstruction
- **LFDecay**: non-prompt physics objects originating from light flavour decays. The IFFTRUTHCLASSIFIER does not provide a distinction between physics object types, but this studies separates this category into **LFDecay(had)** which represents hadrons and **LFDecay(bkg)** which covers all other object types
- **BHadDecay**: non-prompt electrons originating from decays of b-hadrons
- **CHadDecay**: non-prompt electrons originating from decays of c-hadrons
- **TauHadDecay**: non-prompt electrons originating from hadronic τ decays
- **PromptMuon**: prompt muons that are properly isolated
- **PromptPhConv**: electrons originating from the conversion prompt photons
- **EleFromMuon**: muons that are wrongly reconstructed as an electrons
- **KnownUnknown**: any lepton that the tool should be able to classify, but there is not enough information contained by the event to do so
- **Unknown**: leptons that can not be attributed to any other category

[2] The IFF tool labels consist of numbered classes, but, for readability, in this thesis they will be represented as abbreviated text

The MC sample was a $t\bar{t}$ **EGAM1** derivation, which selects events with at least one well-identified (*Medium Likelihood*) electron, deeming it not a perfect fit for this study. **EGAM1** requirements imply that we can only access probes from the *el+el*, *el+mu* and *single el* decay channels of $t\bar{t}$. Unfortunately, no derivation covers all of the *semileptonic* and *dileptonic* channels while also providing all of the necessary variables necessary for this selection model. Thus, it was settled on **EGAM1** for the most part, but there was also some success in probing the muon channels with **EXOT17** derivations. This other derivation is similar to **EGAM1**, but its selection criteria apply to muons instead of electrons, requiring one prompt muon with $p_{\mathrm{T}} \geq 50\,\mathrm{GeV}$.

The heavy-flavour content of the *probes* passing the selection criteria is shown in Figure 4.1. The results were obtained after applying the algorithm to a sample of 29.5 million **EGAM1** $t\bar{t}$ events. In the Monte Carlo studies the heavy-flavour purity was found to have a value of 90.88%. To put this number in perspective, by counting all of the *probes* before applying the conditional selection, we would have gotten a heavy-flavour content of about 3%. The set of cuts applied on the *probes* was specifically designed to maximize purity, thus, this selection model operated at a very high background rejection point ($\approx$ 99.996%), representing the percentage of heavy-flavour probes that are rejected by the cuts. However, the stringency of the requirements also translates to a relatively low heavy-flavour efficiency: just shy of 1.3%.

Considering the numbers presented in the last paragraph, it is easy to conclude that a lot of data is necessary in order to produce a consistent sample of high-purity heavy-flavour electrons. Table 4.3 shows the yields in both real data and MC. The event yield is notably lower in data compared to the Monte Carlo derivation, but this is easily explainable by the fact that simulated data is comprised only of $t\bar{t}$ events, whilst real pp collisions can have a multitude of other outcomes beside top quark pair production.

Dataset	Total EGAM1 events	Events Selected	Probes Selected
MC	29,575,867	6,411,542	29,192
	100%	21.67 %	$9.87 \times 10^{-3}\ e/\mathrm{evt}$
DATA18	245,106,731	9,061,847	23,856
	100%	3.69 %	$9.73 \times 10^{-4}\ e/\mathrm{evt}$

Table 4.3.: Probe yields in real data and MC. The second column shows the total number of events that went in the event selection. Column three presents in both relative and absolute terms the number of events that passed the initial selection (i.e. a *tag* object was found). In the last column we show the number of *probes* passing the cuts and their occurrence rate relative to the total number of events

Probe yields are also greatly diminished in the data sample compared to the MC reference, which likely means a low $t\bar{t}$ selection performance of the event-level cuts. The requirement of a prompt lepton and two $p_{\mathrm{T}}^{\mathrm{jet}} \geq 20\,\mathrm{GeV}$ jets is more accommodating to W+jets events

than $t\bar{t}$, even at the upper end of the missing transverse energy spectrum ($E_{\mathrm{T}}^{\mathrm{miss}} \geq 50\,\mathrm{GeV}$) [103]. Cleaner $t\bar{t}$ signal regions in the lepton+jets channel could be achieved by requiring a high overall jet multiplicity (≥ 4) in lieu of an invariant mass reconstruction procedure, as exemplified by Reference [104]. In the context of the current analysis, the event-level cuts aren't stringent enough to cleanly separate $t\bar{t}$ from some of its main sources of background like W+jets, Multijet or Diboson (WW, ZZ, WZ). However, the additional constraints introduced by *tag* selection and the subsequent search for *probes* are expected to provide further background rejection that would be reflected in lower probe yields.

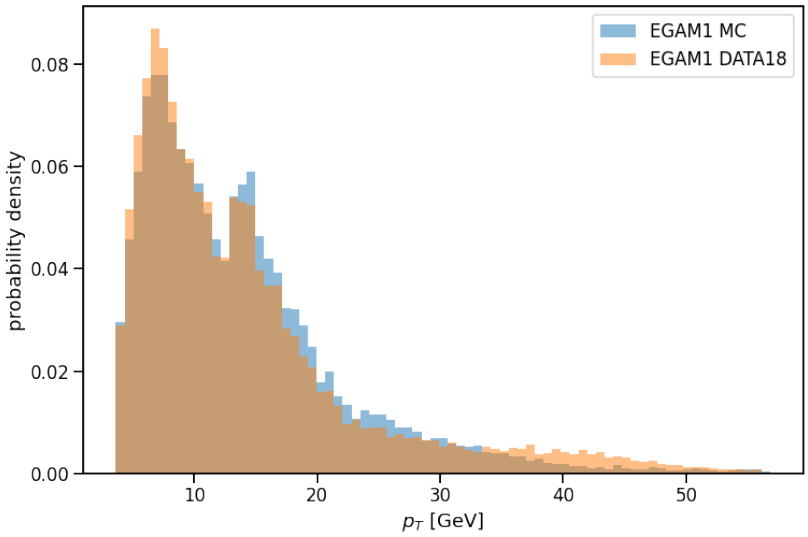
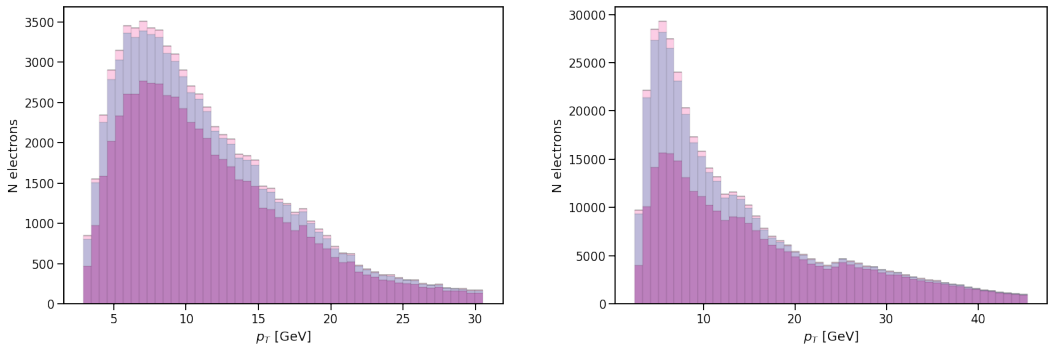


Figure 4.2.: The normalized distributions of heavy-flavour *probes'* p_{T} for **MC** (blue) vs **DATA18** (orange).

Having established how this selection model behaves when applied to simulated data, it is worthwhile to check if the same applies to real data. Figure 4.2 shows the normalized p_{T} distributions for the selected *probes*. It is important to note that *probes* coming from the *single muon* and *mu+mu* channels have been excluded from the **DATA18** histograms. Due to the **EGAM1** selection criteria, at least one *tag* electron should be present in every event. This means that any contribution from events with only muon *tags* are unlikely to pass the **EGAM1** selection. The p_{T} spectra in Figure 4.2 shows a good match between data and MC, with the exception of a slight excess in real data *probes* above 35 GeV. One obvious feature of these spectra is the sharp peak around 15 GeV. This shaping was found to be caused exclusively by the combined effects of the *probe* cuts. When checking the probe p_{T} distribution in MC, before the cuts and including only the heavy-flavour electrons tagged by their IFF type, there is no deviation from the expected spectrum shape. Figure 4.3(a) shows the effects of all of the cuts, except for *Very Loose Likelihood*, on the p_{T} distribution.

Analyzing this, we deduced that the biggest culprit is the likelihood cut, although its effects alone on the p_T spectrum shape (shown in Figure 4.3(b)) were not as pronounced as one would have expected. We concluded that the shaping effects of the *Very Loose Likelihood* are somewhat amplified by the rest of the cuts, thus obtaining the distributions exhibited in Figure 4.2.



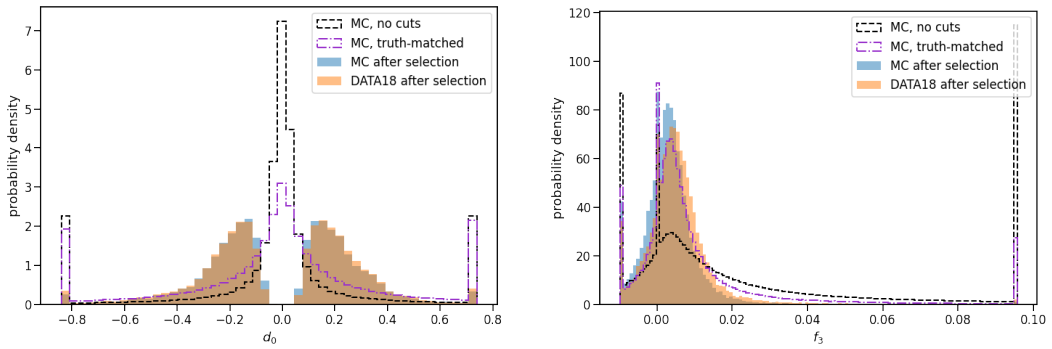
(a) (b)

Figure 4.3.: Stacked p_T histograms for three different sources of $t\bar{t}$ Monte Carlo heavy-flavour non-prompt electrons: b-decays (violet), c-decays (liliac), non-prompt τ decays (pink). Here, (a) shows the spectrum when applying all of the cuts except the likelihood and (b) shows what happens when only the *Very Loose Likelihood* requirement is applied.

4.3.1. Shower Shape Variables

In this subsection, we show the spectra for several variables of importance in the electron-ID process. The plots present the regions where the electrons picked by the heavy-flavour non-prompt selection model are found in the shower shape variable space. We compared the normalized distributions of those quantities after the probe selection for both simulated and real data. For reference, the distributions for all probes before the selection cuts and truth-matched probes to b-quarks (in MC only) are also included. The plots in this section have been put together in order to arrive at a qualitative representation of the algorithm's effects on the variables relevant to electron identification. As stated in the previous section **DATA18** probes have been chosen only from events with at least one *tag* electron, the exclusion of the other channels being justified by considering the **EGAM1** skimming.

Figure 4.4(a) displays the distribution of transverse impact parameter values. Without the heavy-flavour selection criteria, the shape of this spectrum resembles a zero-centred and narrow Cauchy distribution. It appears the application of cuts resulted in the exclusion of electrons whose d_0 is in the vicinity of zero. Since there is a requirement specifically targeting electrons for high significance values of this impact parameter, this shape of the d_0 distribution is hardly surprising. The distribution of truth-matched electrons to b-hadron



(a) Transverse impact parameter

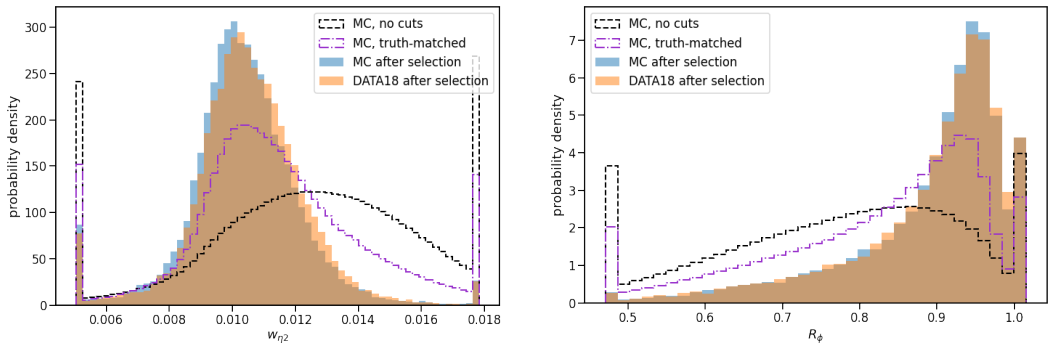
(b) Ratio of the energy in the third layer to the total energy in the EM calorimeter

Figure 4.4.: Normalized distributions of d_0 (a) and f_3 (b), electron- ID variables. Electron samples have been obtained by applying the selection model to **EGAM1** derivations of **Monte Carlo** (blue) and **DATA18** (orange). The reference distributions are presented for all probes before the selection (black) and probes truth-matched to b-quarks (Purple) for **MC** events only.

decays shows that indeed heavy-flavour *probes* should be more spread in terms of d_0 but there is no inherent discontinuity around the zero value.

The electron identification variable associated with the 3rd layer of the EM calorimeter is f_3 energy ratio, whose distribution is shown in Figure 4.4(b). The range of this quantity, if we were to look at all electrons indiscriminately, would be much broader ($-0.03 \leq f_3 \leq 0.47$), which is what we have in the histogram associated with all *probes* before the selection criteria are applied. The heavy flavour non-prompt electrons selected with this method, in **MC**, seem to settle around a mean of $\bar{f}_3 \approx 0.003$ which is one order of magnitude lower than the average f_3 for all *probe* electrons in $t\bar{t}$. **DATA18** electrons' distribution is slightly shifted to the right, but this difference is present regardless of the existence of *probe* selection criteria. A narrower lateral shower profile in simulation is a behaviour consistently found in most studies (e.g. Ref. [105]). Another feature worthy of note is the high count of zero values. The most likely cause of those was the electrons in the crack region ($1.375 \leq |\eta| \leq 1.5$), where there is no 3rd layer of the calorimeter. The true distribution for electrons from b-jets shows a similar shape relative to the one obtained by applying the cuts, with roughly the same mean value, but a bit more spread.

Moving forward to variables concerning the 2nd layer of the EM calorimeter, the distribution of lateral shower widths is presented in Figure 4.5(a). This distribution suffers several changes after the application of *probe* selection cuts. Its mean gets slightly shifted towards the left and there is a considerable narrowing effect induced by the cut-flow. The distribution median for **DATA18** is slightly shifted towards the right, but again this effect is present even without the application of cuts. Regarding the true distribution of heavy-flavour electrons from b-jets, the plot shows a similar density curve compared with the outcome of the *probe*



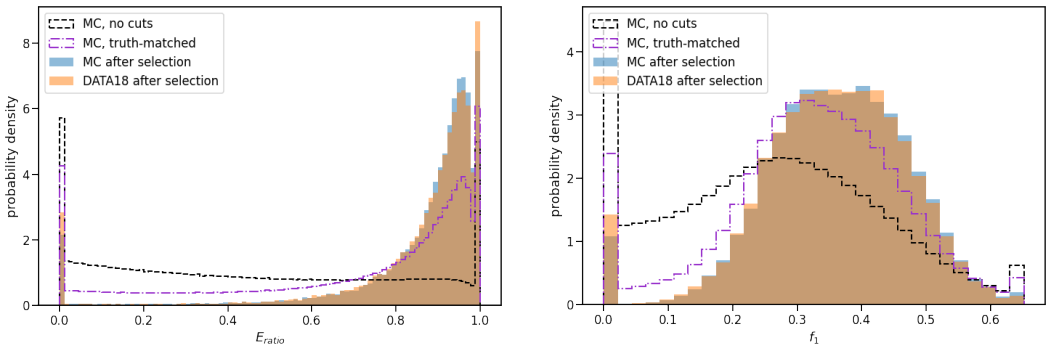
(a) Lateral shower width along η

(b) Ratio of the sum of the energies of the cells contained in a $3 \times 3 \times (\eta \times \phi)$ rectangle to the sum of the cell energies in a 3×7 rectangle, both centred around the most energetic cell

Figure 4.5.: Normalized distributions of $w_{\eta 2}$ (a) and R_ϕ (b), electron identification variables related to the second layer of the Electromagnetic (EM) calorimeter. Electron samples have been obtained by applying the selection model to **EGAM1** derivations of **Monte Carlo** (blue) and **DATA18** (orange). The reference distributions are presented for all probes before the selection (black) and probes truth-matched to b-quarks (Purple) for **MC** events only.

selection, but there is a notable difference concerning the right tail, which is much fatter for the truth-matched electrons. When it comes to the other variable in this category, shown in Figure 4.5(b), the distribution is being affected by selection cuts in a milder fashion. Its left tail would have been concave were the probe cuts not been applied. On the other hand, the narrowing effect of the cuts on the spectrum's shape is once again present. Adding the truth information into the mix, it is apparent that the selection cuts produce a stronger shaping, manifested as a skimming of the left tail when compared to the true distribution.

The electron identification variables derived from energy deposits in the first EM calorimeter layer are displayed in Figure 4.6. The plot on the left, namely Figure 4.6(a), shows the distribution of the ratio of the energy difference between the maximum energy deposit and the energy deposit in a secondary maximum in the cluster to the sum of these energies. Judging by how left-skewed this distribution is, we can safely conclude that the heavy-flavour non-prompt electrons are most likely to have values of E_{ratio} closer to 1. This statement was confirmed by the shape of the true distribution, which is left-skewed as well, but with a significantly larger spread. In Figure 4.6(b) the distribution of ratios of the energy measured in the first layer of the Electromagnetic calorimeter to the total energy of the EM cluster. For the heavy-flavour electrons selected, this distribution becomes much more light-tailed compared to the general case. Most *probes* were found around the mean value $\bar{f}_1 \approx 0.36$, but for a significant fraction of the electrons, the value of f_1 is zero. Note that for prompt electrons, at least, the mean value of f_1 has a negative correlation with p_{T}. The values of zero or below are associated with the electrons either in the crack region or closer



(a) (b)

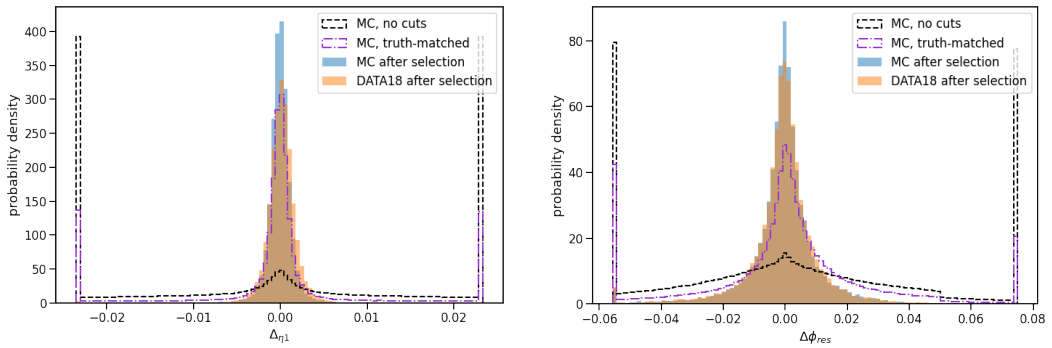
Figure 4.6.: Normalized distributions of E_{ratio} (a) and f_1, (b), electron- ID variables related to the first layer of the EM calorimeter. Electron samples have been obtained by applying the selection model to **EGAM1** derivations of **MC** (blue) and **DATA18** (orange). The reference distributions are also presented for all *probes* before the selection (black) and *probes* truth-matched to b-quarks (Purple) for **MC** events only.

to the beamline ($|\eta| \geq 2.4$). The shape of the truth-matched electrons' distribution is similar to the output of the selection cuts, but the mean is a bit closer to zero and again there is more spread associated with the true distribution.

Finally, we looked at shower shape variables associated with track-cluster matching. Figure 4.7(a) presents the distribution of $\Delta\eta_1$. In **MC**, this is a symmetric, zero-centred distribution with a standard deviation of $\sigma = 0.001$. Compared to the same plot but for all $t\bar{t}$ *probes*, this $\Delta\eta_1$ distribution has a roughly 20 times smaller σ, meaning that heavy-flavour electrons tend to find themselves closer to the mean. However, **DATA18** electrons seem to be a bit more spread, the distribution showing some positive skew. This disagreement between real data and MC is well known for prompt electrons and is caused by the deformations of the calorimeter under its own weight, which is not taken into account by the simulation tools. Regarding $\Delta\phi_{\text{res}}$, in Figure 4.7(b), the same behaviour was observed. The region richest in heavy-flavour non-prompts is proximal to the value zero and the data-simulation mismatch is way less pronounced. For both plots in Figure 4.7 the distribution of truth-matched electrons to b-jets confirms that heavy-flavour non-prompts lie close to the value of zero in terms of both $\Delta\eta_1$ and $\Delta\phi_{\text{res}}$. However, the probe selection cuts seem to be pretty stringent, selecting a much narrower vicinity around zero, compared to the truth.

4.4. Conclusion of the Qualification Task

Exploring the different sources of electron fakes may result in compelling advantages including, but not limited to, the improvement of electron identification tools. The Qualific-



(a) $\Delta\eta$ between the cluster position in the first layer of the EM calorimeter and the extrapolated track

(b) $\Delta\phi$ between the cluster position in the second layer of the EM calorimeter and the momentum-rescaled track, extrapolated from the perigee, times the charge q

Figure 4.7.: Normalized distributions of $\Delta\eta_1$ (a) and $\Delta\phi_{\text{res}}$ (b), shower shape electron identification variables related track-cluster matching. Electron samples have been obtained by applying the selection model to **EGAM1** derivations of **MC** (blue) and **DATA18** (orange). The reference distributions are presented for all probes before the selection (black) and probes truth-matched to b-quarks (Purple) for **MC** events only.

ation Task was focused on the heavy-flavour non-prompt electrons and concluded with the development of an event selection model aimed at creating high-purity samples of electrons of this type. With a minimal set of cuts comprised of *Very Loose Likelihood*, *Loose Isolation*, the significance of d_0 and proximity to a b-jet, purities of over 90% were achieved, although, with some compromises in heavy-flavour efficiency. This model was designed for targeting the *semileptonic* and *dilepton* final states of $t\bar{t}$, but it may be adapted to some other processes with relative ease. Analyzing the distributions of shower shape variables provided a useful understanding of the control regions where heavy-flavour electrons can be found, while also highlighting the relevant features of those particular types of fakes. The study was concluded with a presentation of this analysis within the e/γ group's plenary meeting after which my qualification as an ATLAS author was complete.

Both the selection model and the samples themselves have the potential to be improved and extended. With the expected widening of adoption for Neural Network (NN) models by Combined Performance groups, high-purity background data samples would become increasingly valuable. As Run 3 starts, the ATLAS Collaboration will collect more and more data, which would allow for even bigger heavy-flavour background samples. At the same time, there is room for improvement in the event selection criteria. A subsequent more in-depth study may be able to increase the signal efficiency of this method while minimizing the shaping effects over the shower shape variables' distributions.

5. Data Analysis Tools and Machine Learning

The bulk of my individual contribution is closely related to the application of Machine Learning (ML) methods to physics problems. This chapter is meant to provide a general introduction to this class of analysis models and methodologies. Section 5.1 aims to give an overview of Machine Learning (ML) methods and their vast areas of applications, followed by Section 5.2 which introduces Neural Network (NN) models. Lastly, Section 5.3 covers the process of *training* ML models with a focus on gradient-based optimization. The content of this chapter is closely related to my teaching activity, which I undertook for the duration of my PhD program. I passed on this knowledge to several bachelor's degree physics students who showed a keen interest in computational physics and choose to follow internships at my home institute **IFIN-HH**. In the elaboration of this chapter, the most relevant bibliographical reference is the book *Deep Learning* by Goodfellow, Bengio and Courville [106].

5.1. Machine Learning Methods

Analyzing data from observations has always been an integral part of the scientific method, regardless of the field. At the most fundamental level, the process of successfully formulating and testing hypotheses is inextricably linked to the capabilities of identifying the underlying patterns behind the observed data. As more and more knowledge is acquired in any particular field of research, subsequent refinements of our understanding require increasingly better technology for more precise measurements as well as a greater volume of (often more complex) data. Data analysis itself can become progressively more demanding, making exhaustive testing of the many possible underlying models to be unfeasible.

Pattern recognition concerns algorithmically uncovering underlying data regularities [107], commonly by using a highly-parameterizable statistical model. The design and study of those types of models fall under the umbrella of Machine Learning (ML), which also defines the procedures for **learning** those parameters based on the available data. In most cases, *learning* amounts to an optimization problem for a given objective. One such objective may be **classification**, where the model is optimized to assign correct categorical labels to the data points. However, the objective could just as well amount to mapping the data point to a real-valued scalar, also known as **regression**. Generally, methods involving a known desired output are categorized as **Supervised Learning**. There are applications where such *labels* could not be known from the start and they pertain to the domain of **Unsupervised**

Learning. Label availability is not necessarily a binary characteristic of a dataset. There are middle-ground cases where labels could be partially available, covering only fractions of the dataset. Methods specifically designed for such scenarios do exist and are usually referred to as **Semi-supervised Learning** [108].

5.1.1. General Formalism and Terminology

A Machine Learning problem involves searching for the unknown mapping between a set of input objects $\mathcal{X}$ and a set of targets $\mathcal{Y}$, displayed in Equation 5.1. Most often than not, the exact function g is not tractable, so the solution aims at finding an approximation f, as in Equation 5.1, that is as close as possible to the truth.

$$\text{Exact function: } g : \mathcal{X} \to \mathcal{Y} \qquad \mathbf{y} = g(\mathbf{x}) \tag{5.1}$$

$$\text{Approximation: } f : \mathcal{X} \to \mathcal{Y} \qquad f \approx g \tag{5.2}$$

This approximation f is *learned* based on the available examples in the dataset through a process commonly referred to as **training**. For supervised learning models, the dataset, defined in Equation 5.3, contains n pairs of examples $\mathbf{x}_i \in \mathcal{X}$ and $\mathbf{y}_i \in \mathcal{Y}$, with $\mathbf{y}_i = g(\mathbf{x}_i)$.

$$\mathbf{D} = \{(\mathbf{x}_i, \mathbf{y}_i);\ i = \overline{1, n}\} \tag{5.3}$$

The input objects $\mathbf{x}_i$ may take many different forms, based on the nature of the recorded observations, but they need to be **independent**. In most cases, observations themselves are made up of multiple variables, conventionally called **features**. As Equation 5.4 describes, an observation $\mathbf{x}_i$ can be defined as a d-dimensional vector of scalar features x_i^j. Most ML models require the entire dataset to have uniform dimensionality, thus allowing the input objects to be represented as points in $\mathbb{R}^d$ space (or even $\mathbb{C}^d$ in some cases [109]).

$$\mathbf{x}_i = (x_i^{(1)}, x_i^{(2)}, \ldots, x_i^{(d)}) \tag{5.4}$$

Thus, the entire dataset, assuming a constant number of features, can be easily expressed as a matrix. Equation 5.5 shows the matrix representation of the same dataset from Equation 5.3. This object is often called **model matrix** or **design matrix**.

$$\mathbf{D} = \begin{bmatrix} x_1^{(1)} & x_1^{(2)} & \ldots & x_1^{(d)} \\ x_2^{(1)} & x_2^{(2)} & \ldots & x_2^{(d)} \\ \vdots & \vdots & \ddots & \vdots \\ x_n^{(1)} & x_n^{(2)} & \ldots & x_n^{(d)} \end{bmatrix} \tag{5.5}$$

Data representations aside, a model also needs a way to generate predictions. Typically, this involves a set of model parameters $\mathcal{P}$ and a collection of mathematical rules that describe

how the output is calculated for given inputs and said parameters. While the model's computation logic may be rigid, its parameters are variables that are being *learned* during training, in an attempt to **fit** the dataset provided. The *training* of *fitting* process translates to an optimization task whose end goal is to find the best possible set of parameters allowing the model to approximate the true mapping g.

Lastly, we need to define what constitutes a good *fit* and how to measure its quality. The function describing the *goodness of fit* may bear different but equivalent names across literature, such as **objective function**, **cost function** or **loss function**. It has the important role of quantifying the errors of the predictions and is often the quantity undergoing the optimization process during training. Denoting the model under the set of parameters $\mathcal{P}$ by $\mathbf{f}_{\mathcal{P}}$, a Mean Absolute Error (MAE) loss function could be written as shown in Equation 5.6.

$$\mathcal{L}\left(\mathbf{f}_{\mathcal{P}}(\mathbf{x}_i), \mathbf{y}_i\right) = \frac{1}{d}\sum_{j=1}^{d}\left|f_{\mathcal{P}}^{(j)}(x_i^j) - y_i^{(j)}\right| \tag{5.6}$$

Depending on the data, other loss functions better reflect the prediction error or be easier to optimize. Regardless of the choice in the loss function, as long as it can be optimized, it can also be used in finding the model parameters $\tilde{\mathcal{P}}$ that best approximate the true underlying mapping. Equation 5.7 is illustrative of the loss minimization procedure for model training. Some of the more common loss functions are presented and discussed in Appendix B.

$$\mathbf{f}_{\tilde{\mathcal{P}}} = \underset{\mathcal{P}}{\operatorname{argmin}}\frac{1}{n}\sum_{i=1}^{n}\mathcal{L}\left(\mathbf{f}_{\mathcal{P}}(\mathbf{x}_i), \mathbf{y}_i\right) \tag{5.7}$$

5.1.2. Models and Assumptions

Any algorithm using **trainable** parameters to infer *a priori* unknown information about new data points is a Machine Learning model. One of the simplest classes of ML architectures assumes a linear relationship between the input objects and the targets and is thus named **linear models**. Equation 5.8 shows a linear model with d-dimensional inputs and a scalar output.

$$\mathbf{f}_{\mathcal{P}}(\mathbf{x}_i) = \mathbf{w}^T\mathbf{x}_i + b = \begin{bmatrix} w^{(1)} & w^{(2)} & \dots & w^{(d)} \end{bmatrix}\begin{bmatrix} x_i^{(1)} \\ x_i^{(2)} \\ \vdots \\ x_i^{(d)} \end{bmatrix} + b \tag{5.8}$$

Linear models, for every component of the output vector, have a set of weights $\mathbf{w}$ assigned to each input feature and a bias term b. In the most general case, for d_1 inputs and d_2

outputs, the model parameters are represented by the weights matrix $\mathbf{W} \in \mathbb{R}^{d_1 \times d_2}$ and the bias vector $\mathbf{b} \in \mathbb{R}^{d_2}$, as in Equation 5.9 They can be used for **regression** problems as well as **classification**, by setting a threshold for the output.

$$\mathcal{P}_{\text{regression}} = \{W_{ij};\ i = \overline{1, d_1},\ j = \overline{1, d_2}\} \cup \{b_j;\ j = \overline{1, d_2}\} \tag{5.9}$$

Other approaches involve different assumptions about the data. For example, one may assume that for inputs with small distances between them, the targets would also be proximal to each other. This kind of assumption is perfectly embodied by the *k*-Nearest Neighbors (KNN) algorithm [110]. Denoting the set of k objects closest to the $\mathbf{x}_i$ object in the feature space by $D^k_{\mathbf{x}_i}$, Equation 5.10 shows how such a model would generate predictions.

$$\mathbf{f}_{\mathcal{P}}(\mathbf{x}_i) = \frac{1}{k} \sum_{i:\mathbf{x}_i \in D^k_{\mathbf{x}_i}} y_i \tag{5.10}$$

KNN is particularly interesting because the training dataset is used as the set model parameters $\mathcal{P}_{\text{KNN}} = \mathbf{D}$. There is no training process involved, but the value of k can be tuned by the user. In this case, k is what is known as a hyperparameter.

If the opposite assumption from KNN is considered true, one might use a **decision tree** algorithm instead. Those allow for predictions that are not continuous or uniform in any way. There are several ways of creating decision trees, the most used one involving Iterative Dichotomizer 3 (ID3) [111].

Choosing ML algorithms and models is synonymous with making implicit assumptions about the problem and the data. On average, tested on all possible sets of problems, all models and optimization methods will perform the same, as stated by the *No Free Lunch* theorem [112].

5.2. Neural Networks

Many Machine Learning (ML) models can act as universal function approximators and model architecture choice is data-assumption dependent. There are, however, other desirable properties of ML models, like generalization power. Even a regressor decision tree, provided enough depth, can approximate any arbitrary function, but it is unlikely to generalize well when applied to data far outside the domain of the training set.

NN were developed in an attempt to increase the generalization power of the previous ML models by borrowing ideas from biological networks of neurons. They are created by stacking multiple linear models on top of each other and adding non-linear functions to every layer's output. The introduction of non-linear functions is mandatory since multiple

linear models stacked together by themselves would be equivalent to a single-layered linear model with different parameters.

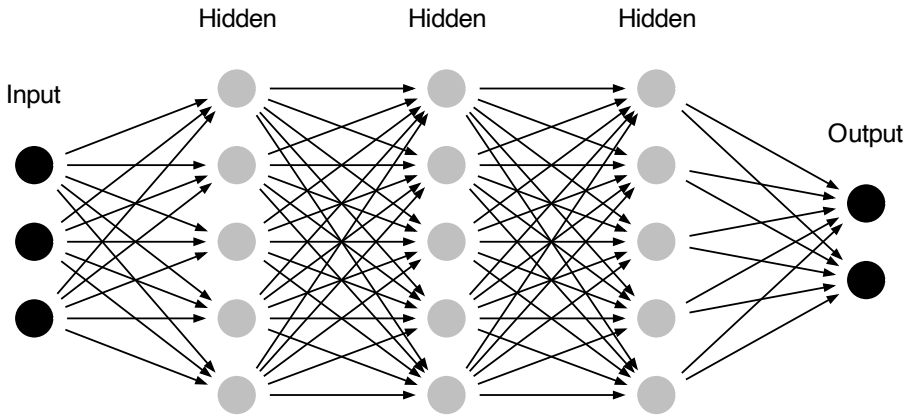


Figure 5.1.: Diagram of a Neural Network (NN) with three hidden layers

Figure 5.1 shows a **fully-connected** NN with inputs $\mathbf{x} \in \mathbb{R}^3$, outputs $\mathbf{y} \in \mathbb{R}^2$ and three *hidden layers* in-between. The hidden layers contain intermediate representations of the input data and a single such layer is necessary for the network to become a universal function approximator [113]. In this graph representation of the network, layers are vectors and their components are represented as nodes, which sometimes may also be called **neurons**. Graph edges are unidirectional and represent the data flow through the network, they are associated with the weights and biases of the previous nodes.

The model presented in Figure 5.1 is known as a **feed-forward** NN, meaning that the values within a n-dimensional layer k are dependent on the values of the previous layer $k-1$. Values of the neurons are obtained by applying the **weights** and summing the contributions from the neurons in the previous layer, followed by adding the **bias** term and applying the non-linear function (called **activation function**). This process is described by Equation 5.11. Note that for the following equations, the subscript would be used to index vector components while the superscript indicates the layer number.

$$x_i^k = f\left(\sum_{i=1}^{n} w_{ij}^{k-1} x_i^{k-1} + b_i\right) \tag{5.11}$$

Equation 5.12 generalizes Equation 5.11 to an entire layer. This generalization is depicted in

matrix form by Equation 5.13.

$$\mathbf{x}^k = f\left(\mathbf{W}^{k\,T} \times \mathbf{x}^{k-1} - \mathbf{b}^k\right) \tag{5.12}$$

$$\begin{bmatrix} x_1^k \\ x_2^k \\ x_3^k \\ \vdots \\ x_n^k \end{bmatrix} = f\left(\begin{bmatrix} w_{1,1}^k & w_{2,1}^k & w_{3,1}^k & \dots & w_{m,1}^k \\ w_{1,2}^k & w_{2,2}^k & w_{3,2}^k & \dots & w_{m,2}^k \\ w_{1,3}^k & w_{2,3}^k & w_{3,3}^k & \dots & w_{m,3}^k \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ w_{1,n}^k & w_{2,n}^k & w_{3,n}^k & \dots & w_{m,n}^k \end{bmatrix} \begin{bmatrix} x_1^{k-1} \\ x_2^{k-1} \\ x_3^{k-1} \\ \vdots \\ x_m^{k-1} \end{bmatrix} - \begin{bmatrix} b_1^k \\ b_2^k \\ b_3^k \\ \vdots \\ b_n^k \end{bmatrix} \right) \tag{5.13}$$

In terms of non-linear activation functions, **sigmoid**, the **hyperbolic tangent** and Rectified Linear Unit (ReLU) are the most commonly used in NN models. The choice of activation functions has a significant impact on the optimization process as well as overall model performance. More extensive coverage of the definitions and properties of activation functions can be found in Appendix C.

Even though a single hidden layer network can approximate arbitrary functions, deeper models obtained by increasing the number of layers provide better outcomes than wide models with more nodes per layer [114]. Deep Neural Network (DNN) models also have an advantage in terms of parameter count scaling. Two fully connected layers with n nodes require n^2 parameters for weights and n bias terms. A change of a factor $\alpha \in \mathbb{N}^*$ in the layer's widths will require more parameters than using α layers, as shown by Equation 5.14.

$$(\alpha n)^2 + \alpha n \geq \alpha(n^2 + n),\ \alpha \geq 1 \tag{5.14}$$

5.3. Gradient-Based Optimization

The training process involves iteratively adjusting the parameters of the model in order to get closer to the desired results. This translates into an optimization procedure that attempts to find the minima of the model's loss function and retrieve the set of parameters $\hat{\mathcal{P}}$ associated with the minimum loss. This process was previously shown in Equation 5.7, but this section is going to provide further details about the optimization process itself. Several optimization algorithms could be used but in most cases, gradient-based optimization provides the fastest convergence rate, which is important, especially when exploring sizable parameter spaces.

Computer science developments such as **automatic differentiation** [115] allowed complex ML models to be easier to design and train. Since NN models are differentiable, gradient descent is one of the best-known ways of training them [116–118]. It also allows for highly

parallelizable model training, facilitated by modern Graphics Processing Unit (GPU) hardware. Gradient descent involves using the gradient of the loss function to update model parameters in an iterative process, as shown in Equation 5.15. The current iteration number is represented by i, while $\mathbf{D}$ denotes the training dataset (defined in Equation 5.3)

$$\mathcal{P}_i = \mathcal{P}_{i-1} - \eta \nabla \mathcal{L} \left(\mathcal{P}_{i-1} \,|\, \mathbf{D} \right) \tag{5.15}$$

The η parameter is called *learning rate* and sets the magnitude of the parameter update. For a Neural Network with n layers, the parameters would be the weights and biases of those layers, as shown in Equation 5.16

$$\mathcal{P} = \left\{ \mathbf{W}_1, \mathbf{W}_2, \mathbf{W}_3, \ldots, \mathbf{W}_n, \mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3, \ldots, \mathbf{b}_n \right\} \tag{5.16}$$

5.3.1. Gradient Descent Algorithms

As with simpler types of models, DNN training requires computing the loss function gradient $\nabla \mathcal{L}$ with respect to every model parameter. For model inference the data passes through the network sequentially, the values of a layer k depending on the values of the previous layer. Gradient calculations, however, have to start from the output and propagate backwards. Thus, this process is named **backpropagation**. An example of analytical gradient computation using **backpropagation** is provided in Appendix D.

No gradient-based optimization method is guaranteed to locate the local minima of the loss function, but the ML community is constantly developing better methods of navigating the parameter space. Those algorithms using the gradient of the loss to apply parameter updates are referred to as **optimizeers**.

Stochastic Gradient Descent (SGD) is one of the simpler **optimizers**, relying on the logic described by Equation 5.15, but using only a subset of the training data at each step. Splitting a dataset in **batches** has become a procedure almost universally used in ML because it accelerates the convergence towards the minima by significantly speeding up computations.

Notable extensions of SGD exist, such as Root Mean Square Propagation (RMSProp) [119], which on top of using bach computations, also adapts the learning rate η to each parameter's gradient magnitude. In order to avoid saddle points in the parameter space and increase the likelihood of finding the global minima, some optimizers use **momentum** in the parameter space to adjust the direction and magnitude of the gradient. The *Adam* optimizer [15] is one of the most widely used because it keeps the idea of parameter-specific learning rates while also using momentum.

Figure 5.2 shows how SGD, RMSProp and *Adam* attempt to find the minima of a 2-dimensional parameter space, starting from the same point. Given the simplicity of the problem, all methods arrive at the minima within the 100 optimization steps, but the paths

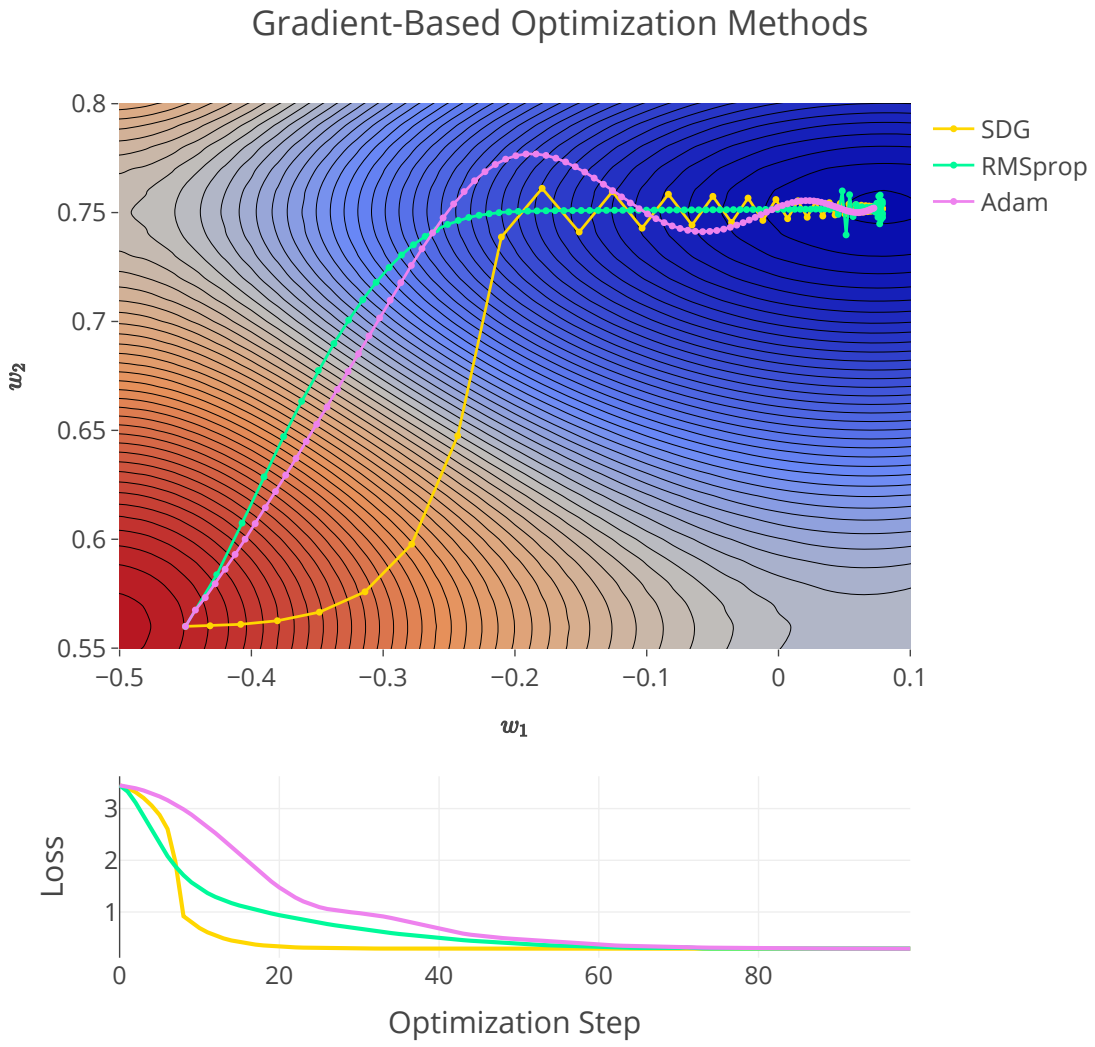


Figure 5.2.: Toy example of different optimization algorithms attempting to find the local minima of a 2D parameter space. The upper plot shows the algorithm's path through the parameter space for 100 optimization steps. In the bottom plot, the value of the loss function is compared between the three approaches, at every step.

they take are significantly different. In the making of Figure 5.2, the reference learning rate η was the same for all optimizers. SGD does not use adaptive learning rate, so the steps it took were the largest of the bunch. As a consequence, SGD showed the fastest-decreasing loss values, but it became increasingly less efficient after it arrived closer to the minima. In contrast, RMSProp showed a smoother convergence due to the adaptive step size. *Adam*'s behaviour is close to that of RMSProp, but the additional inclusion of momenta makes its trajectory resemble a **damped oscillation**. While *Adam* was the slowest converging of the three optimizers, it is also the one that landed closest to the point of minima, without showing major fluctuations around its vicinity.

5.3.2. Model Training Heuristics

Optimizing a model to achieve minimum loss on the *training* dataset does not always relate to good generalization power. In order to test this, the available data is split into multiple datasets. The training dataset is used for gradient-based optimization, while there is another dataset for testing purposes. *Test* data is never used in the training process but rather evaluated after the training is complete. A model whose performance is similar on both the training and testing datasets shows the ability to generalize well to new data points. The opposite scenario is called **overfitting**, where a model may perform well on training data, however, the performance is never translated to the *test* dataset evaluations.

Most models need more optimization steps than the number of batches in the training dataset. It is a common practice to use the same data points multiple times during training. SGD optimizers, as well as their extensions, involve *batching* the dataset, each batch corresponding to a loss evaluation followed by a gradient-based parameter update. After all current batches have been used, the training data is shuffled and new batches are created. This process is repeated several times until the loss value stops decreasing. A full pass through the training dataset's batches is known as an **epoch**.

Overfitting can be assessed even during training, by designating another smaller dataset, commonly referred to as *validation data*, which is used to evaluate the loss function at the end of each epoch. The loss evaluation on the validation dataset is compared with the average loss on the training dataset. As such, when those loss values start diverging, it is a clear indication that the model is overfitted.

Early stopping the training process is one of the ways to prevent overfitting. Another way would be to introduce model regularization. By introducing an additional penalty term to the loss function, the model parameters are prevented from taking on large values. Regularization terms added to the loss function are dependent on the model's weights. $L1$-regularization involves adding together the absolute values of the weights, as shown in Equation 5.17. This type of penalty incentives sparsity throughout the model's weights,

effectively acting as a feature selection mechanism.

$$\mathcal{L}_{L1} = \mathcal{L} + \lambda \sum_{j=1}^{p} |w_j| \tag{5.17}$$

On the other hand, there is $L2$-regularization, which instead of introducing sparsity, enforces the weights to have smaller values. Equation 5.18 displays a loss function with an $L2$-penalty term.

$$\mathcal{L}_{L2} = \mathcal{L} + \lambda \sum_{j=1}^{p} w_j^2 \tag{5.18}$$

In both scenarios, the λ hyperparameter controls how much regularization is applied. Penalty terms corresponding to $L1$ and $L2$ are not mutually exclusive and can be applied together, each with a different regularization parameter λ.

6. New Small Wheel Machine Learning Trigger

This chapter covers the Research and Development (R&D) for a New Small Wheel (NSW) trigger algorithm based on Machine Learning (ML) deployed at the hardware level and targeting the *phase II* upgrade of the ATLAS Detector. Before diving deeper into it, Section 6.1 will provide the general aspects of the muon trigger design, before and after the installation of the New Small Wheel (NSW). The main bibliographical resources consist of the relevant Technical Design Report (TDR) documents [42, 120, 121]. The next part, Section 6.2 will showcase the research going into the design and testing of a muon trigger prototype, using Machine Learning, for the MicroMegas NSW data.

6.1. ATLAS Muon Trigger

As previously discussed in Secion 2.2.5, the Level-1 (L1) trigger searches for, among others, signatures from high-p_{T} muons. This section is going to cover more in-depth the inner workings of the ATLAS muon trigger focusing on the L1 component.

6.1.1. Original Design

In their first iteration, the Level-1 muon trigger used information from the Resistive-Plate Chambers (RPC) and the Thin-Gap Chambers (TGC), while L2 trigger received information from the regions containing possible trigger objects. These regions are called Regions of Interest (RoIs) and the L2 trigger uses them for information on coordinates, energy, and type of signatures in order to filter out unnecessary detector signals.

In terms of detector design, muon trigger chambers were chosen to have a timing accuracy compatible with a clear identification of the bunch-crossing containing the muon candidate. The trigger in both the barrel and the end-cap regions is based on signals from several stations each. Muon candidates are found by searching for a coincidence of hits in the different trigger stations within a road tracing back to the interaction point (IP). The system was designed for concurrent operations within six thresholds, three associated with the

low-p_T trigger ($6 - 9$ GeV) and three associated with the high-p_T trigger ($9 - 35$ GeV) [41], as it is shown in Figure 6.1.

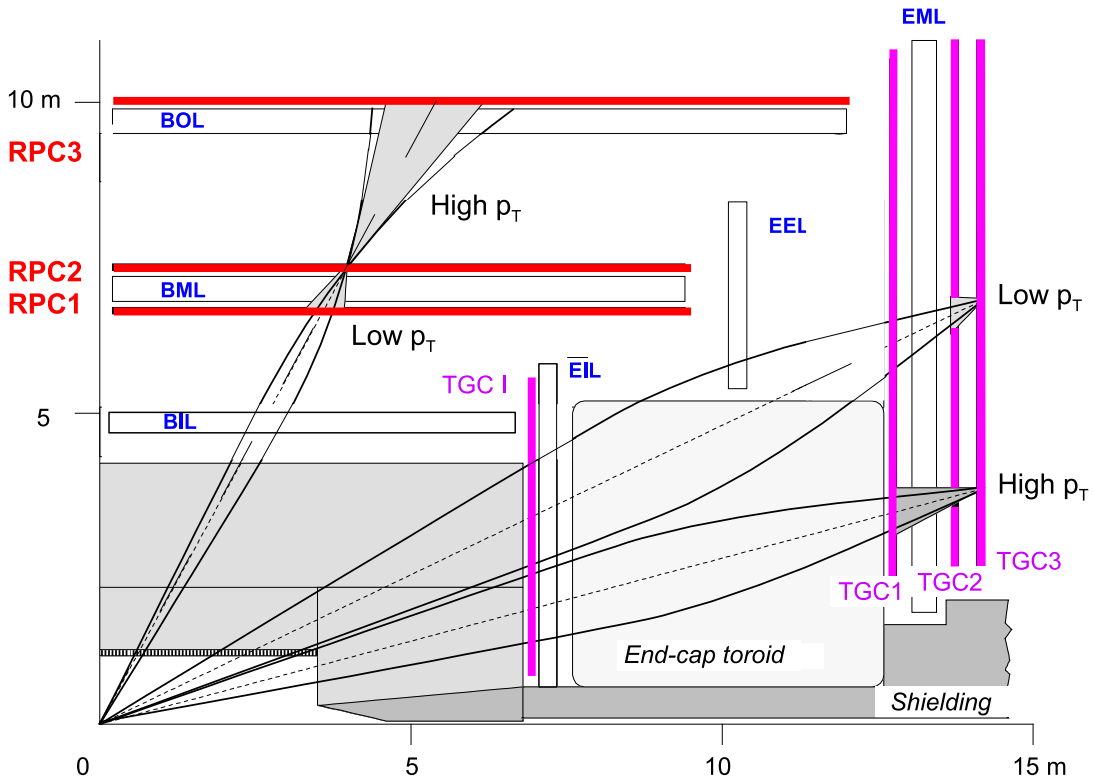


Figure 6.1.: Schematic representation of the L1 muon barrel and end-cap triggers showing tracks for the low-p_T and high-p_T muons. [41]

Resistive-Plate Chambers (RPC) detectors in the barrel region have a space-time resolution of 1 cm $\times$ 1 ns. They can work at an event rate of up to $1\,kHz/cm^2$. As Figure 6.1 shows, the Resistive-Plate Chambers stations are disposed as follows. There are two **Barrel Middle** stations (RPC1 and RPC2), arranged on either side of a Monitored Drift Tubes (MDT) and one **Barrel Outer** station, (RPC3) mounted on the inside MDT barrel outer stations. When a hit is generated in the second RPC station, a search for a corresponding hit is made in the first RPC station within a road tracing back to the IP. The width of the road is a function of the cut on p_T: the smaller the width of the road, the higher the cut on p_T. This is motivated by low-p_T muons being subject to larger trajectory changes due to magnetic fields. Figure 6.1 also showcases muon roads associated with different p_T thresholds. For the high-p_T muons, information from the third RPC chamber is used in conjunction with information from the first two in order to establish their tracks throughout the detector.

On the other hand, the end-cap trigger uses the signals generated by Thin-Gap Chambers (TGC) detectors, which have a weaker time resolution than the 2 ns of the RPCs but provide

a bunch-crossing identification efficiency greater than 99% for the original 25 ns bunch spacing of ATLAS. Thin-Gap Chambers detectors have a great rate capability of more than 20 kHz/cm^2 and are arranged in nine layers of gas volumes, grouped into four planes. Located seven meters in the z-direction, the TGC inner station (TGCI) consists of one plane of doublet units. Seven meters further in the z-direction, there are seven layers arranged in one plane of triplet chambers (TGC1, closest to the IP) and two planes of doublet chambers (TGC2, TGC3). This layout is included in Figure 6.1, as well.

The results from the muon barrel and end-cap trigger processors form the input to the Muon to Central Trigger Processor Interface (MuCTPI). The information includes the position and p_{T} threshold for up to two muon track candidates. Additional functions of the MuCTPI are to provide data to the L2 trigger and the data acquisition system for events selected by L1. The Level-2 (L2) trigger system receives a subset of all muon candidate's information which is decoded as Regions of Interest. Those RoIs sent to L2 are ordered according to decreasing p_{T}. In the original ATLAS design, during the 2.5μs delay of L1-trigger, the event data is buffered in the detector electronics' memory. If the L1-trigger selects the event, this data will be transferred to the High-Level Trigger for further processing.

6.1.2. New Small Wheel Phase I Trigger

With the installation of the New Small Wheel detectors, the barrel region is largely unaffected, but the end-caps receive a significant upgrade. The trigger rate in this region is expected to be reduced by about 30% by minimizing fake triggers coming from non-pointing tracks [42], as indicated in Figure 6.2.

The track segments for the trigger system are produced by the newly deployed MicroMegas (MM) and the small-strip Thin-Gap Chambers (sTGC) detectors, presented more thoroughly in Subsection 2.3.2 and Subsection 2.3.3 respectively. In terms of the trigger algorithms, they both work on similar principles, attempting to extrapolate local muon tracks back to the IP.

Micromegas Trigger Logic

Since MicroMegas strip data has a good resolution provided by the 0.5 mm strip pitch, it is not necessary to calculate layer centroids. Figure 6.3 gives a general overview of the trigger logic, with 6.3 (a) reminding detector operating principles and 6.3 (b) showing the strip configuration for the planes forming a quadruplet.

The MicroMegas trigger algorithm is illustrated in Figure 6.3 (c). At first, hits are transformed into slopes of infinite-momentum tracks using LUTs. The entire area covered by the plane is divided into *slope roads* which are binned in terms of pseudorapidity values of straight tracks connected with the IP. The slope bins are mapped to several strip addresses

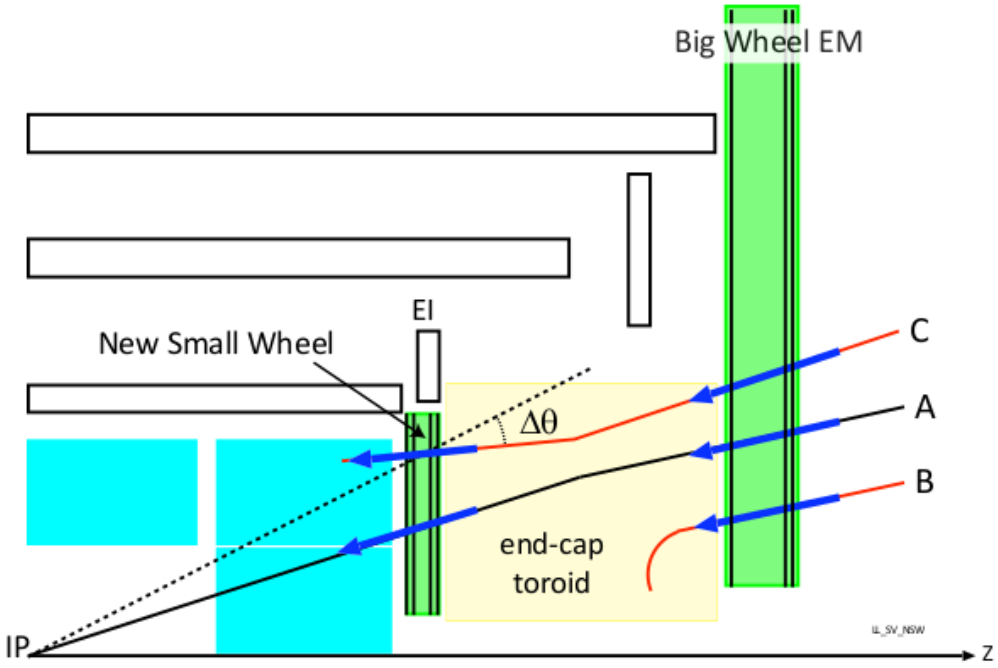


Figure 6.2.: Schematic representation of the muon end-cap trigger. Out of the tracks shown in this figure, only the one labelled *A* will be accepted because it connects with the interaction point [42]

each. Every bunch-crossing the number of strip hits within all *slope roads* is recomputed, with hits remaining buffered for three bunch-crossings before being discarded. A track candidate is defined as a multiple-layer coincidence within the same slope roads. Every bunch-crossing, a coincidence check takes place and once a track candidate is found, its local slope is calculated using the least square fit. Accounting for the known geometric relationships between the IP and the MM planes, track candidate quality is assessed by computing $\Delta\theta = |\theta_{\text{global}} - \theta_{\text{local}}|$. The higher the value of $\Delta\theta$, the less likely it is for the muon track to originate from the IP. After being cross-checked with sTGC data, the candidate muon track is projected through and matched with the segments from the Big Wheels [122].

sTGC Trigger Logic

The sTGC detector is primarily used for triggering which means its associated trigger logic is more complex than the one deployed for MM. A general overview of the sTGC trigger is shown in Figure 6.4, with 6.4 (a) showcasing the elements of a layer and 6.4 (b) displaying the quadruplet arrangement.

General trigger operation principles are schematically depicted in Figure 6.4 (c). Pads play

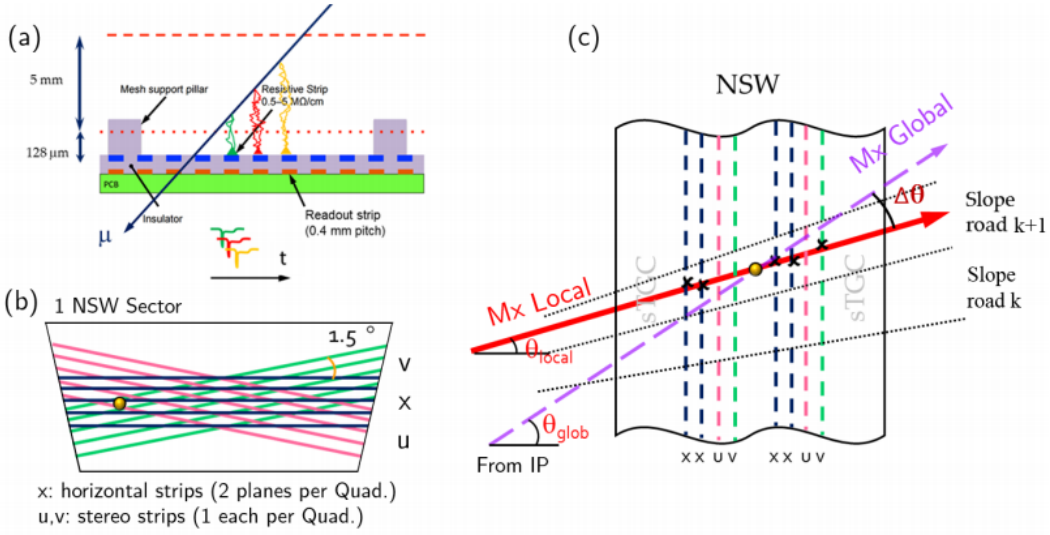


Figure 6.3.: (a) Micromegas detector structure, (b) readout strips in quadruplet, (c) illustration of segment reconstruction achieved with eight MM strip planes [122]. The *Mx* local track is reconstructed from the strip hits while the *Mx* global track is similarly reconstructed, but accounts for the interaction point as well.

an important role in this process, covering large areas and thus allowing their signals to be used in narrowing down the region where strip readout happens. From those strip signals, layer centroids are found by analyzing the amount and distribution of collected charge. In the next step, quadruplet centroids are calculated by averaging the layer centroids, with Look-up Tables (LUTs) further used to find the value of $\Delta\theta$ for the segment. Similarly, the *R*-index for the segment is determined by a different set of LUTs [122].

6.1.3. Phase II Trigger Prospects

While *phase I* brought major changes to the muon trigger system by introducing new detectors, *phase II* does not aim to change detector layout and instead will improve the Trigger and Data Acquisition (TDAQ) system. After the *phase I* upgrade, the entire trigger process should be completed within a 1025 ns window starting from the time of the interaction. *Phase II* operation is expected to accommodate longer latency in order to provide time for more complex trigger algorithms to run.

The trigger scheme planned for the *phase II* upgrade foresees a potential two-level hardware-based trigger system alongside the software-based HLT. This means the addition of a Level-0 (L0) step before the L1. The L0 combined with L1 will have an increased latency compared to the current system (6 μs compared with 2.5 μs) and also an increased acceptance rate of 1 MHz (ten times more than before the upgrade). Thus, it will be possible to maintain

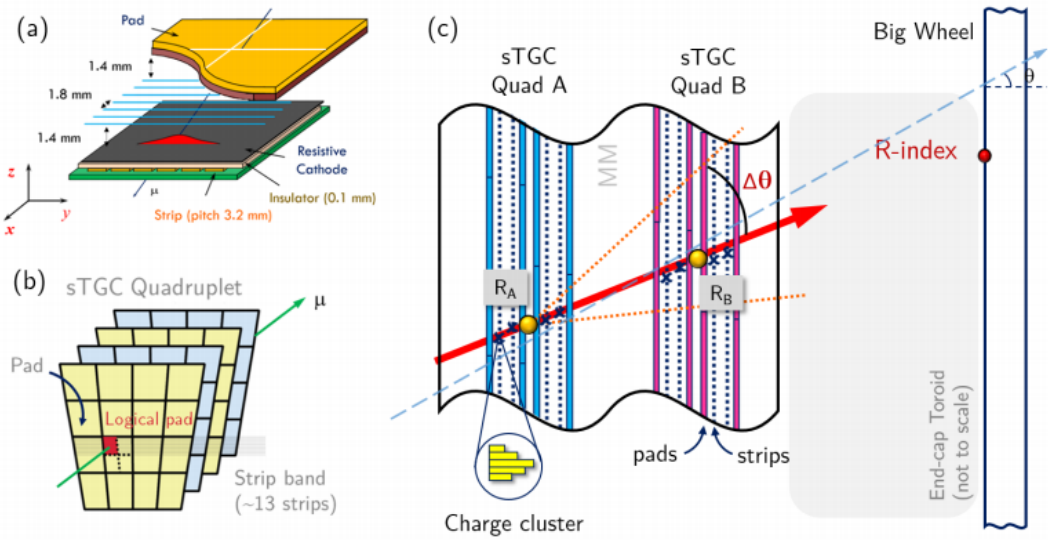


Figure 6.4.: (a) sTGC detector structure, (b) readout strips and pads in quadruplet, (c) illustration of segment reconstruction achieved with eight sTGC strip planes [122]

a p_T threshold as low as $p_T \geq 4\,\text{GeV}$ despite the increased luminosity. In this scenario, the L0 trigger rate should be 4 MHz and around 800 kHz for L1.

Using the increased latency of the first level trigger, the system could be improved by using the Monitored Drift Tubes (MDT) data in the L0 muon trigger. This addition has the potential to improve trigger selectivity. Fake coincidences may easily be rejected when checked against the MDT data.

For the L0 muon trigger, as can be shown in Figure 6.5, the data goes from the muon chambers to the trigger and readout electronics in the counting room. Detector data is sent to the end-cap Sector Logic (SL) and barrel SL boards, which determine the trigger candidates that are sent over to MDT trigger processor in order to obtain an improved measurement of the muon momentum. Based on the values of the momentum, the muon candidate can be accepted or rejected. The final muon L0 trigger decision is sent to the ATLAS central trigger through the Muon to Central Trigger Processor Interface (MuCTPI) board. [42]

During Run 3, the signals from the MicroMegas detectors and sTGC are processed separately. Each NSW sector is making use of two Field-Programmable Gate Arrays (FPGAs), one for each detector technology. FPGAs are preferred for this task because the trigger decision requires very narrow time frames, considering a time between bunch-crossings of 50 ns. The *phase II* upgrade presents an opportunity for data coming from both detector technologies to be processed together on the same FPGA. A block diagram of the

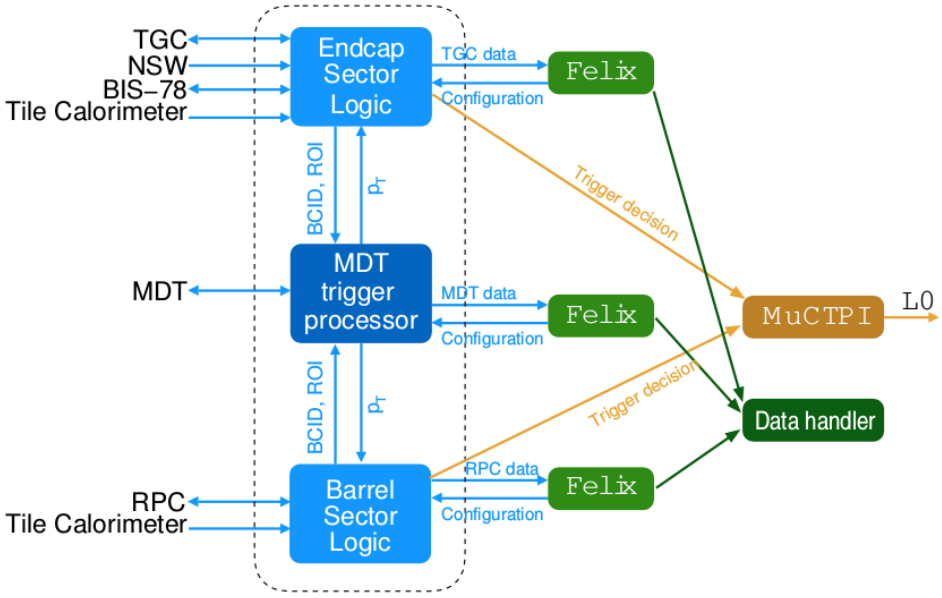


Figure 6.5.: Diagram representing the data flow for the L0 trigger. The muon trigger decision is made by the SL boards from the barrel and end-cap. These boards use data collected from the muon trigger chambers (TGC and RPC), calorimeters and the MDT trigger processor. The trigger decisions made by different SL modules are sent to the MuCTPI and the data associated with them is read out through the Front End Link Interface eXchange (FELIX) and passed from the FELIX modules to the HLT. [42]

NSW sector trigger processor off-detector board is shown in Figure 6.6. This implies that the firmware needs to be redesigned. [123].

6.2. Hardware-Level Machine Learning Micromegas Trigger

With the *phase II* upgrade scheduled for the end of Run 3, planned software and hardware contributions are already actively being developed. Regarding the NSW trigger algorithm, there are several worthwhile potential options. In recent years there have been many successful attempts at implementing Machine Learning (ML) algorithms on FPGAs and this type of approach is worth exploring for such trigger applications as well. FPGA firmware is designed at Register Transfer Level (RTL) which is inherently a very complex low-level task. Using a Hardware Design Language (HDL) to implement ML would be a cumbersome endeavour since a high-level software framework is invaluable to efficiently designing Neural Network (NN) architectures. However, newer technologies are constantly bridging this gap with tools like HLS4ML [124]. The idea of translating algorithms written in higher-level languages into HDL code is not new. There have been tools for High-Level

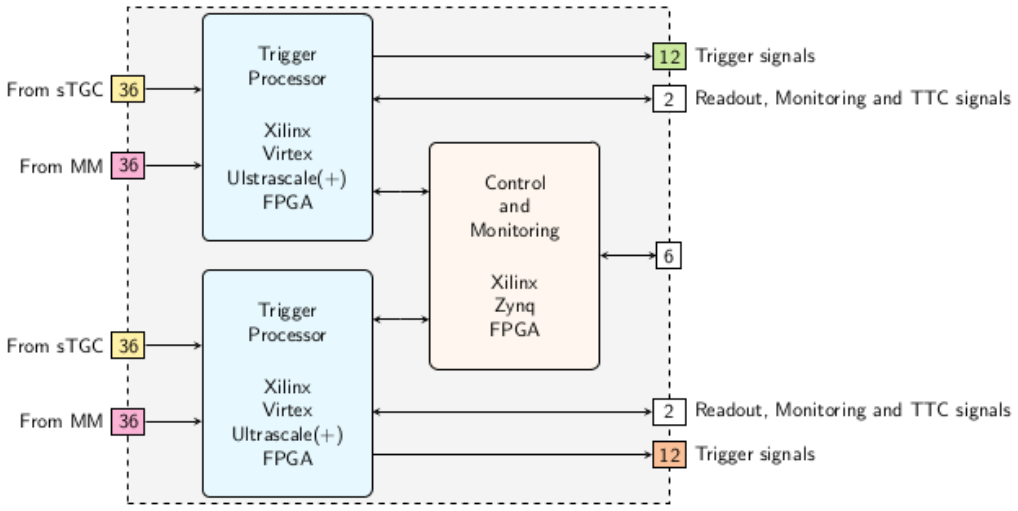


Figure 6.6.: Block diagram of the NSW trigger processor off-detector board after phase-II upgrade [123]

Synthesis (HLS) being used both in industry and research during the last decade. HLS4ML is just one of the currently available tools that push this concept into the realm of Machine Learning. A conceptual representation of its workflow can be seen in Figure 6.7.

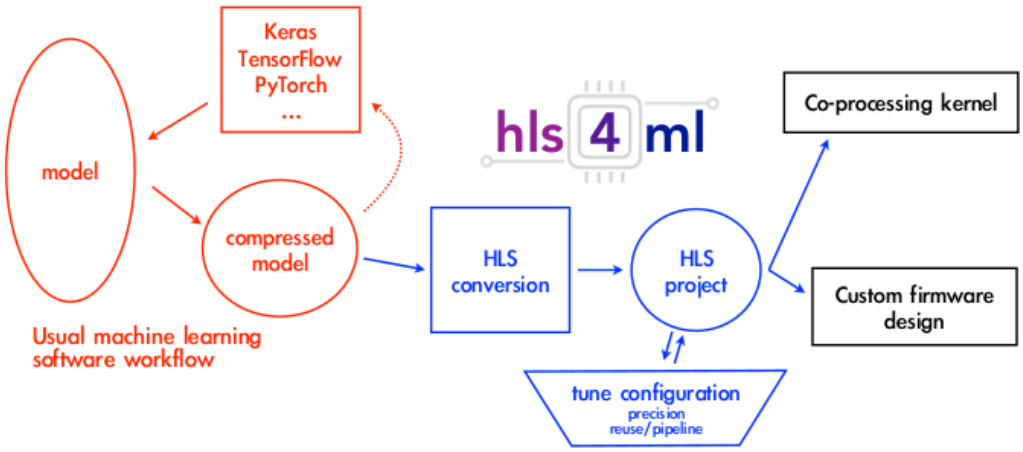


Figure 6.7.: Workflow for translating a machine learning model to an FPGA using HLS4ML [124]

Considering how recent this type of software technology is and the fact that current generation FPGA resources are still significantly limited when compared to dedicated ML hardware, the first milestone of such Research and Development needs to be a proof-of-concept. With this scope in mind, the initial prototype aims at deploying hardware-level

Neural Network models that, using just the FPGA's compute capabilities, successfully manage to reproduce the behaviour of the currently implemented *phase I* trigger without exceeding the time budget. To the extent of this thesis's contribution, only the MicroMegas detectors were modelled so far, since failure with this simplified setup would invalidate the method for the more complex sTGC.

6.2.1. Data generation

The NSW geometry is already implemented within the ATHENA [10] framework, allowing for Monte Carlo (MC) GEANT4 simulations of muon tracks through the newly installed detectors. Figure 6.8 shows the outcome of such a simulation in terms of MicroMegas hits, 2D projected on the xy-plane. Small sectors are depicted with blue shades while big sectors are shown in shades of red. Regardless of the sector type, all the MM Printed Circuit Boards are outlined individually.

As Section 6.1 covered, a MM trigger processor FPGA receives information from an entire sector, meaning that the MM hardware-trigger algorithm is not affected by what happens in other parts of the NSW. The trigger logic for MM signals can be easily reduced to a purely geometric problem involving computing slopes of local and global tracks. One such track is represented in Figure 6.9. It is apparent that several strips end up firing in the vicinity of the muon hit and the strip number and plane number are linearly dependent for the four **outer** MM layers. Given the stereo angles for the **U** and **V inner** four layers, the strip number are slightly shifted up and down respectively, when compared with the **X** layers.

In order to further simplify the proof-of-concept setup, only one sector is being considered and planes are given rectangular shapes, keeping the 0.4 mm strip pitch. The number of strips is rounded to 8800 while keeping the stereo plane's configuration with 1.5° and -1.5° inclined strips (relative to the x-axis). Those changes are reflected in Figure 6.10, where the *slate blue* shows X planes, the U *orange* depicted planes have strips angled at 1.5° and the V represented with *pink* have -1.5° inlined strips.

Using this setup also required fewer computational resources for simulation by allowing the use of lighter software tools. The public GitHub code-base of A. Wang [11] was used as a starting point for developing the simulation code that replicates the chosen simplified configuration. A data sample of MM detector responses for one million simulated muon tracks was generated, with ROOT [12] outputs containing hit locations in terms of the strip and plane number as well as the $\Delta\theta$ associated with the track.

6.2.2. Neural Network Trigger

Muon tracks through a MM sector show a well-defined **pattern** in terms of the strip and layer number and it can be visually represented, as Figure 6.9 shows. This prompted the

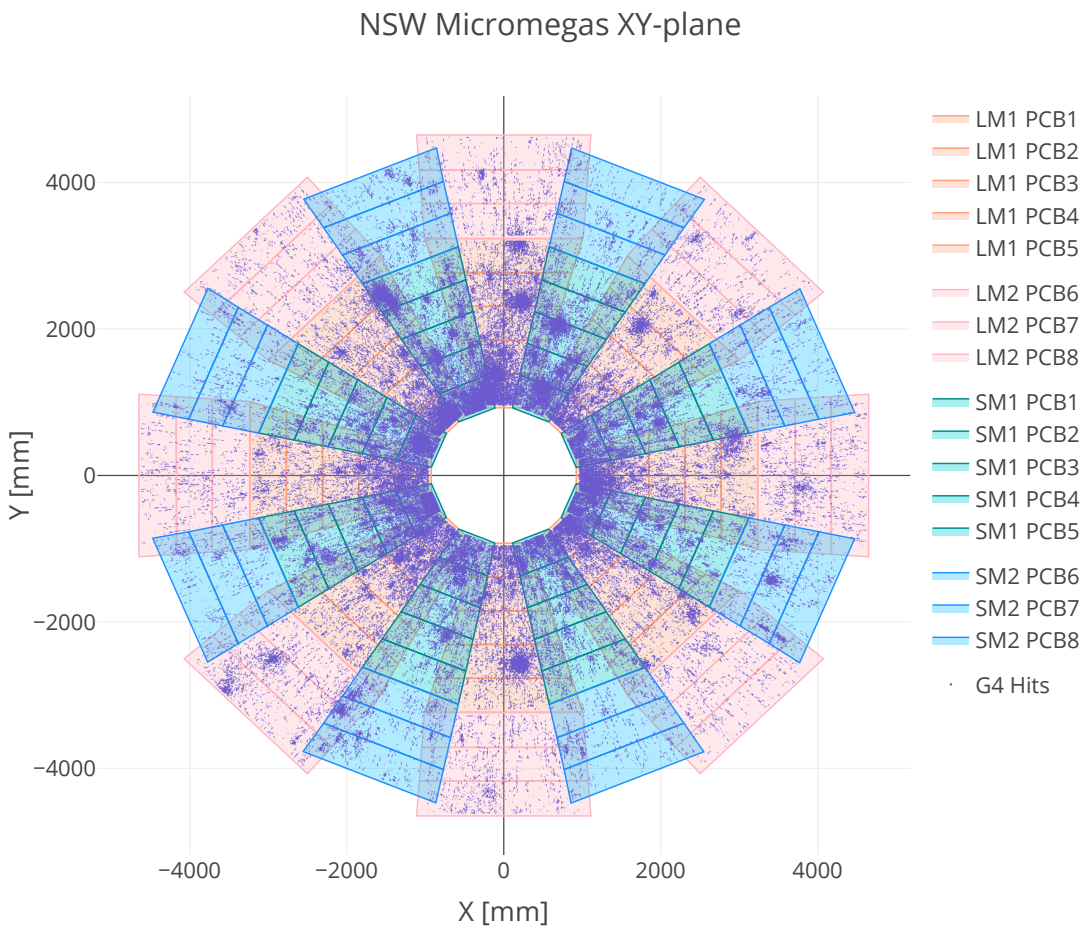


Figure 6.8.: GEANT4-simulated muon hits' position within an *xy*-view of the NSW

exploration of *computer vision* algorithms which are specialized in identifying geometrical patterns. In terms of Machine Learning, this can be achieved by Convolutional Neural Network (CNN) models which take images as input.

The *state* of the detector can be encoded as an image if we account for all of the strips across the eight layers at a given time. A sparse binary matrix $M_{8800\times 8}$ can encompass all of the MM information since it holds a distinct element for every strip. Thus, if the strip i within the plane j registered a hit $M_{i,j} = 1$ and otherwise the value would be 0. Treating all those matrix elements as pixels results in a very narrow image containing the location of every registered hit.

With the $M_{8800\times 8}$ matrix as the model input, the next step is to define a target. Early iterations tested this approach in a **classification** scenario, with the network being tasked with differentiating cosmic tracks from prompt muons. Such a task proved to be trivial for a

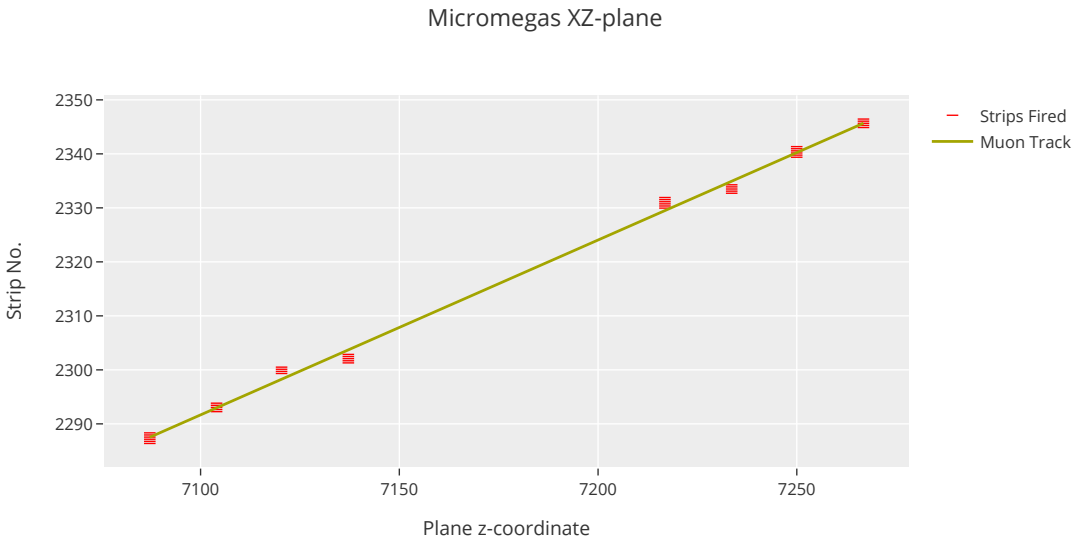


Figure 6.9.: Muon track through a MicroMegas sector shown in green, with the fired strips shown as red horizontal lines and the x-axis is labelled in mm.

CNN so the next step was to attempt a **regression** of the track quality measure $\Delta\theta$.

Fitting within the FPGA resource limitations implies finding a balance between the number of model parameters and the model's performance. The hit matrix being quite large, accommodating 8800 strips, prompted creative dimensionality reduction techniques. Thus, the building blocks for the model were *convolutional blocks* which contain a convolution layer, a max pooling layer Rectified Linear Unit (ReLU) activation. Figure 6.11 shows the first such *block*, applied to the input matrix.

While prototyping, a network with three *convolutional blocks* seemed to be the best compromise between regression performance and model complexity. Max pooling was applied heavily in order to reduce the vertical dimensionality of the input by a factor of four, after each block. The full model is represented in Figure 6.12, containing a chain of three convolutional blocks whose final output is flattened and used to compute the predicted target: $\Delta\theta$.

The design behind this architecture prioritized heavily a small number of parameters, to facilitate FPGA deployment. Table 6.1 shows every layer of the model alongside how many trainable parameters are associated with it. *Batch normalization* [13] and *dropout* [14] were used for regularization purposes.

Model training made use of the Adam optimizer [15] with default parameters and was facilitated by Graphics Processing Unit (GPU) acceleration by an *Nvidia Tesla V100S* board.

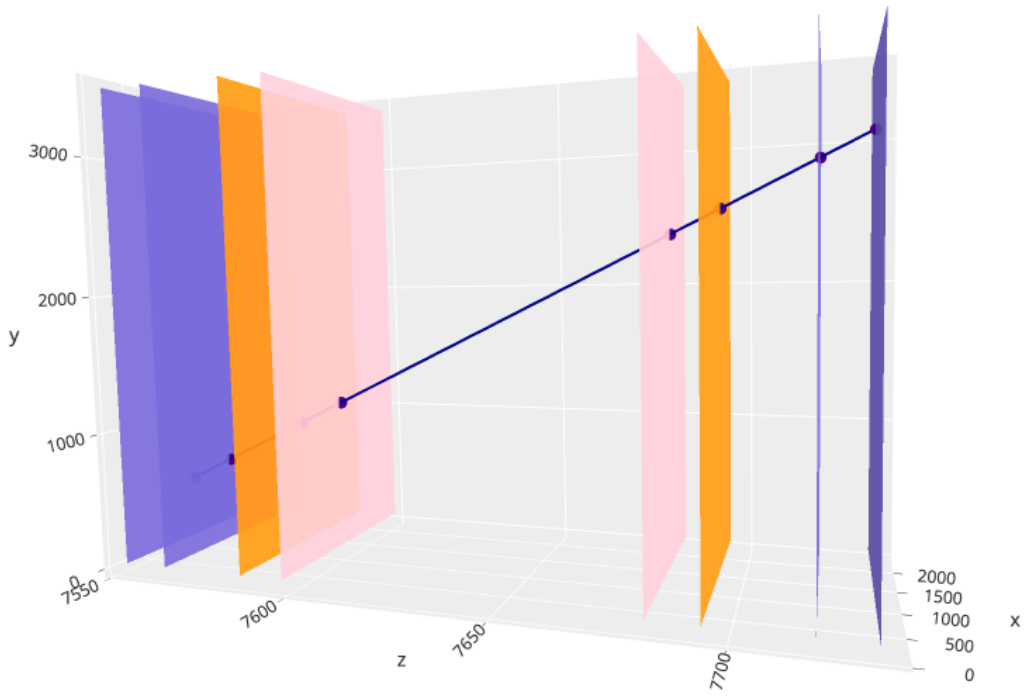


Figure 6.10.: 3D representation of a muon track within the simplified MM assembly setup. The values on all of the three axes are measured in mm.

Batch sizes were set to 500 while the dataset was split 80%/10%/10% between training, validation and testing. The loss function was Mean Squared Error (MSE) and the training process showed no signs of overfitting in terms of the difference between training and validation loss.

6.2.3. Results

The trained model was evaluated on the data reserved for testing purposes, measuring the Mean Absolute Error (MAE) of the predictions. Mean Absolute Percentage Error (MAPE) could not have been used reliably considering how small the target values are. Distributions for the absolute values of the target $\Delta\theta$ versus the absolute values of its prediction errors are shown individually in Figure 6.13. In terms of absolute errors, nine out of ten muon tracks have absolute prediction errors less than 0.01 rad, as Figure 6.13(b) shows. This cumulative distribution of errors features logarithmic binning and expresses the bin counts as a cumulative fraction of the total number of data points. A closer look at Figure 6.13(a) shows that $\Delta\theta$ values for 50% of tracks are less than the 0.01 rad. The distribution here is again cumulatively counting the fraction of events in log-binning.

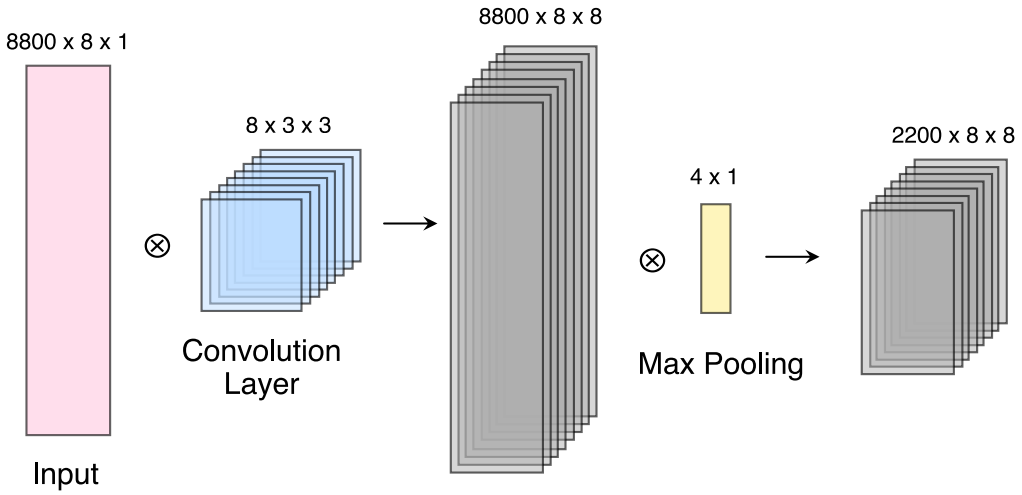


Figure 6.11.: Convolutional block showing the input and output of the convolution operation and max pooling

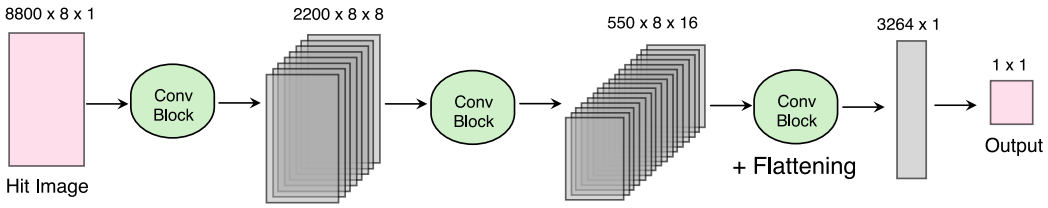


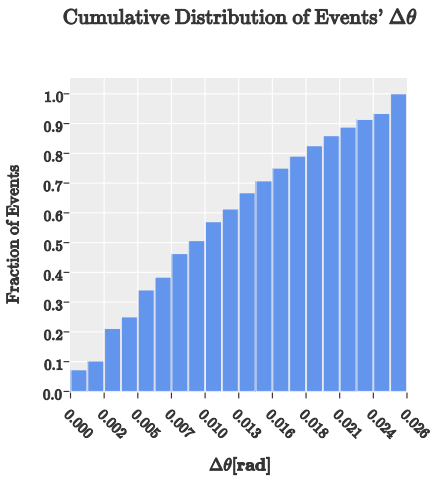
Figure 6.12.: Neural Network architecture deployed in $\Delta\Theta$ regression

A representation of the relationship between those quantities is provided in Figure 6.14, which summarises the regression performance as a two-dimensional histogram of targets versus errors. It is apparent that the regression model's prediction errors take a broad range of values, $|\Delta\theta_{\text{error}}|$ sometimes exceeding the reference value altogether. It is also noteworthy that, the smaller a track's $\Delta\theta$ is, the more error tolerance it has, as well, since $\Delta\theta$'s role is to determine if the track passes the trigger decision and low $\Delta\theta$ tracks are already the most likely to trigger anyway.

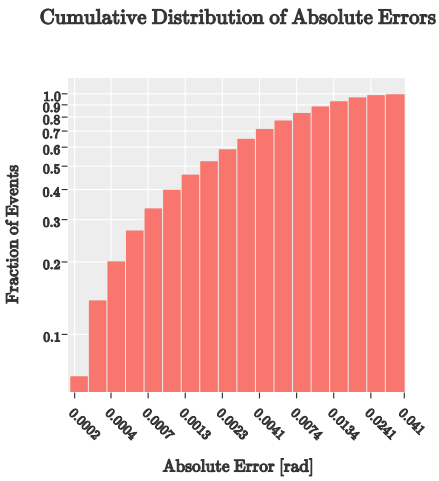
Those results can be better placed in the context of the New Small Wheel (NSW) trigger requirements through the lens of angular resolution. This is computed as the standard deviation of Gaussian fit of the $\Delta\theta_{\text{error}}$ distribution. Both the error distribution and the fit are shown in Figure 6.15. The fitted probability density function is decisively centred at 0, having a standard deviation $\sigma(\Delta\theta_{\text{error}}) = 19$ mrad. For reference, the MicroMegas track angular resolution is 70 mrad [42], but the NSW trigger requirements (which includes sTGC

Type	Shape	Parameters
InputLayer	[(None, 8800, 8, 1)]	0
BatchNormalization	(None, 8800, 8, 1)	4
Conv2D	(None, 8800, 8, 8)	80
MaxPooling2D	(None, 2200, 8, 8)	0
ReLU	(None, 2200, 8, 8)	0
Conv2D	(None, 2200, 8, 16)	1168
MaxPooling2D	(None, 550, 8, 16)	0
ReLU	(None, 550, 8, 16)	0
Conv2D	(None, 548, 6, 32)	4640
ReLU	(None, 548, 6, 32)	0
MaxPooling2D	(None, 34, 3, 32)	0
Flatten	(None, 3264)	0
Dropout	(None, 3264)	0
Dense	(None, 1)	3265
Rescaling	(None, 1)	0

Table 6.1.: Layers and parameter counts of the CNN $\Delta\theta$ regression



(a) Cumulative distribution of test $\Delta\theta$'s absolute value



(b) Cumulative log-binning distribution of test absolute prediction error of $\Delta\theta$

Figure 6.13.: The distribution $\Delta\theta$ values of the test dataset alongside the distribution of $\Delta\theta$'s absolute prediction errors from the model

Figure 6.14.: 2D histogram of absolute value for $\Delta\theta$ versus the absolute values of the regression error. The colour scale and bin labels denote the number of occurrences.

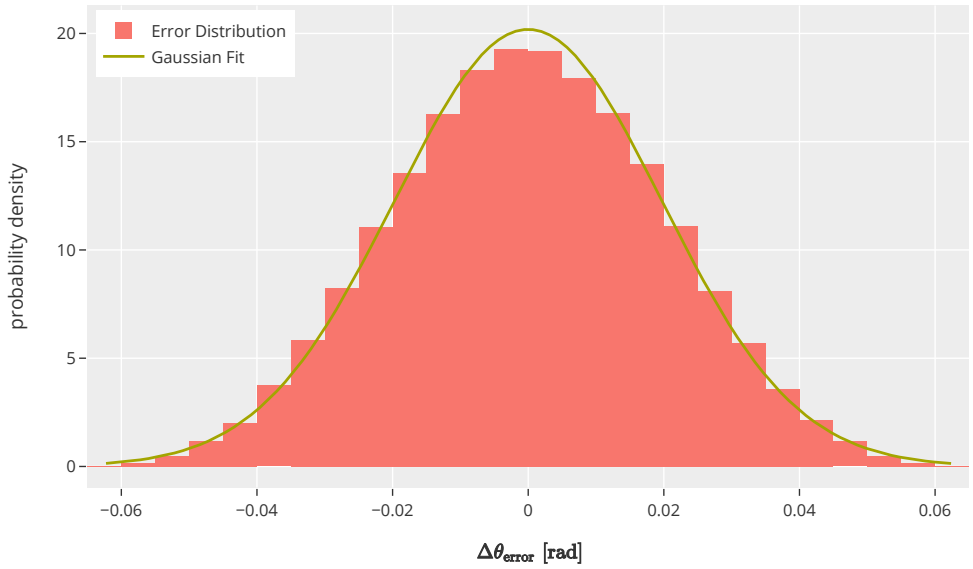


Figure 6.15.: Distribution of $\Delta\theta_{\text{error}}$ alongside a Gaussian fit

as well) target an angular resolution of 1mrad. The NSW small-strip Thin-Gap Chambers hardware is better suited for triggering, having ≤ 1 mrad average angular resolution in segment reconstruction [42]. Thus, the current resolution of $\Delta\theta$ regression may not be close to fulfilling the requirements but is likely to be greatly improved by the inclusion of sTGC information.

As this proof-of-concept attempts to validate the workflow from a technical perspective, the next logical step is deployment on simulated and then physical FPGA hardware. First FPGA migration attempts were successful on *Xilinx Alveo U220* accelerator cards. This process required model quantization and excluding the regularization layers (dropout and batch normalization), but it did not come at a significant cost in regression accuracy. At the time of writing, inference timing studies are still being conducted, with the data pipeline bottlenecks of the Alveo software posing slight technical challenges.

Conclusions and Further Outlook

This chapter presents a feasibility study of a Machine Learning based trigger algorithm that has an adequate size to be implemented on FPGA hardware for an eight-plane MicroMegas detector assembly representative of the New Small Wheel (NSW) layout. With simulated muon tracks and a Convolutional Neural Network approach, I have trained and evaluated a trigger configuration at the software level. The results show that the convolutional method is able to infer the most important trigger variable $\Delta\theta$ up to a resolution of 19 mrad, while not requiring a great number of parameters. The prototype's hardware-level deployment still needs some additional optimization to reduce inference time. On its own, the current iteration of the model is not yet competitive with the existing trigger implementation but still achieves its goal of validating the technical feasibility of the method. The findings are indicative of a relevant potential for the CNN approach to eventually meet the requirements for the NSW trigger processor. This work is going to be continued and extended to the full NSW trapezoidal ATLAS geometry. Future iterations would incorporate information from the small-strip Thin-Gap Chambers (sTGC) detectors, which is expected to improve resolution.

7. BSM Searches using Anomaly Detection

Searches for Beyond Standard Model (BSM) physics phenomena are one of the most active areas of research in the field of High-Energy Particle Physics (HEPP). The established way of performing these kinds of searches has primarily depended on the use of complex targeted analyses that attempt to verify the predictions of certain theoretical models, using real experimental data. With a plethora of models to choose from and progressively larger amounts of data, the range of possible BSM searches grows wider and becomes increasingly difficult to navigate. Although highly successful sometimes, as in the case of the Higgs boson's discovery, this approach requires a lot of highly specialized work that may, in fact, not pay off, unless the underlying model's predictions could be successfully demonstrated or disproved. In light of those aforementioned challenges, some of the particle physics community is turning to alternative ways of searching for BSM signals.

Unsupervised and semi-supervised **anomaly detection** methods, based on Machine Learning (ML), proved highly useful across a wide range of applications. This begs the question if such an approach could be developed for model-independent searches of BSM signals. Considering the vast amounts of data collected by ATLAS throughout its years of operation, purely data-driven anomaly detection could potentially find outliers indicative of new physics.

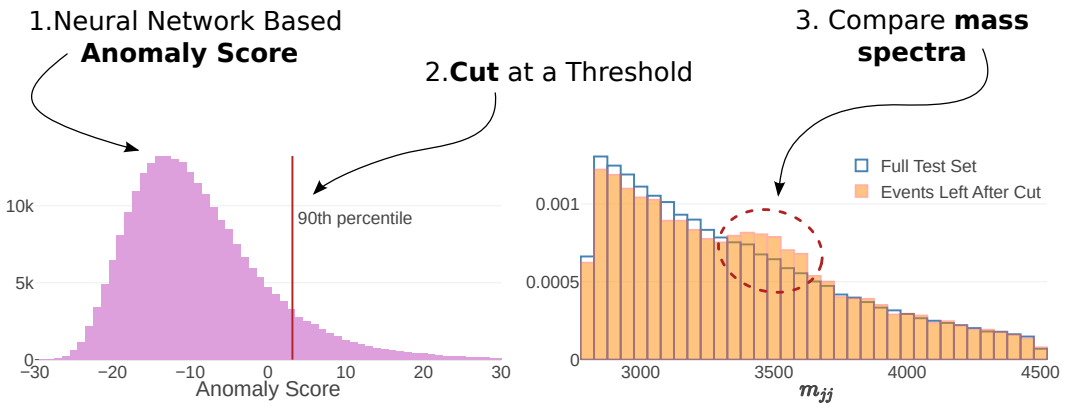


Figure 7.1.: BSM search strategy using NN-based anomaly detection

A major research objective for this thesis was the exploration of unsupervised ML-based methods for anomalous jet tagging. An overview of the BSM search strategy using this type of framework is represented in Figure 7.1, which showcases three main steps:

- A Neural Network (NN) model is trained to generate *anomaly scores* that represent a quantitive assessment of how dissimilar certain events are, relative to the others.
- The application of a user-defined threshold on the distribution of anomaly scores results in a subsample containing the most *anomalous* data points.
- The properties of the original data set and the anomalous subsampled are compared. In the BSM search scenario, the physics object's mass distributions when put side-by-side may uncover a resonance related to a new particle.

7.1. Unsupervised Learning

The supervised approach to Machine Learning makes use of known output targets for the training data, but such information is not always available. ML models attempt to identify data **patterns** that are closely related to the prediction targets, but those inherent patterns could exist even in the absence of a well-defined training label. Unsupervised learning refers to this exact scenario where ML is deployed to look for underlying data structure and relationships between observations, without having access to any dependent variables. As well as providing a general introduction to the subject, this section will describe in more detail the specific methods, relying on information from relevant bibliographical resources [18, 106].

Unsupervised learning has two major areas of applications, one being **outlier detectioin** [125] and the other closely related one being *clustering*. D. Hawkins [126] defines an outlier as *an observation that deviates so significantly from other observations as to arouse suspicion that it was generated by a different mechanism*. They may be the result of errors and faults in the data and/or model, but there are also cases when the outliers are caused by an unknown underlying process.

Within the domain of anomaly detection methods, one has several options to choose from. The two main categories of models could be thought of as reconstruction-based and density-based. Reconstruction refers to the ML model's objective, meaning the reconstruction of the entire event using a limited subset of the available information. The most notable NN architecture design for reconstruction-based anomaly detection is the Autoencoder [106, 127], further discussed in Subsection 7.1.1. The other category of models involves density estimation, the simplest example being Kernel Density Estimation (KDE) [128, 129].

7.1.1. Autoencoders

An Autoencoder (AE) is a deep Neural Network that has input and output layers with the same number of dimensions, with an intermediate *informational bottleneck* in between. This bottleneck is materialized as a smaller-dimensional layer, holding a *compressed* representation of the input data. The output of this layer often called the *latent representation* is used by subsequent parts of the network in order to reconstruct the original input. Figure 7.2 shows a visual diagram of an AE network. The training target is, thus, the original input, meaning that no additional data labels are necessary [106].

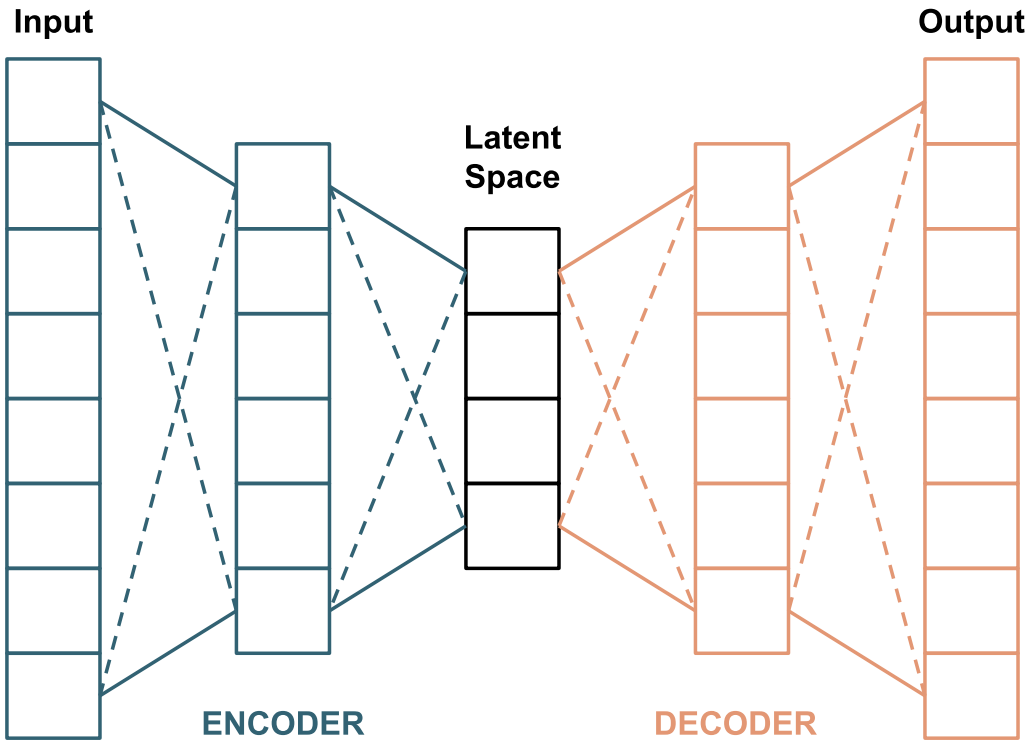


Figure 7.2.: Schematic representation of the Autoencoder architecture

Autoencoder are useful for *representation learning* [130]. They may be conceived as having two parts due to their symmetrical structure. The **encoder** takes the input and maps it to the lower-dimensional *latent space* and the **decoder** uses the *latent representation* of the data and attempts to map it back to the original space. Constraining the latent space's dimensionality forces the model to make assumptions about the data in order to efficiently compress it. It may remove redundancies and correlations between features, similarly to Principal Component Analysis (PCA) [131], but with the added benefit of potentially modelling non-linear correlations as well.

The size of the *latent space* is an important hyperparameter in this type of model. For datasets with highly correlated features, having the *latent* representation half as large compared to input space may show minimal reconstruction errors, but in other cases, even a 10% reduction of the original representation's size may make the model unable to reconstruct the inputs. Unfortunately, there is no established heuristic for determining *a priori* the optimal size of the latent space, which means several configurations need to be tested before arriving at a desirable result. Some metrics for evaluating the AE model's performance may be the average reconstruction error, measured as the value of the **loss** function, as well as qualitative assessments on how *expressive* are the *latent space* variables' distributions. As an extreme example, if the size of the latent space is too small or the data is too noisy, the *latent* variables may *collapse* to constant values which are independent of the input. This lack of *expressivity* would indicate an Autoencoder that was unable to identify any meaningful latent representation and minimized the loss function just by virtue of always outputting the average of the training dataset's features.

Apart from data compression, the Autoencoder may also be applied to outlier detection. The reconstruction error, which is often also the loss function, can effectively be used as an anomaly score. The Autoencoder learns to minimize this loss for the training dataset, which requires extracting the most suitable information as the *latent representation*. This means that the model is making assumptions on feature relevance based on the data it *sees* during training. For out-of-distribution data points, those assumptions may not apply anymore, resulting in erroneous reconstruction. Thus, reconstruction error can be interpreted as a measure of similarity between the observation under evaluation and the data used in training the model.

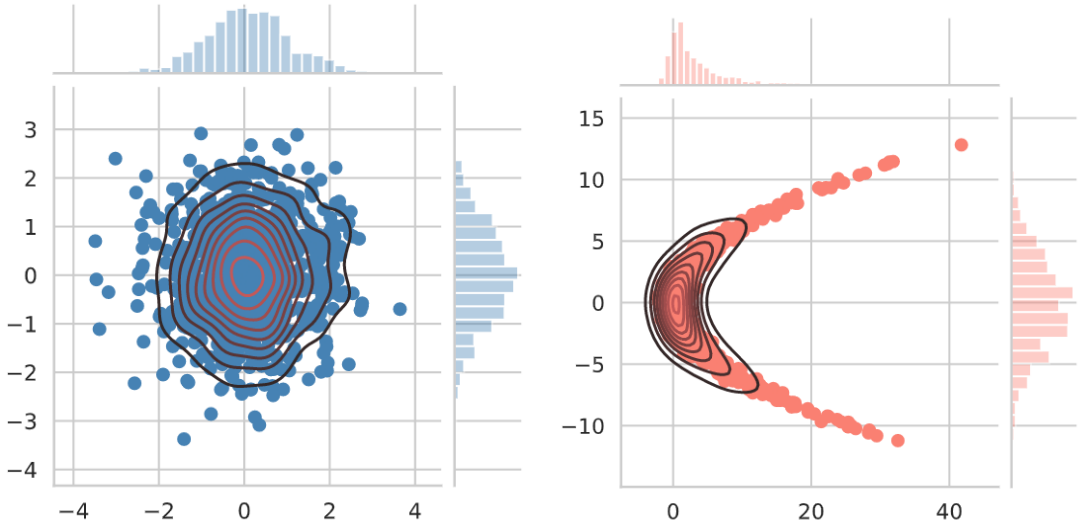
Anomaly detection using Autoencoders may take different forms. If the data available for training already contains unlabeled anomalies, one may train and evaluate the AE on the same dataset, taking a fully **unsupervised** approach. However, some applications may benefit from having anomaly-free data samples and use those for training before evaluating the model on potentially anomaly-ridden datasets, this case falling under **semi-supervised learning**. Use cases such as those require sufficiently large training data samples that reasonably cover the entire space of what would be considered "in distribution" observations.

7.1.2. Normalizing Flows

Non-parametric density estimation is a wide-reaching statistical problem with a long history of developments [132]. Numerous more traditional methods can still seamlessly solve this task today, such as Kernel Density Estimation (KDE) [128, 129] and k-Nearest Neighbors (KNN) [110, 133]. However, as the data complexity increases to higher numbers of dimensions, a lot of such methods can become very computationally demanding and/or start to lose accuracy in their estimations. In an attempt to find better density estimation methods that are also scalable to complex multidimensional datasets, ML-based approaches started

to gain traction as possible solutions. One such model is called a Normalizing Flow (NF), which can approximate an **explicit density function** for multivariate data. This is achieved by *learning* a deterministic mapping between a know *source* $p(\mathbf{z})$ distribution such as a multivariate Gaussian and the *target* $q(\mathbf{x})$ distributions.

The Normalizing Flow mapping can be formalized as a *deterministic bijective transformation* $\mathbf{T}$ between $q(\mathbf{x})$ and $p(\mathbf{z})$. An illustrative example showcasing toy 2D distributions is shown in Figure 7.3. The source distribution is the 2D Gaussian in Figure 7.3(a), and the target distribution is shown in Figure 7.3(b).



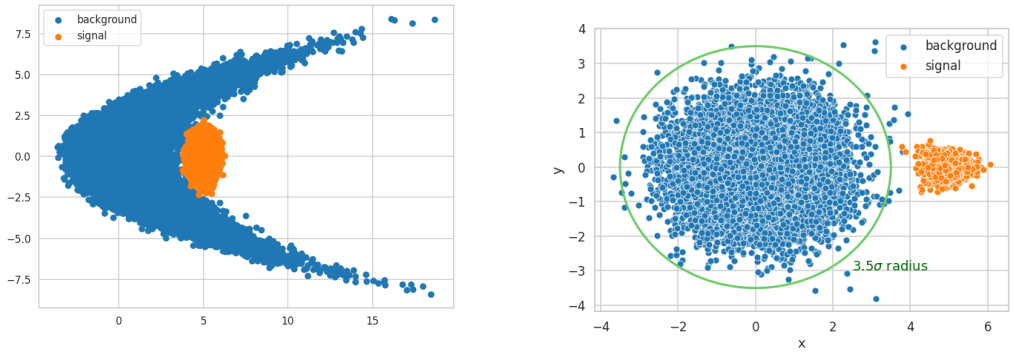
(a) 2D Gaussian

(b) Toy target data

Figure 7.3.: Toy example data with two-dimensional source and target distributions generated with the SCIKIT-LEARN library [134]

Knowing the transformation $\mathbf{T}$ and using the **bijectivity** property allows transforming the data back and forth between the original space and the Gaussian-distributed space. Coming back to the toy example, Figure 7.4(a) introduces out-of-distribution points labelled as *signal*. Although they are not well separated from the *background* points in terms of spatial distance, once everything is transformed to the *source* space, the separation becomes apparent. Figure 7.4(b) shows the *background* and *signal* points after the inverse transformation $\mathbf{T}^{-1}$ was applied.

It is easily apparent how this type of bijective transformation can be useful in anomaly detection, but so far nothing has been said about how it is defined or learned. When defining this bijective mapping the first important aspect is enforcing the **conservation of probability mass**. In short, applying transformations to a distribution should not affect the total probability. Mathematically, this requirement manifests as an additional factor, namely



(a) Original data

(b) Reverse transformation of the data

Figure 7.4.: Example outlier detection application on 2D toy data. Subfigure 7.4(a) shows data points from the target *background* distributions alongside *signal* out-of-distribution data points in the original space, while subfigure 7.4(b) shows the same datapoints after applying the reverse normalizing flow transformation.

the determinant of the transformation's Jacobian, applied to the transformed distribution, as in Equation 7.1.

$$q(\mathbf{x}) = p(\mathbf{z}) \left|\det \mathcal{J}(\mathbf{T}(\mathbf{z}))\right|^{-1} \tag{7.1}$$

From a computational standpoint, calculating the determinant for an $n \times n$ arbitrary matrix is expensive [135], having $\mathcal{O}(n^3)$ complexity in Bachmann-Landau notation [136, 137]. Thus, to ensure the scalability of the method, the bijective transformation needs to be chosen in such a way that the determinant of its Jacobian is significantly faster to compute. Triangular maps are a great solution since the determinant becomes just the product of the diagonal elements and for the vast majority of distributions, there will always exist an **increasing triangular map** facilitating a diffeomorphism between the *source* $p(\mathbf{z})$ and *target* $q(\mathbf{x})$ distributions [138].

While requiring the mapping to be triangular has computational benefits, this translates to using an autoregressive transformation, where the first variable of the target multivariate distribution depends only on the first variable of the source multivariate distribution, while the last variable of the target would depend on an all of the source's variables. This kind of relationship between the source and target distribution's variables is illustrated in Figure 7.5.

In practice, the order of a data point's features is arbitrary, and, in order to prevent any ordering biases, several triangular transformations are stacked together. The order of the variables is permuted between every two mappings, ensuring that the output receives contributions from every one of the input's features. Thus, the Normalizing Flow can be

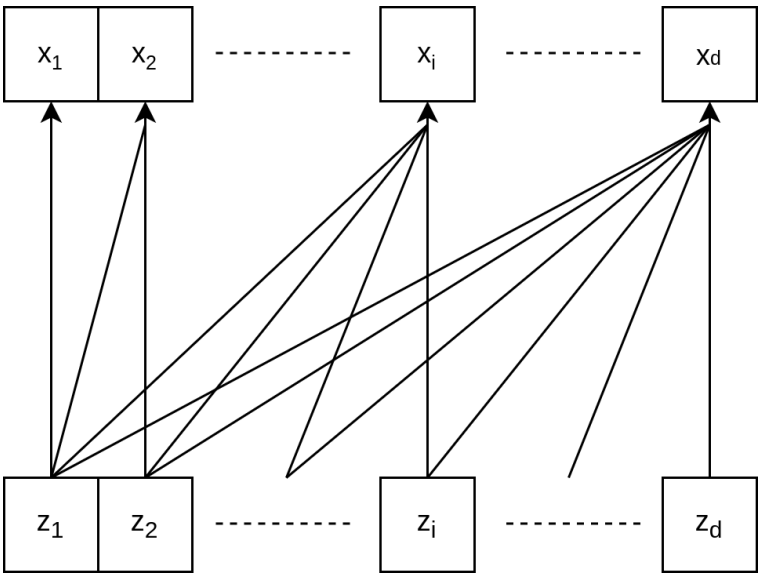


Figure 7.5.: Schematic representation between the input $\mathbf{z}$ and output $\mathbf{x}$ of an autoregressive transformation $\mathbf{T}$, where $\mathbf{x} = \mathbf{T}(\mathbf{z})$.

thought of as a chain of triangular bijective maps between the *source* and *target* distributions, as Equation 7.2 shows.

$$\mathbf{z} \xrightarrow{\mathbf{T}^{(1)}} \mathbf{z_1} \xrightarrow{\mathbf{T}^{(2)}} \mathbf{z_2} \dots \xrightarrow{\mathbf{T}^{(k)}} \mathbf{x} \tag{7.2}$$

With all intermediate bijective transformations being triangular, the conservation of probability mass requirements does not gain any significant complexity. Determinants for all of those maps are needed, as shown in Equation 7.3, but they remain easy to compute due to the triangularity of each individual map.

$$q(\mathbf{x}) = p(\mathbf{z}) \left|\nabla \mathbf{T}^{(1)}\right|^{-1} \left|\nabla \mathbf{T}^{(2)}\right|^{-1} \dots \left|\nabla \mathbf{T}^{(k)}\right|^{-1} \tag{7.3}$$

Before talking about *learning*, one must first define the model's parameters. In the case of the Normalizing Flow, the simplest way of parametrising the transformations would be through the use of **affine functions** [18]. Equation 7.4 exemplifies a simple scalar transformation consisting of a scaling and a translation. In this example, α_i and β_i are the parameters to optimize for a component T_i of one of the triangular maps. However, the autoregressive nature of the transformations implies that α_i and β_i may depend on the previous component's T_{i-1} parameters. A more detailed overview of how autoregressive affine transformations are used by NF models can be found in Appendix E.

$$x_i = T_i(z_i) := \alpha_i z_i + \beta_i \tag{7.4}$$

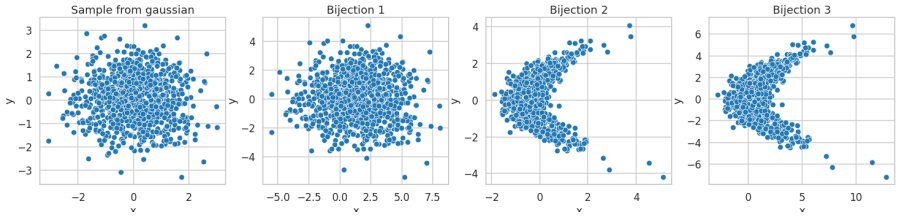


Figure 7.6.: Step-by-step transformation of the toy dataset

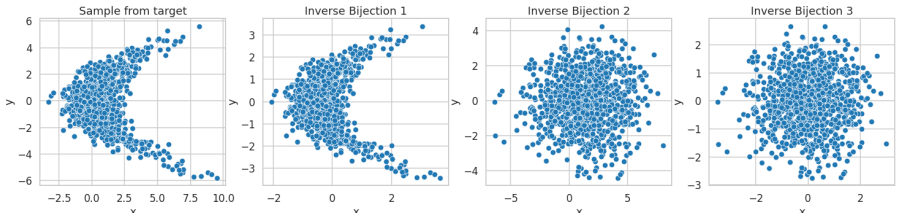


Figure 7.7.: Step-by-step inverse transformation of the toy dataset

Normalizing Flow models using affine transformations are the most widely used [18], with notable literature examples being Non-linear Independent Components Estimation (NICE) [139], Inverse Autoregressive Flow (IAF) [140] and Masked Autoregressive Flow (MAF) [141].

Training a Normalizing Flow is fully **unsupervised**, requiring just a sample from the target distribution. Since the NF model can use its trainable parameters to compute the likelihood of the input data, a great training objective would be to maximize this likelihood over the training dataset, as shown in Equation 7.5. For numerical stability reasons, the logarithm of the likelihood would be preferable instead, thus establishing the Negative log-likelihood (NLL) as a good loss function choice.

$$\hat{\mathbf{T}} = \underset{\mathbf{T}}{\operatorname{argmax}} \prod_{i}^{n} p(\mathbf{z_i}) \left| \det \mathcal{J}(\mathbf{T}(\mathbf{z_i})) \right|^{-1} \tag{7.5}$$

Circling back to the previous toy data displayed in Figure 7.3, the Normalizing Flow can be shown in action. Training a Masked Autoregressive Flow (MAF) containing four stacked transformations on this toy data produces the results displayed in Figure 7.4. Since the affine transformations only allow simple translations and rescaling, several of them are required even in this 2D example.

Diving deeper into how the model works, Figure 7.6 shows the output of every MAF *layer*, while the inverse transformation is shown step-by-step in Figure 7.7. The first layer of the bijective mapping appears to just perform a rescaling of the distribution while the second layer changes its shape before the third one rescales once again.

7.1.3. Probabilistic Autoencoder

As discussed in the two previous subsections, reconstruction-based methods such as the Autoencoder and density-based methods such as the Normalizing Flow, both have clear uses in anomaly detection tasks. In the context of BSM search use, there is research suggesting that a combination of both classes of methods may outperform using either of them individually [16].

There could be many ways of combining an Autoencoder (AE) with density estimation, but one particularly interesting ensemble that achieves this is the Probabilistic Autoencoder (PAE) [17]. The PAE consists of an AE model whose latent space is mapped to normally-distributed space by a Normalizing Flow model. Figure 7.8 shows a diagram of the PAE architecture.

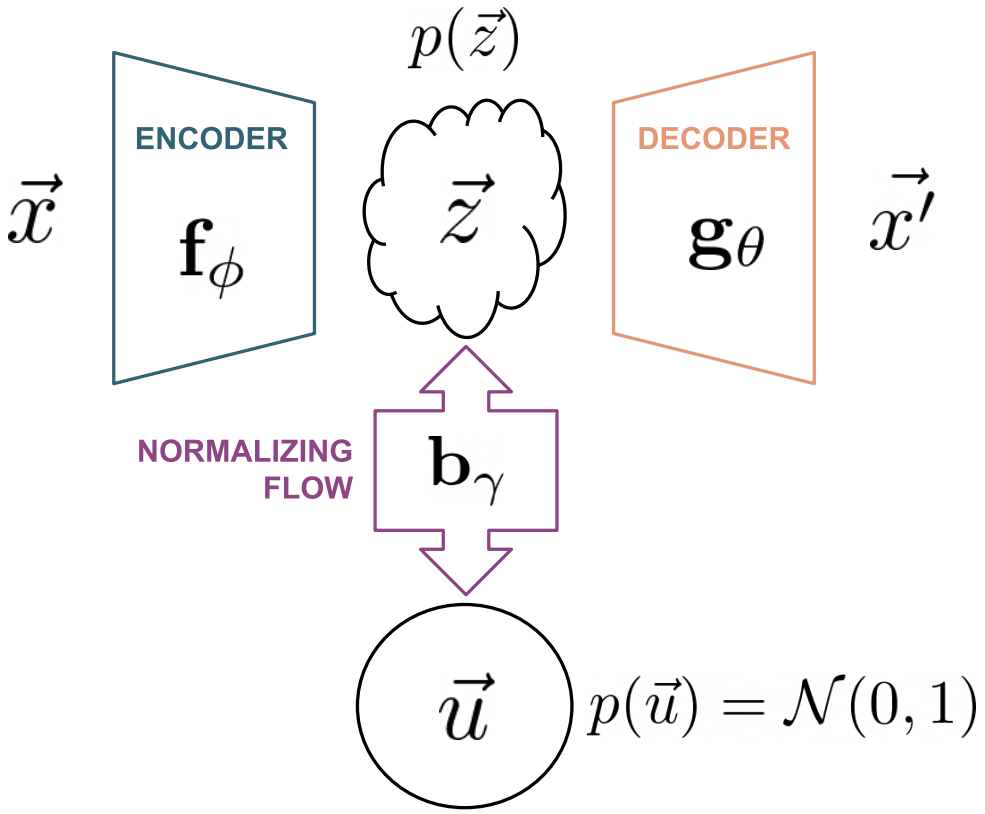


Figure 7.8.: Schematic of the Probabilistic Autoencoder Neural Network ensemble

Within the Probabilistic Autoencoder model, the Autoencoder and Normalizing Flow can work together in an attempt to provide a likelihood estimate of the inputs $p(\vec{x})$. The NF b_γ is already able to provide a tractable density function for the latent representation $p_\gamma(\vec{z})$, but, unfortunately, since the property of injectivity is not enforceable for the encoder

$\mathbf{f}_\phi$, the input and encoding densities are not the same: $p(\vec{x}) \neq p_\gamma(\vec{z})$. To arrive at an approximation of $p(\vec{x})$, one starts by writing the joint probability $p(\vec{x}, \vec{z})$ as a product of conditional probabilities according to Equation 7.6 [17].

$$p(\vec{x}, \vec{z}) = p_\theta(\vec{x}|\vec{z}) p_\gamma(\vec{z}) \tag{7.6}$$

Further next, $p(\vec{x})$ can be isolated by marginalizing over z, as in Equation 7.7. This integral however is not directly tractable and will require some approximations.

$$p(\vec{x}) = \int p_\theta(\vec{x}|\vec{z}) p_\gamma(\vec{z}) d\vec{z} \tag{7.7}$$

The implicit likelihood $p_\theta(\vec{x}|\vec{z})$ is assumed to have a Gaussian form, as Reference [17] dictates, while the covariance matrix σ is diagonal and its non-zero elements are represented by the average per-feature reconstruction errors. Denoting by σ_i the diagonal elements of σ and the decoder by $\mathbf{g}_\theta$, the Gaussian prior can be written as shown in Equation 7.8, where n represents the number of features for $\vec{x}$.

$$p_\theta(\vec{x}|\vec{z}) = (2\pi \prod_{i=1}^{n} \sigma_i^2)^{-\frac{n}{2}} \exp\Big(-\frac{1}{2} \sum_{i=1}^{n} \frac{[x_i - g_{\theta,i}(\vec{z})]^2}{\sigma_i^2}\Big) \tag{7.8}$$

With this likelihood defined, the integral in Equation 7.7 can be computed using the Laplace approximation [19]. According to the authors of the original Probabilistic Autoencoder paper [17], the Hessian terms introduced by the Laplace method are reducing the stability of the likelihood estimate so they will be dropped from this point forward.

Accounting for all this, Equation 7.9 shows the resulting formula for the $p(\vec{x})$ approximation.

$$\ln p(\vec{x}) \approx -\frac{1}{2} ||\vec{x} - \vec{x}'||^2 \vec{\sigma}^{\circ -2} - \frac{1}{2} b_\gamma(\vec{z})^2 + \ln |\det \mathcal{J}_\gamma| \tag{7.9}$$

Unpacking this score, the first term can be recognized as the Mean Squared Error (MSE) loss of the Autoencoder and $\vec{\sigma}$ is the diagonal of the implicit likelihood's covariance matrix. The other term of the anomaly score is the likelihood of the latent space representation, provided by the Normalizing Flow model.

Equation 7.9 contains an additional shortcut compared to the related publication [17], where instead of $\vec{x}'$, which is the AE reconstruction of the input, $\mathbf{g}_\theta(\vec{z}')$ is used. Finding $\vec{z}'$ implies optimizing $\vec{z}$ in order to minimize the negative log-likelihood of the posterior distribution, as defined in Equation 7.10.

$$\text{NLL} = \underset{\vec{z}}{\text{argmin}}(-\log p_\theta(\vec{x}|\vec{z}) - \log p_\gamma(\vec{z})) \tag{7.10}$$

This additional optimization step may be bypassed by assuming $\vec{z}' = \mathbf{f}_\phi(\vec{x})$ resulting in an easier to compute $p(\vec{x})$ estimation. The effect of this assumption on the input likelihood's performance as an anomaly score is going to be discussed in further sections.

Training a Probabilistic Autoencoder ensemble is done sequentially, starting with the Autoencoder being trained normally. The fitted AE is used to infer the latent representation of all training data points. Those latent representations are then used as training data for the Normalizing Flow. Additionally, the average per-feature AE reconstruction error can be evaluated, preferably on the validation dataset, in order to determine the values of $\vec{\sigma}$, which are used for the likelihood estimation of the test data.

For anomaly detection purposes, a trained Probabilistic Autoencoder offers several potential metrics for discrimination. The likelihood of the latent representation as well as the Autoencoder reconstruction error may be considered, or even something like $||\vec{x} - \mathbf{g}_\theta(\vec{z}')||^2$, which uses the NF's density indirectly through the negative log-likelihood optimization necessary for finding $\vec{z}'$. Another obvious choice would be the likelihood estimate of the input $\ln p(\vec{x})$. Given all this flexibility in anomaly scores, the Probabilistic Autoencoder can effectively use both density and reconstruction error information at the same time, which makes it a good worthwhile candidate for potential unsupervised or semi-supervised BSM searches.

7.2. LHC Olympics Challange

Following the increase in interest for anomaly detection Machine Learning as a potential model-independent way of searching for Beyond Standard Model (BSM) signals, the High-Energy Particle Physics community started to create benchmark datasets for evaluating the performance of such approaches. Challenges such as the *Dark Machines Anomaly Score Challange* [142] and *LHC Olympics* [20] provided open-access data samples containing fractions of BSM signals among large amounts of events from simulations of well-understood Standard Model (SM) processes.

The *LHC Olympics* challenge launch coincided with the start of my thesis, and, considering my existing research focus on anomaly detection, it was a great opportunity for putting my work to the test. Thus, the *LHC Olympics* datasets are the main benchmarks used in this work.

Several datasets were provided, some of them labelled and some unlabeled, each of them designed to test a different anomaly detection scenario. All *LHC Olympics* data revolved around jet physics, with every event containing at least two jets in the final state. The data

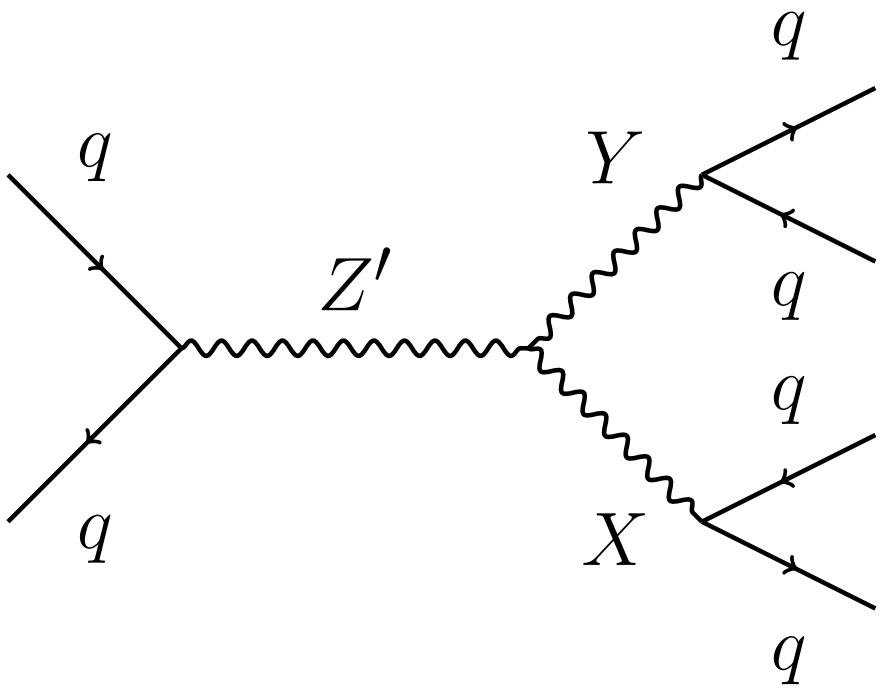


Figure 7.9.: Feynman diagram for the process generating the R&D and *Black Box 1* signals [20].

format consists of four-vector particle flow information of all particles produced in every event. There is no information provided on particle types or their mass. Data samples were generated using PYTHIA [143], HERWIG++ [144] and DELPHES [145] simulations.

7.2.1. Benchmark Datasets

Before the challenge started, the R&D dataset [146] was published, containing one million QCD events labelled as *background* and two smaller samples of 10^5 *signal* events obtained from simulations of BSM processes with two and three final state jets respectively. This data was meant for developing and validating anomaly detection methods, hence the use of *background* and *signal* labels.

The challenging part of the *LHC Olympics* was applying the developed method to find the BSM signal in the *Black Box* datasets. Those used the same data format but were initially unlabeled, forcing any analysis to search for the signal in model-independent ways. As the *LHC Olympics* challenge started, the following datasets were published [147]:

- Background-only training set (10^6 QCD events)
- Three different *Black Boxes* potentially containing BSM signals (10^6 total events each)
  - **Black Box 1**: 3.823 TeV Z' decaying into two jets (834 signal events)

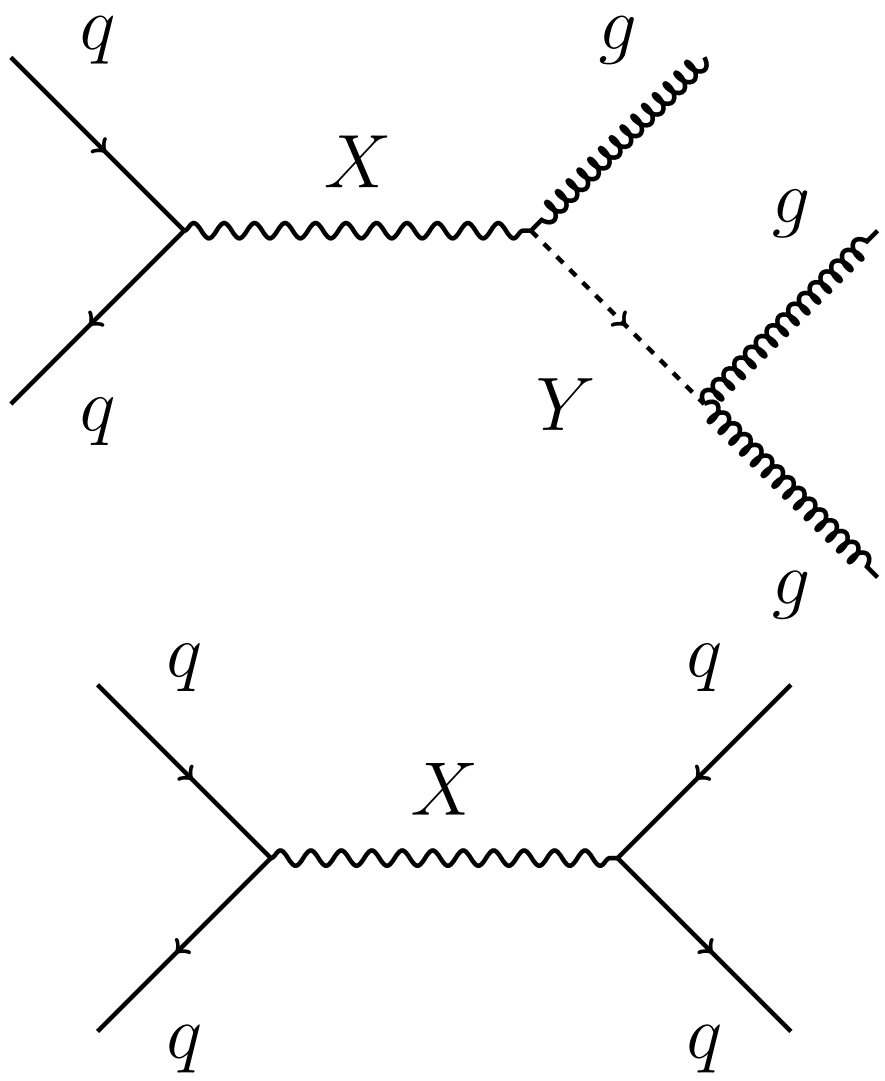


Figure 7.10.: Feynman diagram for the process generating the *Black Box 3* signal [20].

- **Black Box 2**: QCD background only
- **Black Box 3**: 4.2 TeV Kaluza-Klein gravitron [148] decaying in dijet (1200 events) and trijet (2000 events)

The **Black Box 1** was released first containing the same type of signal as the R&D dataset, but in a very small signal-to-background ratio: $S/B \approx 0.084\%$. Figure 7.9 shows the Feynman diagram for the signal-generating process. The main difference between the R&D and **Black Box 1** data is the masses of the Z' and its decay products which are $m_{Z'} = 3.5\,\text{TeV}$, $m_X = 500\,\text{GeV}$, $m_Y = 100\,\text{GeV}$ for the former and $m_{Z'} = 3.823\,\text{TeV}$, $m_X = 732\,\text{GeV}$, $m_Y = 378\,\text{GeV}$ for the latter.

Black Box 2 did not contain any signal making it a great test for model bias. While an anomaly detection method may be very good at identifying outliers in data, it is just as important to be able to tell when there aren't any. Thus, looking for signals in **Black Box 2** is a good sanity check for potential biases of the model.

Lastly, the **Black Box 3** dataset contained a BSM signal from a Kaluza-Klein graviton [148]. While the $S/B \approx 0.321\%$ is larger than what is found in **Black Box 1**, the 4.2 TeV signal particle has two decay modes, shown in Figure 7.10, one with a dijet final state and the other one with three jets. To make things even more difficult the authors specifically tuned the branching ratio in such a way finding only one of those decay modes would not be enough to identify a significant excess.

The *LHC Olympics* challenge datasets also contained a *background*-only sample, which enables the use of semi-supervised methods as well fully unsupervised ones. In practice, one may have access to pure background samples via MC simulations, but there may always be small miss-matches between real data and Monte Carlo. In order to account for this common limitation, the background data is modelled slightly differently across all datasets. Thus, a semi-supervised model trained on the pure background sample needs to be properly adapted to those small differences.

7.2.2. Data Preparation

Datasets published for the *LHC Olympics* contain only low-level event information in the form of lists of particle four-vectors. Using this data directly would be infeasible since the analysis strategy uses properties of higher-level physics objects such as jet invariant mass. Thus, the first step of data preparation should be jet clustering. Towards this end, the authors of the challenge suggest using the anti-k_t clustering algorithm [149] to select $R = 1$ radius jets with a threshold of $p_T \geq 1.2$ TeV. In terms of the software data processing was done in PYTHON [22] with the PYJET library [71] deployed for clustering.

Once the jets are clustered, the next important step concerns feature selection. With PYJET allowing PYTHON access to FASTJET [70] functionalities, a lot of jet features can be extracted such as kinematics and invariant mass. On top of those, jet substructure was expressed in terms of **subjettinesses** τ_3, τ_2 and τ_1 definitions for those observables were introduced in Subsection 3.4.4. Another substructure-related set of features comprised **energy rings** which reflect the constituent's energy distribution in terms of radius. Equation 7.11 shows the formal definition for the energy ring i out of n total equal-thickness rings.

$$E_i = \frac{1}{E_{\text{jet}}} \sum_k E_k, \text{ where } \Delta R_{\text{jet},k} \in \left[\frac{i-1}{n} R_{\text{jet}}; \frac{i}{n} R_{\text{jet}} \right] \text{ and } i = \overline{1, n} \tag{7.11}$$

The full list of jet features goes as follows:

- kinematics: (p_T, η, ϕ, m)
- total energy: E_{jet}
- subjettinesses: τ_1, τ_2, τ_3, τ_{32}, τ_{21}
- inclusive and exclusive subjet multiplicity[1]
- constituent particle multiplicity[2]
- ten energy rings with logarithmical thickness binning

In total there are 22 features per- jet and two event-level features: dijet combined mass m_{jj} and multiplicity of inclusive jets resulting from the initial clustering. Alternatively, the events were also represented as a series of images of the clustered jets [86], following the methodology introduced in Subsecrion 3.4.4. The full code used for clustering and pre-processing of *LHC Olympics* data is publicly available [150].

7.3. Anomaly Detection Results

While several approaches were experimented with, the Probabilistic Autoencoder (PAE) ended up being the main focus, considering the large number of potential anomaly scores that can be derived from it. Using the general strategy introduced at the beginning of this chapter, a full anomaly detection workflow was developed to search for the BSM signals included in the *LHC Olympics* datasets. Since the strategy revolves around searching for resonances on the m_{jj} spectrum after anomaly score-based event selections, it is especially important to ensure that the anomaly score is not correlated with the dijet invariant mass. The methods deployed for mitigating such biases are described in Subsection 7.3.1.

The labelled R&D dataset allows in-depth studies of anomaly score performance. Susection 7.3.2 presents a comparative review of PAE-derived anomaly scores, evaluated by their discrimination power as well as the amount of bias they introduce.

Once the most promising anomaly score is identified, the next step involves deploying the full analysis method on the *Black Boxes*. By applying a cut on the anomaly score's distribution to select a subsample of *most anomalous* events, this subsample is then compared to the full *Black Box* data. Looking at the two normalized m_{jj} distributions side-by-side, one expects to find an excess corresponding to the BSM signal in the anomaly-enriched subsample. This excess is sought after with the **Bump Hunter** algorithm [21], which apart from facilitating the resonance search, is also able to compute the significance of the findings. To seamlessly integrate *bump hunting* into the PYTHON [22] and JUPYTER Notebook [23] analysis codebase, the PYBUMPHUNTER library was used for the convenience and performance that it provides.

[1] The number of inclusive and exclusive ($d_{cut} = 0.2$) subjets as defined in Subsection 3.4.2
[2] The number of constituent particles

The results obtained by applying this analysis workflow to the *Black Boxes* datasets are presented and discussed in Subsection 7.3.3. Finally, additional tweaks and optimizations are discussed in Subsection 7.3.4.

7.3.1. Bias Mitigation Strategies

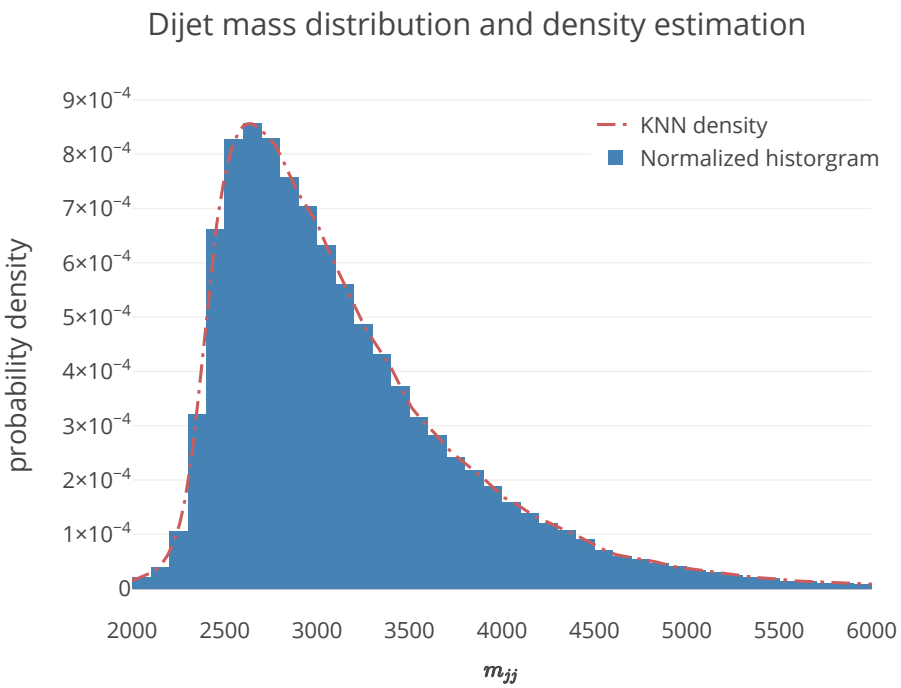


Figure 7.11.: Dijet mass m_{jj} distribution of the R&D dataset's background QCD events (blue) and its KNN density estimate (red)

Among all *LHC Olympics* datasets the events' m_{jj} distribution has a similar shape, peaking around 2.7 TeV and having a long tail extending up to 9.5 TeV. While such distribution makes physical sense, as m_{jj} increases there are fewer and fewer data points left. Thus, a naive density-based anomaly score may flag events at the upper and lower ends of this spectrum as outliers. A reconstruction-based anomaly score may easily acquire a similar bias since it doesn't *see* enough of such examples in order to prioritize learning how to reconstruct them. This means that from the start, m_{jj} should not be used as a feature in training.

In addition to this, the training data can be reweighted in such a way that less common events in terms of m_{jj} are awarded higher priorities in order to bring them in line with the rest of the training dataset. The simplest means to this end would be to use a non-parametric

density estimation algorithm, such as *k*-Nearest Neighbors (KNN) [133] and fit it to the training data's m_{jj} distribution, as exemplified in Figure 7.11.

By inverting the values of the m_{jj} density estimate for each data point a rescaling factor is obtained which is further going to be referred to as the event *weight*. Multiplying the value of the loss function by this weight has the same m_{jj} uniformisation effect as selective oversampling of the training data. A scatter plot of the weights obtained from the previously shown KNN density is shown in Figure 7.12.

More often than not, even with such measures in place, the model may still learn to infer m_{jj} from jet features such as p_{T}, E and η. Initial PAE prototypes suffered from this exact issue. The best solution found was to apply uniformisation to all training features as well. With the `QuantileTransformer` utility from the SCIKIT-LEARN library [134], a dedicated transformation is applied to every quantile of each feature's distribution such that the result is a training dataset having the values of each feature be uniformly distributed without breaking the relationships between features.

All of the aforementioned data manipulations were applied to the dataset used for training the PAE model in order to minimize the biases of the results as much as possible. None of those mitigation strategies seemed to negatively affect anomaly detection performance.

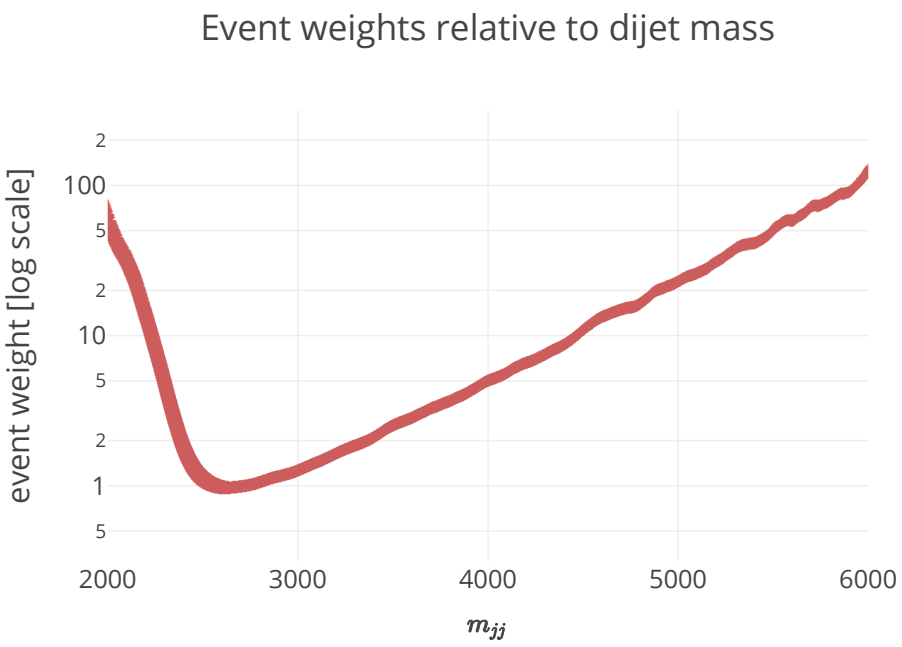


Figure 7.12.: Scatter plot of R&D QCD events' weight as a function of their dijet mass

7.3.2. Anomaly Score Performance on the R&D Dataset

Several anomaly scores were evaluated on the R&D both in terms of their efficiency of signal extraction as well as how much bias they are introducing. The PAE-derived chosen for this comparison were:

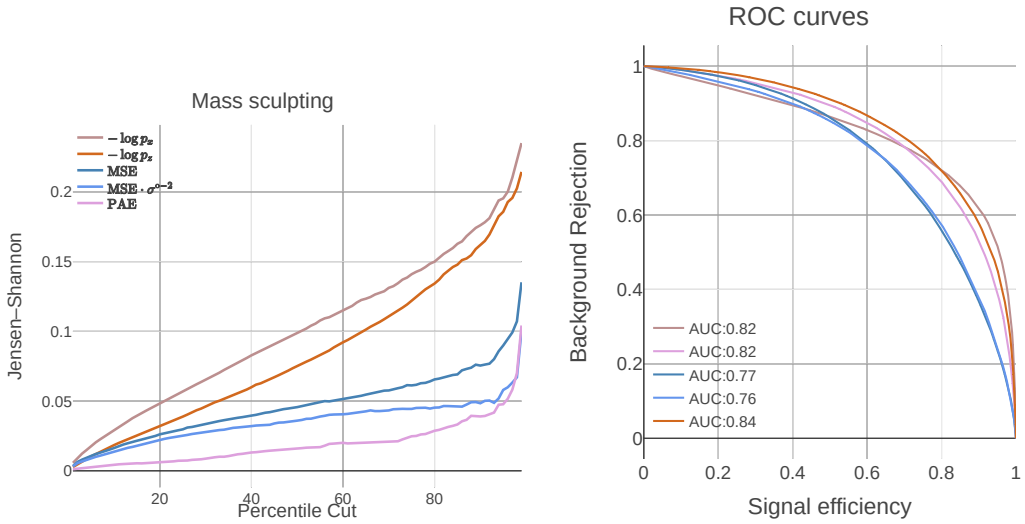
- $-\log p_z$: the negative log-likelihood of the latent representation
- MSE: Mean Squared Error between the input and the PAE reconstruction
- MSE $\cdot\, \sigma^{\circ -2}$: PAE reconstruction MSE divided by the average per-feature reconstruction error evaluated on the validation dataset
- PAE: the PAE estimate of the input data's likelihood (Equation 7.9)

For reference, an additional anomaly score was tested, namely, the likelihood of the input $-\log p_x$ according to a separate Normalizing Flow (NF) model, trained on the input features as opposed to the latent space.

A single PAE model was trained on a synthetic dataset containing only QCD background from the R&D dataset, while validation data consisted of a smaller but still pure background sample. For testing, a separate dataset was prepared to contain equal amounts of QCD background and dijet signal events.

Quantitative evaluation of the anomaly score's bias relative to the m_{jj} involved the Jensen-Shannon (JS) divergence [24], which is defined in Appendix G. After training, the anomaly scores were first inferred on the training set which was designed to be signal-free. Then, successive thresholds were applied covering all percentiles of the anomaly score distribution. For each of those thresholds, the JS-divergence between the m_{jj} spectra of the full training sample and the above-threshold subsample was computed. The goal here is to identify how much is the m_{jj} distribution *sculpted* by the anomaly score cut. Since JS-divergence measures the *distance* between those two distributions, an unbiased score would show low values for the JS-divergence in this scenario, regardless of the threshold, because there is no resonance to be found in this scenario.

Mass sculpting plots of JS-divergence vs anomaly score percentile threshold are shown in Figure 7.13(a). A common behaviour regardless of the score in question is the abrupt increase in JS-divergence at the very last few percentiles. This can most likely be attributed to statistical fluctuations becoming more and more pronounced as fewer and fewer events pass the anomaly score threshold. Making abstraction of this, it is clear that the negative log-likelihoods of both NF models are introducing significant amounts of m_{jj} bias. On the other hand, the reconstruction-based anomaly scores are adding decisively less mass sculpting. Interestingly, the PAE score seems to be overall the least biased of the bunch, less than any of the two scores that it incorporates.



(a) Mass sculpting

(b) Receiver Operating Characteristics (ROC) curves

Figure 7.13.: Performance metrics for the potential anomaly scores accessible through the PAE ensemble and evaluated on the R&D dataset. Subfigure 7.13(a) shows the mass sculpting calculated at every percentile of the anomaly score distribution and expressed in terms of JS-divergence. Subfigure 7.13(b) the anomaly detection performance represented as ROC curves. The colour coding is identical between the two plots and thus, the legend of the right-hand subfigure denotes the Area Under the ROC Curve (AUC).

The anomaly score's signal discrimination power is just as important as having minimal mass sculpting. Provided the labelled class-balanced synthetic test data set, the anomaly scores can be evaluated as classifiers through the lens of conventional performance indicators such as the Receiver Operating Characteristics (ROC) curve and the area underneath it. More detailed information about ROC curves can be found in Appendix F. Judging the anomaly scores by the Area Under the ROC Curve (AUC) figure of merit their performance looks fairly similar, as Figure 7.13(b) shows. While those AUC values may not stack well against a supervised classification model, they are nonetheless respectable considering this is a semi-supervised method. Reconstruction-based anomaly scores are the worst performers while the Normalizing Flow densities and PAE score all appear to be better, with small differences between them.

Summarized by Figure 7.13, the anomaly score performance comparison yielded interesting results. The PAE score is the least biased while still having one of the best figures of merit for discrimination power. Although the PAE estimation of the input data's negative log-likelihood is based on both Autoencoder reconstruction error and Normalizing Flow density, it doesn't inherit either of those components' m_{jj} biases, while also borrowing the classification performance of the best one.

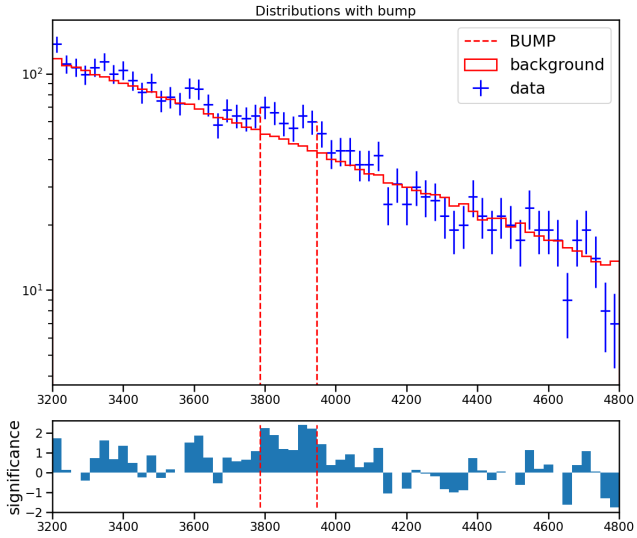


Figure 7.14.: Results of *Bump Hunting* for resonances in **Black Box 1** data, after applying a cut at the 99^{th} percentile of the anomaly score distribution

7.3.3. Results on the Black Box Datasets

The PAE anomaly score performed well in synthetic tests, so the next step is to use it in the full anomaly detection workflow. Anomaly scores were computed for all three *Black Boxes* and after applying a cut on the 99^{th} percentile of the anomaly score distribution, PYBUMPHUNTER was deployed to scan for BSM signal excesses. The bump hunting procedure was attempted on a narrower mass range of $[3.2\,\text{TeV}, 4.2\,\text{TeV}]$ to exclude the peak and the end of the tail of the m_{jj} distribution.

Training the PAE directly on the *Black Boxes* didn't provide the best results in terms of the anomaly score's signal extraction power so the background-only *LHC Olympics* challenge dataset was used instead. To account for the slight differences in modelling between the *Black Boxes* and this pure background sample event reweighting was done a bit differently. Instead of fitting the KNN density estimator on the m_{jj} distribution of the training data, the *Black Boxes* m_{jj} density was the one used. Then, this density estimate was used to compute weights for the training pure-background sample's events. This translates to a **semi-supervised** approach.

Figure 7.14 show the PYBUMPHUNTER output when scanning the $[3.2\,\text{TeV}, 4.2\,\text{TeV}]$ mass range. The *background* in this plot is represented by the m_{jj} distribution of the full **Black Box 1** dataset while the *data* label denotes m_{jj} of the subsample passing the cut 99^{th} percentile of the anomaly score distribution. The most significant *data* excess ($\geq 3\sigma$) is located at 3.86 TeV which matches the **Black Box 1** signal's mass of 3.8 TeV.

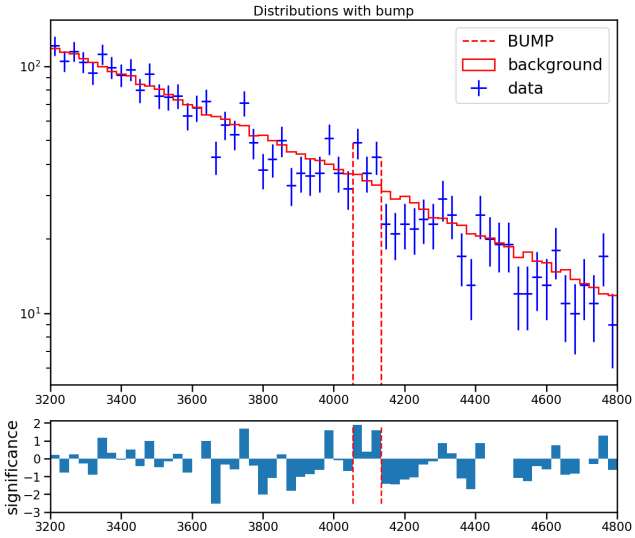


Figure 7.15.: Results of *Bump Hunting* for resonances in **Black Box 2** data, after applying a cut at the 99^{th} percentile of the anomaly score distribution

Noticeable fluctuations are observed in the *data* distribution which may be partly determined by the stringency of the cut, which is discarding 99% of the **Black Box 1** events over the full mass range. However, statistical fluctuations may not be the only factor contributing to this effect. Towards the very end of the selected mass range, the *data-background* mismatch becomes larger, indicating that the PAE's anomaly score still introduces some m_{jj} bias.

The same procedure was applied to **Black Box 2**, and the results are shown in Figure 7.15. The significance of the found excess is much smaller $\leq 1\sigma$, but the *data-background* fluctuations are still present. This further confirms that the model introduces some bias. Since the **Black Box 2** does not contain any signal, the significance of the found excess seems to allow the differentiation between BSM signals and bias-induced signals.

Figure 7.16 shows the result for **Black Box 3**. Unfortunately, the PAE anomaly score is unable to identify the BSM signal, finding no excess more significant than 0.5σ. The challenge posed by the **Black Box 3** lies in the mix of dijet and trijet final states of the bsm signal. Since all models were trained on the features of two leading jets, it is not surprising that the trijet channel can not be found. However, the *data-background* fluctuations seem to be minimal for **Black Box 3**, the bias introduced not being significant enough to generate a *fake* bump.

7.3.4. Further Optimizations

While the Probabilistic Autoencoder (PAE) shows promising results on the *LHC Olympics Black Boxes*, extensive optimization of hyperparameters was attempted in hopes of further mitigating the mass sculpting while increasing the signal extraction power.

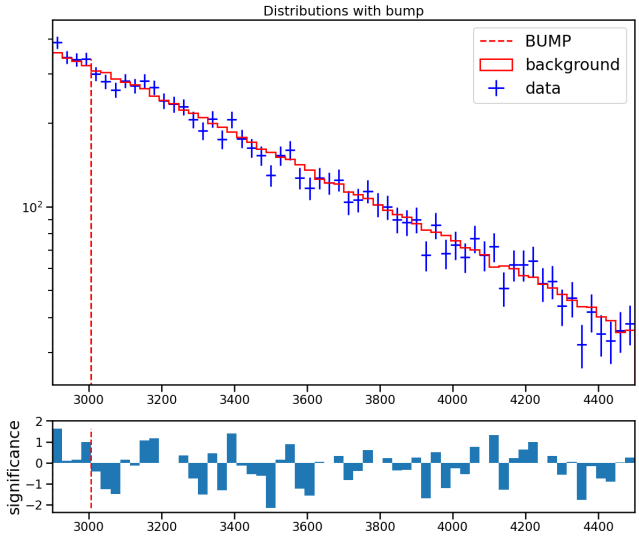


Figure 7.16.: Results of *Bump Hunting* for resonances in **Black Box 3** data, after applying a cut at the 99^{th} percentile of the anomaly score distribution

Several variations of autoencoder depth and latent space size were attempted, as well as introducing **L1** and **L2** regularization. The performance metrics evaluated were JS-divergence at the 99^{th} percentile cut and the S/B in the anomaly-enriched subsample. Additionally, the model's stability was evaluated through the standard deviation of those metrics for the PAE trained using **K-fold cross validation** [151] on the R&D dataset.

Through this process, the optimal configuration for the Autoencoder was a 3-layer encoder and decoder with a 5-dimensional latent space between them. A mix of $\lambda = 10^{-5}$ **L1** and **L2** regularization for both showed benefits in decreasing the variance of the performance metrics between *folds*.

The best performing configuration was tested on **Black Box 1** and the results are shown in Figure 7.17.

Some amount of mass sculpting appears to be unavoidable, regardless of the PAE configuration. Despite this fact, the model is capable of finding the **Black Box 1** signal in a reproducible fashion. After unblinding the dataset, the signal content of the *bump* is immediately apparent. The anomaly-score event selection provides a sample with $S/B = 1.19\%$ which is one whole order of magnitude higher than the initial $S/B = 0.08\%$.

Conclusions and Future Outlook

This chapter introduced the Probabilistic Autoencoder (PAE) and its building blocks while proposing a **semi-supervised** methodology for Beyond Standard Model searches that is

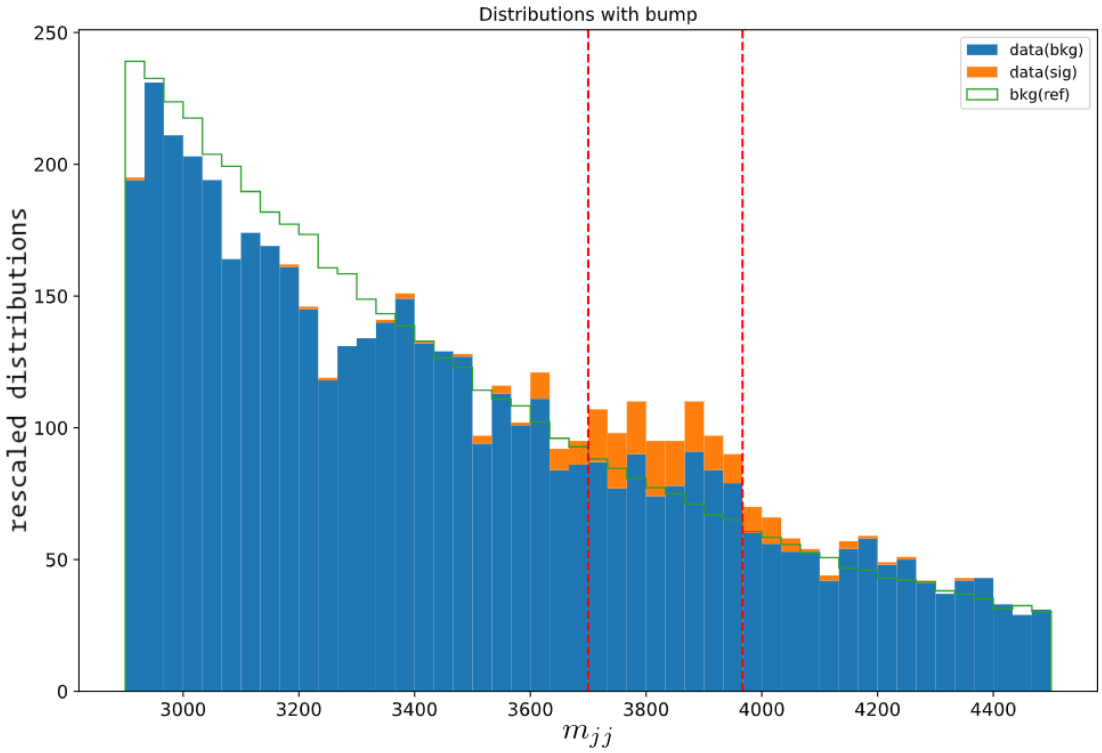


Figure 7.17.: Results of bump hunting after comparing the **Black Box 1** data, depicted as a green line, with the subsample of 99^{th} percentile anomalous events, depicted as coloured bars: blue for background events and orange for signal events. Signal and background labels were added after the fact, the unblinding being used to confirm the presence of real BSM signal.

centred around it. Out of the several anomaly score candidates that could be derived from the PAE, the negative log-likelihood of the inputs was shown to introduce the smallest bias while also being among the best performing in terms of anomaly detection. Furthermore, the Probabilistic Autoencoder offers promising anomaly detection performance in the dijet BSM search benchmark scenario provided by the **Black Box 1**. Mass sculpting remains one of its main issues, but the signal discrimination power seems strong enough to make up for it. Additional refinements of the method may bring further improvements in terms of m_{jj} bias mitigation. A more targeted feature selection and additional mass decorrelation techniques such as **DisCo** [152] have yet to be explored. The method could also be extended to a set of features properly covering trijet final states and thus probe the **Black Box 3**.

Conclusions

There is still plenty of progress potential in the field of High-Energy Particle Physics (HEPP). While the Standard Model (SM) provides an exceptionally good description of the most fundamental aspects of nature, there are some phenomena yet to be accounted for. Competing theories concerning Beyond Standard Model (BSM) physics propose explanations for several remaining mysteries such as neutrino masses, dark matter or quantum gravity. However, the necessary experimental evidence is not yet available for proving or disproving all of them.

Benefiting from and contributing to the A Toroidal LHC Apparatus (ATLAS) experiment at CERN, this thesis covered three major topics relevant to the current state of High-Energy Particle Physics (HEPP).

I successfully concluded and ATLAS authorship Qualification Task (QT) by producing high purity samples for studying electron fakes heavy flavour sources. Alongside developing the selection model, a subsequent study identified the control regions in terms of shower shape variables for the selected heavy-flavour electrons. The results provided key insights into how this type of electron background can be rejected while the sample produced facilitated the training of Machine Learning models aiming to improve electron identification performance.

Regarding data taking and particle detection, as part of the PhD thesis, I developed a prototype for a Machine Learning based trigger algorithm targeting the MicroMegas (MM) detectors deployed with the New Small Wheel (NSW). With the Research and Development (R&D) on this front still being in its earlier stages, the current prototype acts as a proof-of-concept attesting to the feasibility of the approach. The current iteration successfully deploys a Convolutional Neural Network (CNN) which uses the signals from the detector to estimate the most relevant variable involved in the NSW trigger decision.

Lastly, I designed and studied a model-independent anomaly detection analysis workflow targeting applications in searches for Beyond Standard Model physics. The proposed methodology relies on using a Neural Network ensemble to generate *anomaly scores* designed to be sensitive to the presence BSM signals which are not known *a priori*. The Probabilistic Autoencoder (PAE) was chosen as the main model of interest and several potential anomaly scores that could be derived from it were tested on synthetic datasets. A large focus was placed on mitigating potential model biases in order to reduce the likelihood of finding fake signals. Results on the **Black Box 1** dataset, part of the *LHC Olympics* challenge, showcase

the anomaly detection performance of the method. In a semi-supervised scenario, the PAE anomaly score was used to achieve an order of magnitude increase in the ratio between the sample's signal and background content. While there was some bias introduced with respect to the dijet mass, deploying *bump hunting* algorithms to the m_{jj} distribution of the anomaly-enriched sample resulted in a correct identification of the BSM signal against the full **Black Box 1** reference background.

While I have provided tangible results for the topics pursued, my interest in them is far from being exhausted. There are several endeavours that I consider worthwhile as a potential continuation of the work done so far.

Studies on prompt electron identification performance and background sources continue to be relevant in ATLAS' e/γ group while Neural Network (NN) solutions are proposed and studied. In this context, the analysis of heavy-flavour electron background could be further extended to include the other background sources, thus aiding the development of such models.

The proof-of-concept MicroMegas trigger prototype for the ATLAS New Small Wheel (NSW) muon detector enables further trigger algorithm developments. An important next milestone would be the addition of small-strip Thin-Gap Chambers data to the current prototype. With Long Shutdown 3 starting at the end of 2025 and lasting until 2029, a complete Machine Learning based trigger could be completed in time for the next data-taking run.

As for model-independent BSM searches, the Probabilistic Autoencoder approach can further be optimized in terms of reducing the bias introduced by the model. There are potential strategies that haven't yet been explored such as mass decorrelation techniques and using additional jet features.

Research Contributions

During the PhD program, I was recognized as a contributor to the following collection of published research works:

Thesis-related Publications

- G. Kasieczka et al., *The LHC olympics 2020: a community challenge for anomaly detection in high energy physics*, Reports on Progress in Physics (2021)
- I.-M. Dinu, I.-S. Trandafir and C. Alexa, *A Machine Learning Based Muon Trigger Algorithm for an Assembly of Micromegas Detectors*, Romanian Journal of Physics **67**.7-8 (2022) 401
- I.-M. Dinu, 'Event-Level Anomaly Detection for Multijet BSM Searches with Probabilistic Autoencoders', *9th Large Hadron Collider Physics Conference*, 2021 340
- I.-M. Dinu and I.-S. Trandafir, 'A Machine Learning Based Muon Trigger Algorithm for an Assembly of Micromegas Detectors', *20th International Balkan Workshop on Applied Physics and Materials Science*, oral presentation, 2022, URL: http://ibwap.ro/wp-content/uploads/2022/07/book-abstracts_IBWAP2022.pdf
- A. Chirosca, I.-M. Dinu and C. Alexa, 'First Prototype of a Machine Learning Trigger Algorithm on FPGA for Micromegas detectors', *11th International Conference of the Balkan Physical Union*, oral presentation, 2022, URL: https://indico.bpu11.info/event/1/contributions/272/

Other Publications

- G. Aad et al., *Search for lepton-flavor violation in Z-boson decays with τ leptons with the ATLAS detector*, Physical review letters **127**.27 (2021) 271801
- G. Aad et al., *Search for new phenomena in final states with two leptons and one or no b-tagged jets at $\sqrt{s} = 13$ TeV using the ATLAS detector*, Physical review letters **127**.14 (2021) 141801
- G. Aad et al., *Search for R-parity-violating supersymmetry in a final state containing leptons and many jets with the ATLAS experiment using $\sqrt{s} = 13$ TeV proton–proton collision data*, The European Physical Journal C **81**.11 (2021) 1
- G. Aad et al., *Search for exotic decays of the Higgs boson into long-lived particles in pp collisions at $\sqrt{s} = 13$ TeV using displaced vertices in the ATLAS inner detector*, Journal of High Energy Physics **2021**.11 (2021) 1

- G. Aad et al., *Measurement of the production cross section of pairs of isolated photons in pp collisions at 13 TeV with the ATLAS detector*, Journal of High Energy Physics **2021**.11 (2021) 1
- G. Aad et al., *Search for charginos and neutralinos in final states with two boosted hadronically decaying bosons and missing transverse momentum in pp collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, Physical Review D **104**.11 (2021) 112010
- G. Aad et al., *Measurement of the energy response of the ATLAS calorimeter to charged pions from* $W^{\pm} \to \tau^{\pm}(\to \pi^{\pm}\nu_{\tau})\nu_{\tau}$ *events in Run 2 data*, The European Physical Journal C **82**.3 (2022) 1
- G. Aad et al., *Search for heavy particles in the b-tagged dijet mass distribution with additional b-tagged jets in proton-proton collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS experiment*, Physical Review D **105**.1 (2022) 012001
- G. Aad et al., *Measurement of the nuclear modification factor for muons from charm and bottom hadrons in Pb+ Pb collisions at 5.02 TeV with the ATLAS detector*, Physics Letters B **829** (2022) 137077
- G. Aad et al., *Observation of electroweak production of two jets in association with an isolated photon and missing transverse momentum, and search for a Higgs boson decaying into invisible particles at 13 TeV with the ATLAS detector*, The European Physical Journal C **82**.2 (2022) 1
- A. collaboration et al., *Operation and performance of the ATLAS semiconductor tracker in LHC Run 2*, arXiv preprint arXiv:2109.02591 (2021)
- G. Aad et al., *Search for exotic decays of the Higgs boson into* $b\bar{b}$ *and missing transverse momentum in pp collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, Journal of High Energy Physics **2022**.1 (2022) 1
- A. collaboration et al., *AtlFast3: the next generation of fast simulation in ATLAS*, arXiv preprint arXiv:2109.02551 (2021)
- G. Aad et al., *Measurement of the energy asymmetry in* $t\bar{t}j$ *production at 13 TeV with the ATLAS experiment and interpretation in the SMEFT framework*, The European Physical Journal C **82**.4 (2022) 1
- G. Aad et al., *Search for Higgs bosons decaying into new spin-0 or spin-1 particles in four-lepton final states with the ATLAS detector with 139 fb- 1 of pp collision data at* $\sqrt{s} = 13$ *TeV*, Journal of High Energy Physics **2022**.3 (2022) 1
- G. Aad et al., *Measurements of* $W^+W^- + \geq 1$ *jet production cross-sections in* $/pp$ *collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, Journal of High Energy Physics **2021**.6 (2021) 1
- G. Aad et al., *Measurement of Higgs boson decay into b-quarks in associated production with a top-quark pair in* pp *collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, (2021)

- G. Aad et al., *Measurements of azimuthal anisotropies of jet production in Pb+ Pb collisions at* $\sqrt{s_{_{NN}}} = 5.02\,TeV$ *with the ATLAS detector*, Physical Review C **105**.6 (2022) 064903
- G. Aad et al., *Search for flavour-changing neutral-current interactions of a top quark and a gluon in pp collisions at* $\sqrt{s} = 13\,TeV$ *with the ATLAS detector*, The European Physical Journal C **82**.4 (2022) 1
- G. Aad et al., *Determination of the parton distribution functions of the proton using diverse ATLAS data from* pp *collisions at* $\sqrt{s} = 7, 8, 13\,TeV$, The European Physical Journal C **82**.5 (2022) 1
- G. Aad et al., *Search for Higgs boson pair production in the two bottom quarks plus two photons final state in* pp *collisions at* $\sqrt{s} = 13\,TeV$ *with the ATLAS detector*, tech. rep., ATLAS-HDBS-2018-34-003, 2021
- G. Aad et al., *Modelling and computational improvements to the simulation of single vector-boson plus jet processes for the ATLAS experiment*, tech. rep., ATLAS-PMGR-2021-01-003, 2021
- G. Aad et al., *A search for an unexpected asymmetry in the production of* $e^+\mu^-$ *and* $e^-\mu^+$ *pairs in proton–proton collisions recorded by the ATLAS detector at* $\sqrt{s} = 13\,TeV$, Physics Letters B (2022) 137106
- G. Aad et al., *Search for Higgs boson decays into a pair of pseudoscalar particles in the* $bb\mu\mu$ *final state with the ATLAS detector in pp collisions at* $\sqrt{s} = 13\,TeV$, Physical Review D **105**.1 (2022) 012006
- G. Aad et al., *Measurement of the* $t\bar{t}t\bar{t}$ *production cross section in pp collisions at* $\sqrt{s} = 13\,TeV$ *with the ATLAS detector*, Journal of High Energy Physics **2021**.11 (2021) 1
- G. Aad et al., *Search for long-lived charginos based on a disappearing-track signature using* 136fb^{-1} *of pp collisions at* $\sqrt{s} = 13\,TeV$ *with the ATLAS detector*, tech. rep., ATLAS-SUSY-2018-19-003, 2022
- G. Aad et al., *Search for single production of a vectorlike T quark decaying into a Higgs boson and top quark with fully hadronic final states using the ATLAS detector*, Physical Review D **105**.9 (2022) 092012
- G. Aad et al., *Direct constraint on the Higgs-charm coupling from a search for Higgs boson decays into charm quarks with the ATLAS detector*, tech. rep., ATLAS-HIGG-2021-12-003, 2022
- G. Aad et al., *Measurement of the c-jet mistagging efficiency in* $t\bar{t}$ *events using pp collision data at* $\sqrt{s} = 13\,TeV$ *collected with the ATLAS detector*, The European Physical Journal C **82**.1 (2022) 1
- G. Aad et al., *Search for type-III seesaw heavy leptons in leptonic final states in* pp *collisions at* $\sqrt{s} = 13\,TeV$ *with the ATLAS detector*, tech. rep., ATLAS-EXOT-2020-02-003, 2022

- G. Aad et al., *Measurements of the Higgs boson inclusive and differential fiducial cross-sections in the diphoton decay channel with pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, tech. rep., ATLAS-HIGG-2019-13-003, 2022
- G. Aad et al., *Observation of WWW production in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, tech. rep., ATLAS-STDM-2019-09-003, 2022
- G. Aad et al., *Search for invisible Higgs-boson decays in events with vector-boson fusion signatures using* 139fb^{-1} *of proton-proton data recorded by the ATLAS experiment*, tech. rep., ATLAS-EXOT-2020-11-003, 2022
- G. Aad et al., *Measurement of the polarisation of single top quarks and antiquarks produced in the t-channel at $\sqrt{s} = 13$ TeV and bounds on the tWB dipole operator from the ATLAS experiment*, tech. rep., ATLAS-TOPQ-2018-10-003, 2022
- G. Aad et al., *Measurements of differential cross-sections in top-quark pair events with a high transverse momentum top quark and limits on beyond the Standard Model contributions to top-quark pair production with the ATLAS detector at $\sqrt{s} = 13$ TeV*, tech. rep., ATLAS-TOPQ-2019-23-003, 2022
- G. Aad et al., *Search for neutral long-lived particles in pp collisions at $\sqrt{s} = 13$ TeV that decay into displaced hadronic jets in the ATLAS calorimeter*, Journal of High Energy Physics **2022**.6 (2022) 1
- G. Aad et al., *Search for events with a pair of displaced vertices from long-lived neutral particles decaying into hadronic jets in the ATLAS muon spectrometer in pp collisions at $\sqrt{s} = 13$ TeV*, tech. rep., ATLAS-EXOT-2019-24-003, 2022
- G. Aad et al., *Search for associated production of a Z boson with an invisibly decaying Higgs boson or dark matter candidates at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Physics Letters B **829** (2022) 137066
- G. Aad et al., *Constraints on Higgs boson production with large transverse momentum using $H \to b\bar{b}$ decays in the ATLAS detector*, Physical Review D **105**.9 (2022) 092003
- G. Aad et al., *Search for resonant pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state using pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Physical Review D **105**.9 (2022) 092002
- G. Aad et al., *Study of $B_c^+ \to J/\psi D_s^+$ and $B_c^+ \to J/\psi D_s^{*+}$ decays in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Journal of High Energy Physics **2022**.8 (2022) 1
- G. Aad et al., *Measurements of Higgs boson production cross-sections in the $H \to \tau^+\tau^-$ decay channel in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Journal of High Energy Physics **2022**.8 (2022) 1

Appendices

A. New Small Wheel Naming Scheme

This appendix summarises the agreed-upon naming scheme used for defining the components of the New Small Wheel (NSW), as defined in the Technical Design Report (TDR) [42]. A visual representation of this naming scheme is shown in figure A.1, displayed below.

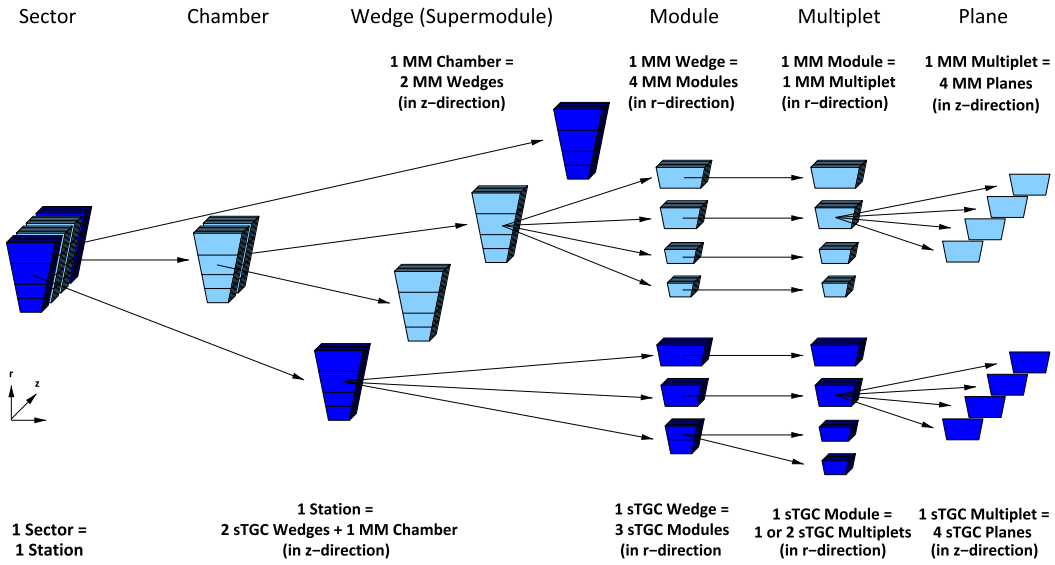


Figure A.1.: Hierarchy of object types making up a NSW sector [42]

.

The explicit definition of the objects schematically represented in Figure A.1 is as follows:

- **plane**: the trapezoidal gas-filled active volume of a single planar detector element using either MM or sTGC technology. **Planes** represent the smallest working particle-detecting units comprising the NSW.
- **multiplet**: a group of four *planes* using identical detector technologies (either MM or sTGC) placed in sequence across the z-direction.
- **module**: an arrangement of one or several *multiplets* along the r-coordinate that constitutes an object independently produced as a single unit.
- **station**: an assembly of four *modules* (two MM and two sTGC stationed along the z-axis)

- **wedge** or **supermodule**: a collection of *modules* sharing the same detector type (either MM or sTGC) and cover an entire *sector* in the $r - \phi$ plane
- **sector**: a 16^{th} of the NSW on either side of the ATLAS Detector.

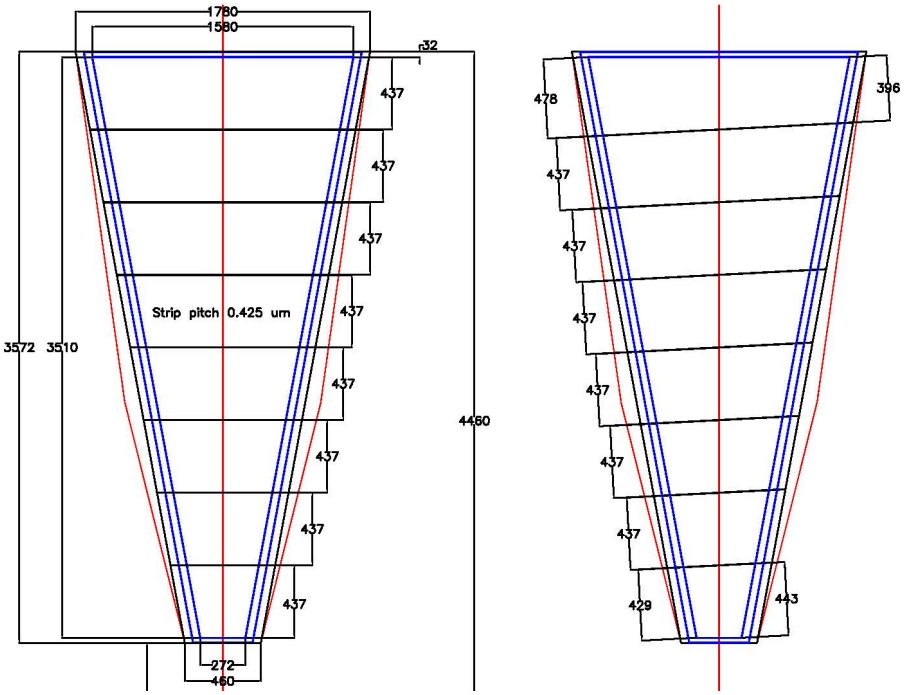
While those terms are equivalent between the two detector types, there is additional terminology associated with each of them, individually.

A.1. MicroMegas

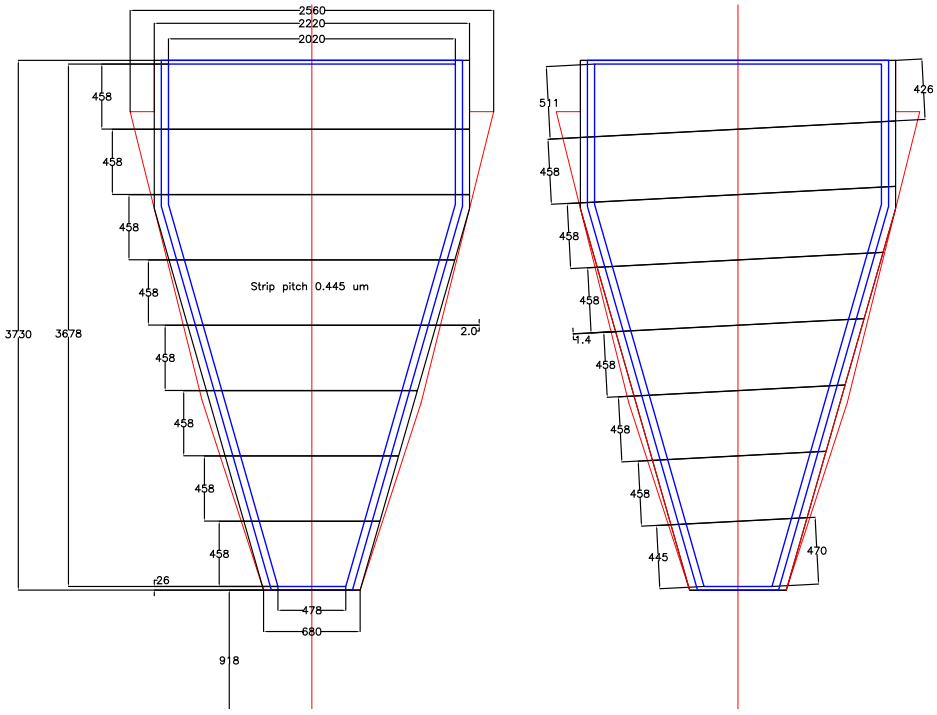
In terms of MicroMegas-specific granularity, the layers making up the multiplets are distinguished by the angle of their strips relative to the perpendicular of the r-axis. Half of the layers are η layers (denoted **X**), which means that their strips are perpendicular to the r-coordinate, as such, the strip numbers are directly related to the η of the particle being detected. Stereo layers feature slightly angled strips (1.5°) relative to the **X** layers and are denoted differently based on the sign of the angle. This results in two types of stereo layers: **U** and **V**, with $+1.5^{\circ}$ and -1.5° strip inclination values. The figure A.2 configuration of PCBs for small A.2(a) and large A.2(b) sectors, featuring an **X** layer on the left and a **U** stereo layer on the right.

A.2. sTGC

The small-strip Thin-Gap Chambers deploys a more straightforward method for measuring a second coordinate. Each sTGC plane contains three separate types of elements that compound each other. The **strips** and **wires** are perpendicular to each other, unlike MM' stereo angle. This enables position measurements in two dimensions but adds a lot of readout channels at the same time. To prevent possible delays of the trigger decisions, the **pads** cover large areas and can register hits quickly, delimiting a region called the **sTGC tower** where readout takes place.



(a) MM small sector planes



(b) MM large sector planes

Figure A.2.: Differences in partitioning of MM planes into PCBs between η and stereo layers. [42].

B. Loss Functions

Loss functions are crucial to model training and optimization. Considering the output of a model $\mathbf{f}$ trained to predict the output $\mathbf{y}$ based on the input $\mathbf{x}$, the following notation will be used for the model prediction $\tilde{\mathbf{y}} = \mathbf{f}(\mathbf{x})$. There are several choices for loss functions, some of the most common being discussed below. Figure B.1 displays the values for those loss functions as a function of the difference between the prediction and the target. A 1D linear regression problem has been used in computing the loss values from Figure B.1. Cosine similarity is computed as in Equation B.5 but without the minus sign.

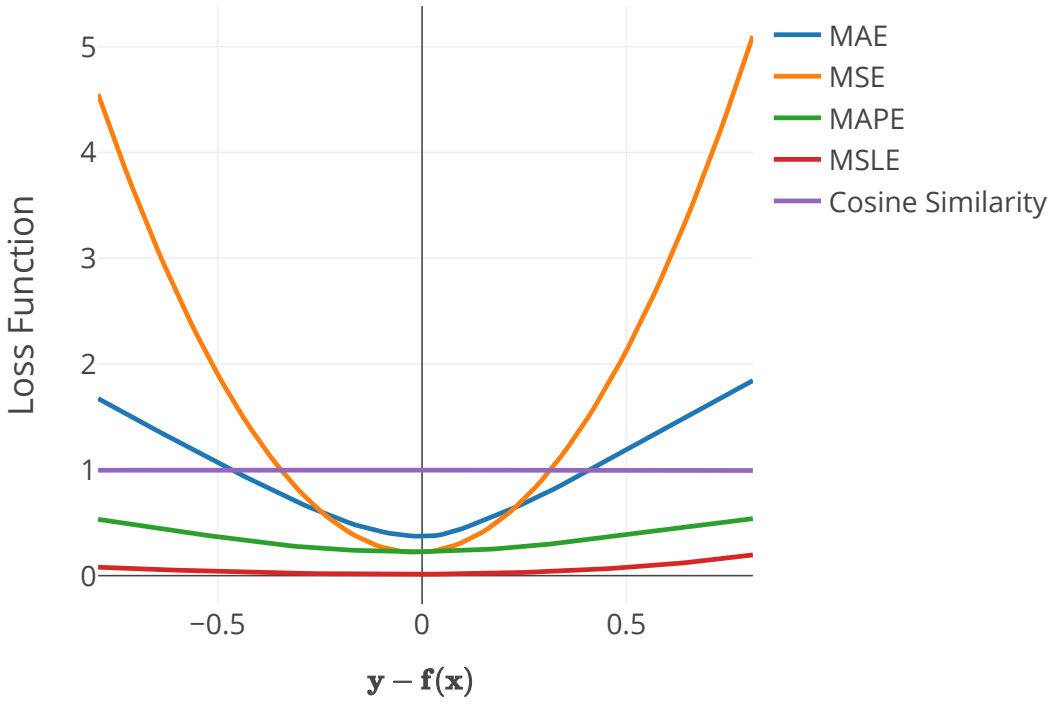


Figure B.1.: Loss function value dependence to the absolute error for a scalar value. Several loss functions are plotted side-by-side.

Mean Absolute Error

Mean Absolute Error (MAE) is defined in Equation B.1 and involves just taking the difference between the model's prediction and the true label in absolute value. It is a linear function and this simplicity makes it easily understandable, but, at the same time, its constant derivative makes it slower to optimize with gradient-based methods.

$$\mathcal{L}_{\text{MAE}}(\tilde{\mathbf{y}}, \mathbf{y}) = \frac{1}{d} \sum_{j=1}^{d} |y^{(j)} - \tilde{y}^{(j)}| \tag{B.1}$$

Mean Squared Error

One of the most popular choices, the Mean Squared Error (MSE) gives the squared value of the prediction's difference from true mapping. Equation B.2 shows how it is calculated. Due to the squaring operation, MSE usually shows fast convergence time towards the optimal parameter values during the training process.

$$\mathcal{L}_{\text{MSE}}(\tilde{\mathbf{y}}, \mathbf{y}) = \frac{1}{d} \sum_{j=1}^{d} (y^{(j)} - \tilde{y}^{(j)})^2 \tag{B.2}$$

Mean Absolute Percentage Error

Equation B.3 describes the Mean Absolute Percentage Error (MAPE). It facilitates interpretability by its values being bound in the $[0, 1]$ interval, thus allowing them to be represented as a percentage. The downside of such an approach becomes apparent when the datasets' target values $\mathbf{y}$ are close to 0. In such circumstances MAPE can go towards infinity while MAE or MSE would be well-behaved functions. MAPE is designed for datasets where loss divergence issues can be avoided and the emphasis falls under the relative value of the error rather than the absolute.

$$\mathcal{L}_{\text{MAPE}}(\tilde{\mathbf{y}}, \mathbf{y}) = \frac{1}{d} \sum_{j=1}^{d} \left| \frac{y^{(j)} - \tilde{y}^{(j)}}{y^{(j)}} \right| \tag{B.3}$$

Mean Squared Logarithmic Error

Some models are required to produce outputs that could span several orders of magnitude. In such cases, a MSE loss would have most of its value coming from errors of the larger-valued targets, assuming a uniform distribution of relative errors. Mean Squared

Logarithmic Error (MSLE) bypasses this issue by applying logarithms before calculating the squared error, just like Equation B.4 is describing.

$$\mathcal{L}_{\text{MSLE}}(\tilde{\mathbf{y}}, \mathbf{y}) = \frac{1}{d} \sum_{j=1}^{d} \left(\log(y^{(j)} + 1) - log(\tilde{y}^{(j)})\right)^2 \tag{B.4}$$

Cosine Similarity

Similarity metrics can be adapted into loss functions by adding a negative sign. Cosine similarity loss is defined in Equation B.5. Cosine similarity ranges between -1 and 1, with 1 meaning that the vectors are completely identical. By adding the minus sign, cosine similarity can be optimized by using the same loss minimization methods as all previous functions.

$$\mathcal{L}_{\text{cosine}}(\tilde{\mathbf{y}}, \mathbf{y}) = -\frac{\tilde{\mathbf{y}} \cdot \mathbf{y}}{|\tilde{\mathbf{y}}||\mathbf{y}|} \tag{B.5}$$

C. Activation Functions for Neural Networks

A Neural Network (NN)'s activation functions have to be non-linear, but there is no ultimately best such function. Historically, the most used such functions were **sigmoid** (Equation C.1) and the hyperbolic tangent (Equation C.2). They have the advantage of constraining the output to the $(0, 1)$ and $(-1, 1)$ intervals, respectively. Constraints such as those are useful for NN outputs because they can be easily rescaled to the interval of the target outputs.

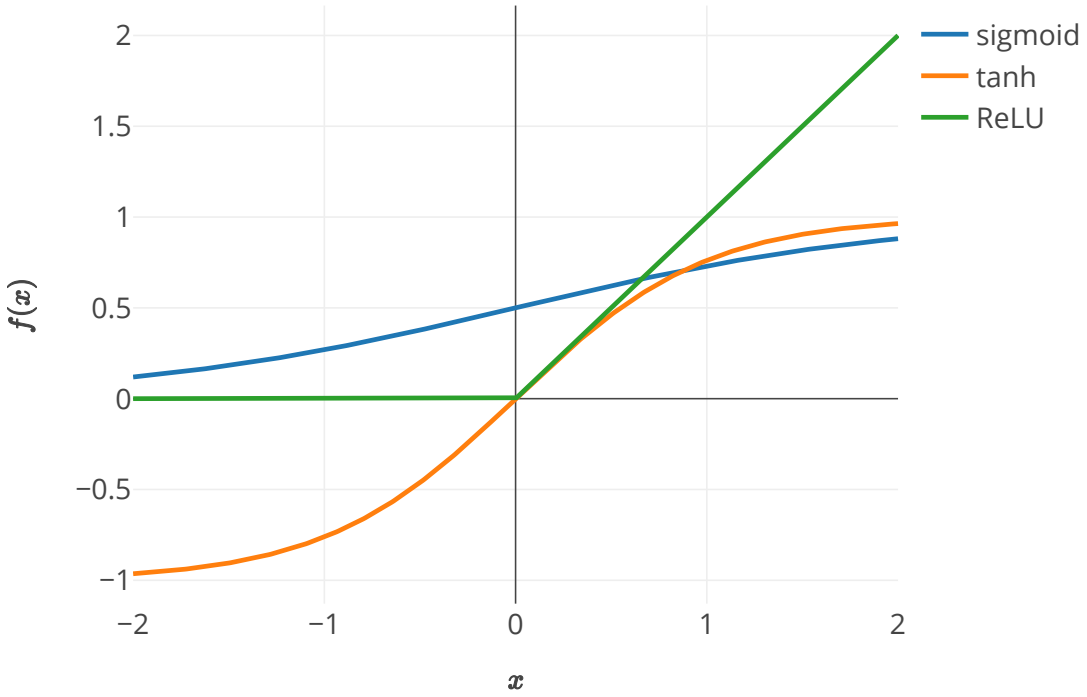


Figure C.1.: Activation functions commonly used in Neural Network

A simpler and computationally inexpensive way of introducing non-linearities would be Rectified Linear Unit (ReLU), which has become increasingly popular in the last decade and is defined in Equation C.3. It behaves like a linear function, but only for positive values, while all negative ones are mapped to zero.

$$\text{sigmoid}(x) = \frac{1}{1+e^{-x}} \tag{C.1}$$

$$\tanh(x) = \frac{2}{1+e^{-2x}} = 2\,\text{sigmoid}(2x) - 1 \tag{C.2}$$

$$\text{ReLU}(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } x \geq 0 \end{cases} \tag{C.3}$$

Figure C.1 shows a plot of the aforementioned activation functions. The **sigmoid** and **tanh** functions converge asymptotically to the bounds of their output interval for large positive and negative values. Computing their derivatives, as displayed in Figure C.2, it is apparent that for increasingly larger values, the gradients of **sigmoid** and **tanh** are vanishing.

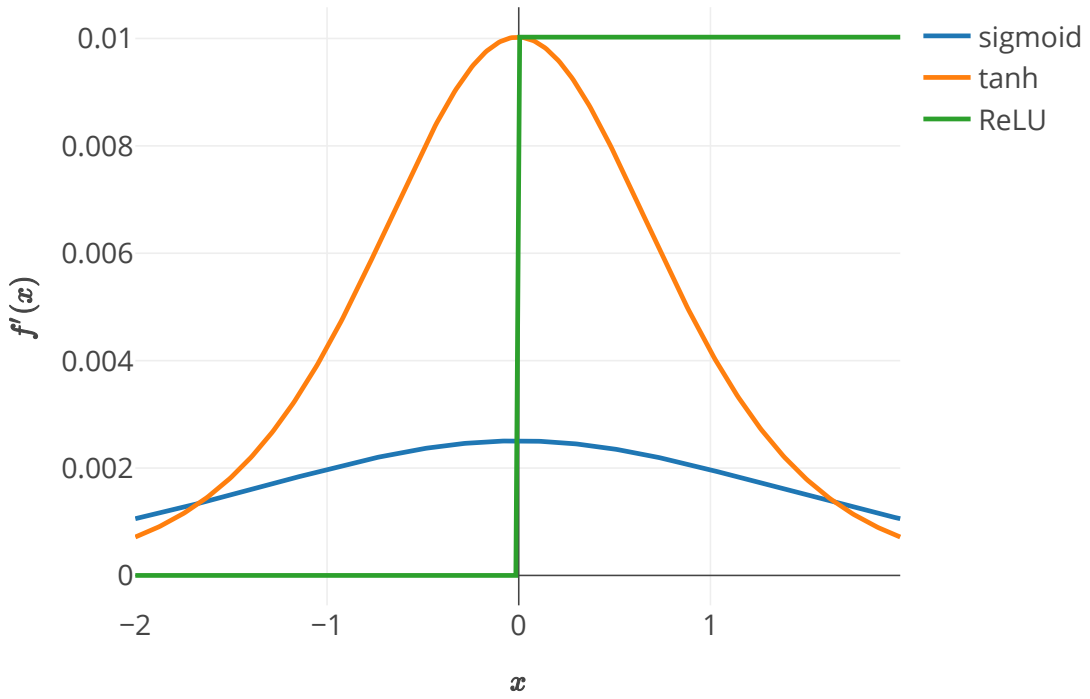


Figure C.2.: Derivatives of common activation functions

Vanishing gradients are a common issue that can negatively interfere with gradient-based optimization methods. In order to prevent this behaviour **sigmoid** and **tanh** tend to not be used for *hidden layers*, being replaced by ReLU. As Figure C.2 shows, ReLU has a constant derivative in the positive domain which prevents gradients from vanishing. There

are, however, potential issues with ReLU in the negative domain. Having a derivative of 0 implies that the optimization algorithm could no longer update the parameters of a negative-output neuron. In such a scenario, if the model assigns a large negative bias to a neuron, it would no longer activate, nor be further optimized by the training process. This is known as the *dead ReLU problem*.

D. Backpropagation in Neural Networks

For this example a neural network with n layers, each having a single node, is considered. The values computed by each neuron will be notated with x_i, where i is the layer index.

Figure D.1.: Neural network with n layers of a single neuron

The loss function $\mathcal{L}_0$ is the Mean Squared Error (MSE) between the output $\tilde{y}(\mathcal{P}) = x_n$ and the expected output y. Equation D.1 shows the loss function and its relation to the model output. The model output is dependent on the parameter set $\mathcal{P} = (w_1, b_1, w_2, b_2, \ldots, w_n, b_n)$, w_i and b_i being the weight and bias associated with the neuron from the i-th layer. Note that the input layer x_0 does not have parameters associated with it.

$$\mathcal{L}_0(\tilde{y}(\mathcal{P}), y) = \left(\tilde{y}(\mathcal{P}) - y\right)^2 \tag{D.1}$$

The output of the neuron in the layer i is calculated using Equation D.2. For simplicity, shorthand notation from Equation D.3 is introduced.

$$x_i = f(w_i x_{i-1} - b_i) \tag{D.2}$$

$$z_i = (w_i x_{(i-1)} - b_i) \implies x_i = f(z_i) \tag{D.3}$$

The parameter dependence of the final neuron is represented in Figure D.2. Gradients of the loss function with respect to the parameters are calculated backwards starting with x_n and moving up the chain.

The general form of the gradient is shown in Equation D.4.

$$\nabla \mathcal{L}_0 = \begin{bmatrix} \frac{\partial \mathcal{L}_0}{\partial w_1} & \frac{\partial \mathcal{L}_0}{\partial b_1} & \frac{\partial \mathcal{L}_0}{\partial w_2} & \frac{\partial \mathcal{L}_0}{\partial b_2} & \cdots & \frac{\partial \mathcal{L}_0}{\partial w_n} & \frac{\partial \mathcal{L}_0}{\partial b_n} \end{bmatrix}^T \tag{D.4}$$

Backpropagation starts by finding the last two elements of our gradient $\frac{\partial \mathcal{L}_0}{\partial w_n}$ and $\frac{\partial \mathcal{L}_0}{\partial b_n}$, using the chain rule. Equation D.5 shows the gradient of the final layer's weight term while equation D.6 corresponds to the bias term.

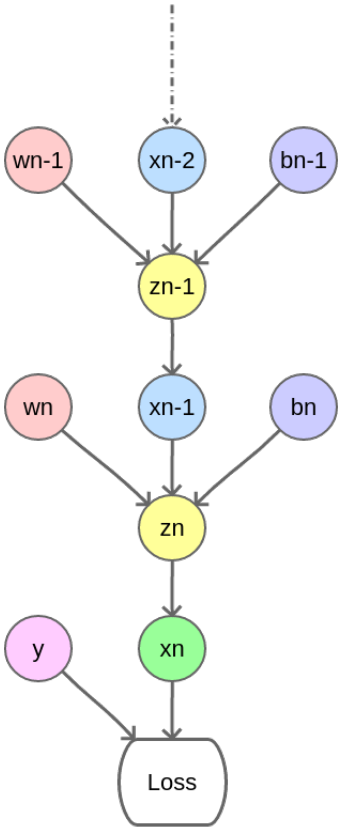


Figure D.2.: Parameter hierarchy relative to the loss function

$$\frac{\partial \mathcal{L}_0}{\partial w_n} = \frac{\partial z_n}{\partial w_n}\frac{\partial x_n}{\partial z_n}\frac{\partial \mathcal{L}_0}{\partial x_n} = x_{n-1}\, f'(x_n)\, 2(x_n - y) \tag{D.5}$$

$$\frac{\partial \mathcal{L}_0}{\partial b_n} = \frac{\partial z_n}{\partial b_n}\frac{\partial x_n}{\partial z_n}\frac{\partial \mathcal{L}_0}{\partial x_n} = 1\, f'(x_n)\, 2(x_n - y) \tag{D.6}$$

From this point, the next gradient vector components are calculated, starting with $\frac{\partial \mathcal{L}_0}{\partial x_{n-1}}$, in Equation D.7.

$$\frac{\partial \mathcal{L}_0}{\partial x_{n-1}} = \frac{\partial z_n}{\partial w_{n-1}}\frac{\partial x_n}{\partial z_n}\frac{\partial \mathcal{L}_0}{\partial x_n} = w_{n-1}\, f'(x_n)\, 2(x_n - y) \tag{D.7}$$

The chain rule is used to find weight $\frac{\partial \mathcal{L}_0}{\partial w_{n-1}}$ and bias $\frac{\partial \mathcal{L}_0}{\partial b_{n-1}}$ terms of the layer $n-1$. Backpropagation continues then with calculating $\frac{\partial \mathcal{L}_0}{\partial x_{n-2}}$ and sequentially moves up the chain until all layers are taken into account and all components of $\nabla \mathcal{L}_0$ are known.

E. Affine Autoregressive Transformations in Normalizing Flows

When defining this bijective mapping the first important aspect is enforcing the **conservation of probability mass**, exemplified in Equation E.1.

$$p(y) = p(x) \left| \frac{\partial f(x)}{\partial x} \right|^{-1} \tag{E.1}$$

For multivariate distributions, this conservation law can be generalized as Equation E.2.

$$p(\mathbf{y}) = p(\mathbf{x}) \left| \det \mathcal{J}(\mathbf{T}(\mathbf{x})) \right|^{-1} \tag{E.2}$$

Notation-wise if we use $\mathbf{x} = (x_1, x_2, x_3, ..., x_d)$ as the target distribution and $\mathbf{z} = (z_1, z_2, z_3, ..., z_d)$ as the source, known-density distribution, the transformation can be defined as in Equation E.3.

$$\mathbf{T} : \mathbb{R}^d \rightarrow \mathbb{R}^d, \quad \mathbf{x} = \mathbf{T}(\mathbf{z}) \tag{E.3}$$

The transformation $\mathbf{T}$ can be thought of as having d scalar components, one for every element of $\mathbf{x}$ such as $x_i = T_i(z_1, z_2, \ldots, z_d)$. Plucking this into the definition of $\mathbf{T}$, Equation E.4 arises.

$$\mathbf{T} = (T_1(\mathbf{z}), T_2(\mathbf{z}), T_3(\mathbf{z}), ..., T_d(\mathbf{z})) \tag{E.4}$$

, where:

$$\begin{aligned} x_1 &= T_1(z_1, z_2, \ldots, z_d) \\ x_2 &= T_2(z_1, z_2, \ldots, z_d) \\ &\ldots \\ x_d &= T_d(z_1, z_2, \ldots, z_d) \end{aligned}$$

Applying transformations requires computing the determinant of the Jacobian shown in Equation E.5.

$$\mathcal{J}(\mathbf{T}(\mathbf{z})) = \nabla_{\mathbf{z}}\mathbf{T}(\mathbf{z}) = \begin{bmatrix} \frac{\partial T_1}{\partial z_1} & \frac{\partial T_1}{\partial z_2} & \cdots & \frac{\partial T_1}{\partial z_d} \\ \frac{\partial T_1}{\partial z_2} & \frac{\partial T_2}{\partial z_2} & \cdots & \frac{\partial T_2}{\partial z_d} \\ \vdots & \vdots & \ddots & \\ \frac{\partial T_d}{\partial z_d} & \frac{\partial T_d}{\partial z_d} & \cdots & \frac{\partial T_d}{\partial z_d} \end{bmatrix} \tag{E.5}$$

However, matrix determinant computation is computationally expensive ($\mathcal{O}(d^3)$ complexity). Since the transformation parameters have to be learned, determinant computations would be required at every step of the training process. This issue can be sidetracked by using **increasing triangular maps**.

The theorem stated in Equation E.6, where $\mathbf{T}$ is a unique increasing triangular map, states that such a triangular map is sure to exist regardless of the distributions chosen.

$$\forall\, p(\mathbf{z}), q(\mathbf{x})\ \exists\, \mathbf{T},\ q(\mathbf{x}) = p(\mathbf{z})\left|\det \mathcal{J}(\mathbf{z})\right|^{-1} \tag{E.6}$$

Assuming a triangular form would imply the following variable transformation components definition:

$$\begin{aligned} x_1 &= T_1(z_1) \\ x_2 &= T_2(z_1, z_2) \\ &\dots \\ x_k &= T_k(z_{i\leq k}) \\ &\dots \\ x_d &= T_d(z_1, z_2, \dots, z_d) \end{aligned}$$

With the triangular Jacobian shown in Equation E.7, the determinant computation can be reduced to the product of the main diagonal elements.

$$\nabla_{\mathbf{z}}\mathbf{T}(\mathbf{z}) = \begin{bmatrix} \frac{\partial T_1}{\partial z_1} & 0 & \cdots & 0 \\ \frac{\partial T_1}{\partial z_2} & \frac{\partial T_2}{\partial z_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \\ \frac{\partial T_d}{\partial z_d} & \frac{\partial T_d}{\partial z_d} & \cdots & \frac{\partial T_d}{\partial z_d} \end{bmatrix} \tag{E.7}$$

The only thing left remains the parametrization of the bijective mapping. Using the probability chain rule the $q(\mathbf{x})$ can be split as Equation E.8

$$q(\mathbf{x}) = q_1(x_1)q_2(x_2|x_1)q_3(x_3|x_1, x_2)\dots q_d(x_d|x_{i<d}) \tag{E.8}$$

Then, each q_i can be parametrized as a gaussian as shown in Equation E.9.

$$q(\mathbf{x}) = \mathcal{N}(\mu_1, \sigma_1^2)\mathcal{N}(\mu_2, \sigma_2^2)\ \ \mathcal{N}(\mu_3, \sigma_3^2)\ldots\mathcal{N}(\mu_d, \sigma_d^2) \tag{E.9}$$

Finally, the transformation can be defined, by components, as:

$$\begin{aligned}
x_1 &= T_1(z_1) := \sigma_1 z_1 + \mu_1 \\
x_2 &= T_2(z_1, z_2) := \sigma_2(z_1) z_2 + \mu_2(z_1) \\
x_3 &= T_3(z_1, z_2, z_3) := \sigma_3(z_1, z_2) z_3 + \mu_3(z_1, z_2) \\
&\ldots \\
x_d &= T_d(z_1, z_2, \ldots, z_d) := \sigma_d(z_{<d}) z_d + \mu_d(z_{<d})
\end{aligned}$$

μ_i, σ_i, with $i \in (1, 2, \ldots, d)$, are the *trainable parameters* while functions $\mu_i(z_{<i})$ and $\sigma_i(z_{<i})$ can be modelled as **autoregressive networks**.

F. Classification Performance Metrics

Binary classification targets usually take on the value 0 to represent members of the *negative class* and the value 1 to represent the members of the *positive class*. Classifier outputs lie within the $[0, 1]$ interval, indicating the likelihood of a data point belonging to a certain class. The Receiver Operating Characteristics (ROC) curve is a common way of representing the performance of a classification algorithm by relating its output to the true target value. Figure F.1 shows an example of a classification algorithm output which will be used as a toy example to illustrate how ROC curves work.

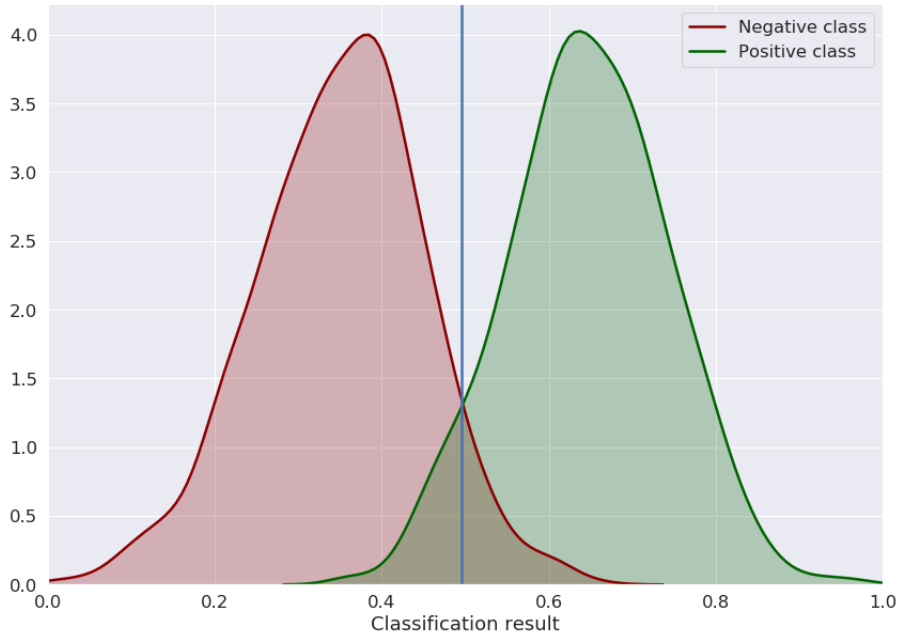


Figure F.1.: An example of a classification algorithm output distribution, the blue line represents the threshold at 0.3

Before the trained model can be used for inference, a **threshold** needs to be applied. The output values greater than the threshold would indicate that the input belongs to the positive class and the output values below the threshold denote the input is a member of the negative class. ROC curves provide a detailed illustration of how the model is expected to perform at all different relevant thresholds.

Before quantifying performance, the following terminology needs to be introduced. True Positive (TP) denotes positive-class samples that are classified correctly while True Negative (TN) corresponds to negative samples classified correctly. Conversely, misclassified points fall into the False Positive (FP) and False Negative (FN) categories.

In the example from Figure F.1, the FP and FN samples are found in the area given by the intersection of the two distributions. FP points would lie on the left side of the threshold whereas FN examples would be on the right side. The areas of the distributions that do not intersect represent the correctly classified events: TN (left of the threshold) and TP (right of the threshold).

More general performance indicators can be obtained by expressing truth in terms of sample fractions rather than absolute counts of observations. Such quantities are the True Positive Rate (TPR) also called **Sensitivity** or **Recall** as in equation F.1.

$$TPR = \frac{TP}{TP + FN} \tag{F.1}$$

Another useful term is **Specificity** defined in equation F.2 and based on it we define in equation F.3 the True Negative Rate (TRN).

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{F.2}$$

$$FPR = 1 - \text{Specificity} = \frac{FP}{TN + FP} \tag{F.3}$$

In order to get the ROC curve, we just need to plot FPR against TPR for every threshold value as in figure F.2. This plot offers an overview of the accuracy of the classification algorithm. In order to make comparisons between classifiers easier, we can use the Area Under the ROC Curve (AUC) **figure of merit**. A value closer to 1 for the AUC relates to a better classifier.

There is an inverse relationship between the amount of overlap between the two distributions and the AUC as shown in the example from Figure F.3

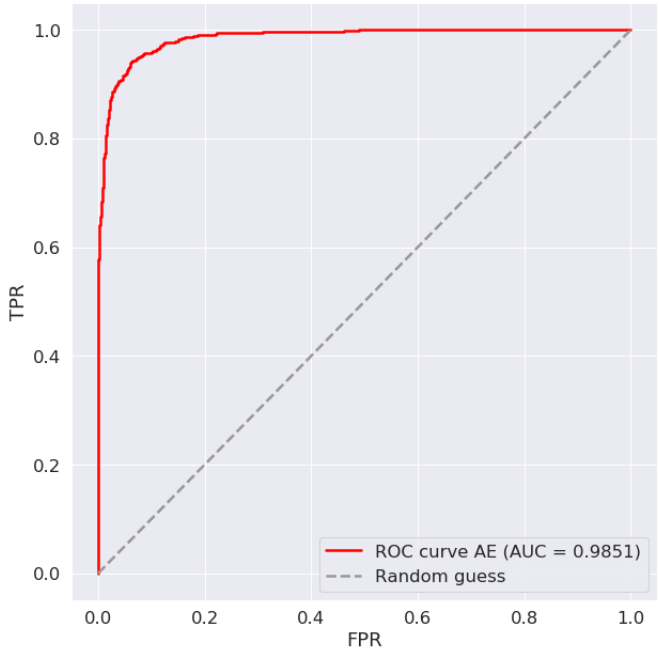


Figure F.2.: The ROC curve calculated for the distributions in figure F.1

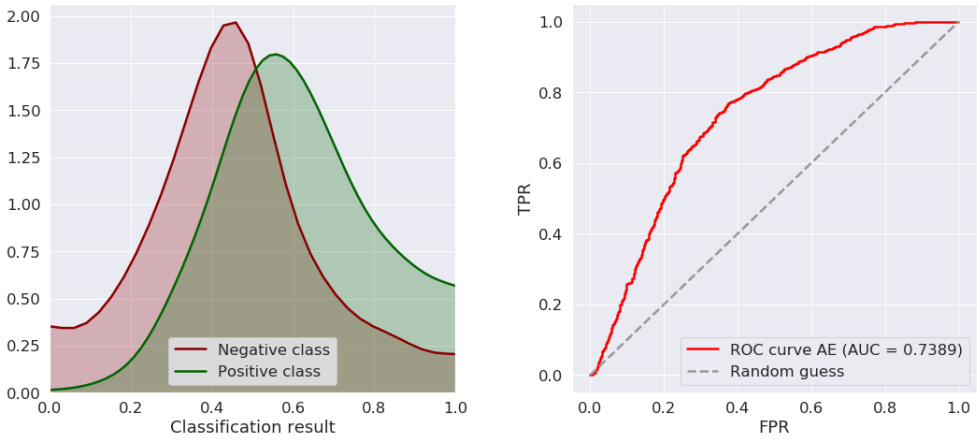


Figure F.3.: A classification output with large overlaps between the distributions of the two classes (left) and the ROC curve calculated for it (right)

G. Elements of Information Theory

In digital communication, *bits* are the unit of measure for quantifying information. Before using bits to transfer information one must first define a protocol for assigning them meaning, a process sometimes referred to as *encoding*. There is much to be considered when encoding data. For example, when transmitting information about an event occurring, it may be useful to take into account its probability of occurrence. This probability is relevant because a binary representation can be thought of as a series of successive *yes/no* questions of which some of which may have varying degrees of redundancy. One may calculate the amount of *useful* information using Equation G.1 [200].

$$\text{amount of useful bits} = -\log_2(p) \tag{G.1}$$

where p is the probability of the event occurring.

As a basic example let's consider we want to encode the colour of a ball drawn from a box. There are four possible colours and the probability and encoding associated with each are shown in Table G.1.

Table G.1.: Example of an encoding table

Colour	Probability	Encoding
Red	6.25%	`00`
Blue	13.75%	`01`
Green	30%	`10`
Yellow	50%	`11`

When a yellow ball is drawn, the two bits used to send the message are `11`, but the amount of useful bits is $-\log_2(0.5) = 1$, as obtained from equation G.1. Conversely, if a red ball is drawn, the binary representation of the message would be `00`, but the amount of useful bits of information received is $-\log_2(\frac{1}{16}) = 4$.

Information **entropy** of a data probability distribution is defined as the average amount of information (in bits), obtained from sampling the probability distribution. If the distribution is discrete, Equation G.2 should be applied.

$$H(P) = -\sum_i p_i \log_2(p_i) \tag{G.2}$$

where P is the distribution, i is the index for the possible values and p_i is the probability of sampling the value indexed by i.

For the example shown in Table G.1 we can calculate:

$$\begin{aligned} H &\approx 0.0625 * 4 + 0.1375 * 2.862 + 0.3 * 1.7369 + 0.5 * 1 \\ &\approx 1.66 \end{aligned}$$

Although we use two bits to send information, on average, 1.66 **useful** bits are received.

If we assume a different distribution Q for the data following distribution P the expected message length can be obtained by calculating the **cross-entropy** as shown in equation G.3.

$$H(P,Q) = -\sum_i p_i \log_2(q_i) \tag{G.3}$$

Using the difference between entropy and cross-entropy, we can now define a measure of how much the distributions P and Q are different from one another; this is called the Kullback-Leibler (KL) divergence [201]. Equation G.4 shows the definition KL-divergence.

$$D_{KL}(P||Q) = H(P,Q) - H(P) = -\sum_i p_i \log_2(\frac{q_i}{p_i}) \tag{G.4}$$

D_{KL} is neither smooth nor upper bounded, which makes it sometimes difficult to use in data analysis. In an attempt to avoid those inconveniences, a symmetric, $[0,1]$-bounded version of the KL-divergence can be defined, known as the Jensen-Shannon (JS) divergence [24] and displayed in Equation G.5.

$$D_{JS}(P||Q) = \frac{1}{2} D_{KL}(P||\frac{P+Q}{2}) + \frac{1}{2} D_{KL}(Q||\frac{P+Q}{2}) \tag{G.5}$$

Glossary

Symbols

IRC safety property of a quantity to be insensitive to the presence of soft QCD emissions or collinear branching. 66

***b*-tagging** the process of identification for jets containing *b*-hadrons, namely *b*-jets. 68, 69, 76, 200

DATA18 all data collected by the ATLAS Detector in 2018, as part of Run 2 . 80– 86, 201

A

ATLAS Collaboration the international group of physicists belonging to different universities and research centres who built and run the ATLAS Detector . 40, 45, 68, 72, 75, 86, 169

ATLAS Detector CERN's general-purpose detector, designed to measure the broadest possible range of signals at the LHC. 21, 35, 40, 41, 45, 47, 48, 50, 54, 55, 57, 59, 60, 63, 65, 73, 97, 146, 169, 172, 200

B

bunch at CERN's LHC, the proton beams are not divided into *chunks* separated by gaps. Those *chunks* are called commonly known as **bunches**. . 169

bunch-crossing the interaction between two bunches of protons, consisting of several *pp* collisions. . 47, 97, 99, 100, 102, 170, 175

C

Combined Performance within the ATLAS Collaboration there are several so-called Combined Performance (CP) groups, whose primary role is developing and maintaining algorithms for identifying, measuring and reconstructing physics objects . 61, 75, 86, 169, 176

crack region The transition region between sub-systems where the detector performance may be affected. For the ATLAS Electromagnetic (EM) calorimeter, it covers the $1.37 \geq |\eta| \leq 1.52$ pseudorapidity interval . 77, 83, 84

cut a selection criterion applied to data during analysis. *Cutting* on the distribution of an observable involves choosing a threshold and selecting only the events having the observable's value above or below it.. 69, 76–86, 98, 127, 130, 132–134, 201, 203, 205

D

derivation a pre-selection of recorded events in order to cater for specific physics uses. It involves **skimming** (removal of whole events based on specific criteria), **thinning** (removal of individual physics objects that are not particularly aligned with the derivation's use case) and **slimming** (removal of variables associated with certain objects, that are deemed irrelevant) . 73, 80, 85, 86, 201

E

emittance the area occupied by the beam's particles in the phase space of position and momentum. 38, 173

event the outcome of a fundamental interaction taking place between subatomic particles, occurring in a very small time frame and a well-defined region of space. In ATLAS, an event is the result of a bunch-crossing. 21, 36, 41, 48, 57, 58, 61, 67–69, 72, 73, 75–86, 99, 114, 123–130, 132, 170, 172, 200, 201, 203, 205

F

fake physics objects that are miss-reconstructed and/or miss-identified as something different than what they are. 43, 50, 59, 85, 86, 99, 102, 137, 171

FODO cell a structure of magnets containing alternating between focusing and defocusing beam quadrupole *lenses*. 37, 199

G

Good Run List a set of Luminosity Blocks () that passed the data quality requirements necessary to be labeled as *good for physics analysis use* . 76, 176

H

hermeticity a hermetic detector is designed to observe as many decay products as possible from any event, through and expended coverage around the interaction point and the inclusion of several different detector systems . 46

hyperparameter parameters associated with the configuration of a Machine Learning (ML) model or with the learning process. They are not learned from data but rather user-specified.. 90, 96, 116

J

jet physics object representative of a narrow cone of particles resulting from the hadronisation of a quark of gluon . 19, 22, 57, 65–71, 76, 80, 81, 114, 123, 124, 126, 127, 129, 133, 138, 169, 171, 200

jet tagging a jet classification task based on the underlying particle that produced the jet and/or the properties of its constituent particles . 73

L

Long Shutdown a time during which the Large Hadron Collider (LHC) has no active beam. During those times, the Large Hadron Collider (LHC) and its experiments are usually undergoing significant hardware upgrades. . 19, 42, 50, 72, 138

Luminosity Block a time interval of data-taking where the experimental recording conditions, specifically the instantaneous luminosity, are constant over the entire duration . 72, 170, 178

N

natural related to theoretical physics, naturalness is the property of models to have constant parameters whose ratios don't span several orders of magnitude and don't require very precise tuning in order to reliably explain observed phenomena. 33

non-prompt label assigned to a physics object, indicating that it's not originating from the main interaction generated by the collision. This includes both objects that are correctly identified but are decay products of the prompt objects and fakes . 59, 61, 75, 79, 82, 83, 85, 86, 201

O

overlap removal an analysis procedure whose goal is to ensure that there aren't multiple physics objects reconstructed from the same detector signals. In jet reconstruction, **overlap removal** enforces that no constituent particle is included more than a single jet . 65, 76, 77

P

physics object entity reconstructed with the detector that corresponds to one or many physical particles. They often are characterized by properties such as four-momentum, electric charge and synthetic variables associated with the confidence of the reconstruction's outcome . 21, 41, 43, 57–59, 61, 63–65, 70, 75–77, 79, 114, 126, 170–172, 200

prompt a physics object is considered *prompt* if it originated from the **primary vertex** of the collision, as opposed to being a decay product of another *prompt* physics objects . 12, 59– 61, 63, 64, 76, 79, 80, 84, 85, 138, 171

R

Run 1 the first data-taking period was performed between 2010-2011 at 7 TeV and during 2012 at 8 TeV . 45, 49, 68

Run 2 the second data-taking period performed between 2015 and 2018 at the same centre-of-mass energy of $\sqrt{s} = 13$ TeV . 36, 45, 60, 61, 68, 169, 200

Run 3 the third data-taking period scheduled between 2021 and 2025 at a constant expected centre-of-mass energy of $\sqrt{s} = 13.6$ TeV . 47, 49– 51, 86, 102, 103

S

shower a cascade of secondary particles produced in the interaction of a high-energy particle with matter or by a hard scattering process started by elementary particles carrying colour charge. . 65, 82– 86, 201

side in the context of the ATLAS Detector, the two ends of the detector all called **sides**. Those are called **A-side** and **C-side**, **B** denoting the central barrel region. As a point of reference, **A-side** is located along the positive z-axis, in the direction of the Geneva Airport.. 49, 50, 146, 200

stereo a **stereo** view generally denotes two different representations of the same underlying entity from distinct viewpoints. In the context of the MicroMegas detector planes deployed within the New Small Wheel (NSW), the **stereo** planes are referencing the same concept. The observed entity is the muon in this case, and due to the different angles of the two **stereo** detector layers, two unique *viewpoints* of the same muon track are recorded at the same time.. 54, 55, 105, 146, 147, 203

T

trigger system designed for preliminary physics object reconstruction and identification. Its main use consists of quickly singling out event that may be interesting for physics analysis. The *trigger decision* determines if the recorded events are kept or discarded. . 18, 21, 45, 76, 137

Twiss parameters also referred to as *Courant-Snyder parameters* [202] describe a particle beam in terms of the positions and velocities of its constituent particles. While moving through the accelerator ring, every particle's position x and velocity $\dot{x}$ alongside a given axis will always be bound by an ellipse in the phase space. The equation of

this ellipse is given by $\gamma x^2 + 2\alpha x\dot{x} + \beta\dot{x}^2 = \epsilon$, where ϵ in the emittance and the **twiss parameters** are: γ, α and β.. 38

Acronyms

A

AE Autoencoder. 1, 14, 114–116, 121–123, 131, 134, 203

ALICE A Large Ion Collider Experiment. 36, 39

AOD Analysis Object Data. 73, 176

ASIC Application-Specific Integrated Circuit. 48

ATLAS A Toroidal LHC Apparatus. 1, 3–6, 18, 19, 21, 40, 42–47, 50–52, 54, 55, 57, 58, 60, 61, 63, 64, 67, 68, 72, 73, 75–78, 86, 97, 99, 102, 112, 113, 137, 138, 169, 170, 199–201

AUC Area Under the ROC Curve. 17, 131, 164, 203

B

BC bunch-crossing. 100, 102

BDT Boosted Decision Tree. 68, 69

BOOSTER Proton Synchrotron Booster. 35

BSM Beyond Standard Model. 1, 3, 12–15, 17–19, 21, 22, 34, 40, 69, 113, 114, 121, 123, 124, 126, 127, 132–135, 137, 138, 199, 203

BT Barrel Toroid. 5, 45

C

CERN European Organization for Nuclear Research. 18, 35, 36, 48, 72–74, 137, 169, 176, 200

CI Continuous Integration. 74

CMS Compcat Muon Solenoid. 40

CNN Convolutional Neural Network. 1, 8, 11, 12, 18, 21, 72, 106, 107, 110, 112, 137, 205

CP Combined Performance. 61, 69, 75, 76, 169

CPU Central Processing Unit. 48

CSC Cathode Strip Chambers. 5, 45, 46

CVMFS CERN Virtual Machine File System. 74

D

DAOD Derived Analysis Object Data. 73

DNN Deep Neural Network. 69, 92, 93

E

ECTs End-Cap Toroids. 5, 45

EDM Event Data Model. 72, 73

EM Electromagnetic. 4, 41, 43–45, 48, 59, 62, 63, 83–86, 169, 176, 201

EMEC LAr EM end-cap. 45

ESD Event Summary Data. 73

EW Electroweak. 29, 30

F

FCal LAr forward calorimeter. 45

FE Front End. 48

FELIX Front End Link Interface eXchange. 103, 202

FN False Negative. 164

FP False Positive. 164

FPGA Field-Programmable Gate Array. 1, 5, 9, 12, 21, 48, 102–105, 107, 112, 202

FPR False Positive Rate. 164

G

GPU Graphics Processing Unit. 1, 10, 21, 93, 107

GRL Good Run List. 76

GSF Gaussian-sum filter. 60

GUT Grand Unified Theory. 3, 34

H

HDL Hardware Design Language. 103

HEC LAr hadronic end-cap. 45

HEPP High-Energy Particle Physics. 1, 12, 18, 21, 35, 66, 113, 123, 137

HL-LHC High-Luminosity LHC. 48, 51, 200

HLS High-Level Synthesis. 103

HLT High-Level Trigger. 48, 49, 99, 101, 103, 202

I

IAF Inverse Autoregressive Flow. 120

IBL Insertable B-Layer. 42

ID identification. 59, 61–64, 75, 82, 83, 85, 200, 201, 205

ID3 Iterative Dichotomizer 3. 90

IFF Isolation and Fake Forum. 79, 81, 201

IP interaction point. 1, 4, 21, 38–42, 64, 97–100, 170, 202

IRC Infrared and Collinear. 66, 70, 169

J

JS Jensen-Shannon. 16, 130, 131, 134, 168, 203

K

KDE Kernel Density Estimation. 114, 116

KL Kullback-Leibler. 168

KNN k-Nearest Neighbors. 90, 116, 128, 129, 132, 203

L

L0 Level-0. 1, 101–103, 202

L1 Level-1. 48, 97–99, 101, 102, 202

L2 Level-2. 48, 97, 99

LAr Liquid Argon. 4, 41, 43–45, 59, 63, 176, 177

LB Luminosity Block. 170

LEIR Low Energy Ion Ring. 36

LHC Large Hadron Collider. 3, 5, 15, 18, 21, 35–42, 45, 47–51, 72, 137, 169, 171, 175, 177, 180, 199, 200, 205

LQG Loop Quantum Gravity. 3, 34

LUT Look-up Table. 99, 101

M

MAE Mean Absolute Error. 10, 89, 108, 150

MAF Masked Autoregressive Flow. 120

MAPE Mean Absolute Percentage Error. 10, 108, 150

MC Monte Carlo. 6, 68, 72, 73, 78–86, 105, 126, 201, 205

MDT Monitored Drift Tubes. 5, 45, 46, 98, 102, 103, 202

MET Missing Transverse Energy. 43

MicroMegas MICRO-MEsh GAseous Structure. 52, 172

ML Machine Learning. 1, 5, 8, 13, 14, 18, 19, 21, 69, 72, 75, 87–90, 92, 93, 97, 103, 104, 106, 112–114, 116, 123, 137, 138, 170

MM MicroMegas. 1, 6–8, 12, 18, 19, 21, 50–55, 97, 99–102, 105–109, 112, 137, 138, 145–147, 200, 202, 203, 205

MSE Mean Squared Error. 15, 108, 122, 130, 150, 157

MSLE Mean Squared Logarithmic Error. 150

MSSM Minimal Supersymmetric Standard Model. 3, 34, 178

MuCTPI Muon to Central Trigger Processor Interface. 99, 102, 103, 202

N

NF Normalizing Flow. 14, 15, 17, 117–123, 130, 131

NICE Non-linear Independent Components Estimation. 120

NLL Negative log-likelihood. 120

NMSSM Next-to- Minimal Supersymmetric Standard Model. 3, 34

NN Neural Network. 1, 13, 19, 21, 22, 58, 67, 69, 75, 86, 87, 90–93, 103, 105, 109, 113–115, 121, 137, 138, 153, 176, 201–204

NSW New Small Wheel. 1, 6, 7, 12, 18, 19, 21, 35, 49–55, 97, 99, 102, 103, 105, 106, 109, 112, 137, 138, 145, 146, 172, 199, 200, 202, 203

O

OS opposite sign. 78

P

PAE Probabilistic Autoencoder. 2, 14–19, 22, 121–123, 127, 129–135, 137, 138, 199, 203

PCA Principal Component Analysis. 115

PCB Printed Circuit Board. 54, 105, 146, 147, 203

PDF Probability Density Function. 60, 64, 68

PS Proton Synchrotron. 35, 36

Q

QCD Quantum Chromodynamics. 2, 28, 29, 33, 65, 67, 124, 125, 128–130, 169, 203

QED Quantum Electrodynamics. 2, 26–28, 31, 33

QFT Quantum Field Theory. 2, 23, 26, 33

QT Qualification Task. 75, 85, 137

R

R&D Research and Development. 18, 97, 104, 124, 125, 127–131, 134, 137, 203

RDO Raw Data Object. 73

ReLU Rectified Linear Unit. 9, 92, 107, 153–155

RMS Root Mean Squared. 38, 39

RMSProp Root Mean Square Propagation. 93, 95

ROC Receiver Operating Characteristics. 16, 17, 131, 163–165, 175, 203, 204

ROD Readout Driver. 48

RoI Region of Interest. 48, 97, 99

RPC Resistive-Plate Chambers. 45, 46, 97, 98, 103, 202

RTL Register Transfer Level. 103

S

SCT Semiconductor Tracker. 42, 43, 57

SGD Stochastic Gradient Descent. 93, 95

SiF sillicon-associdated forward muon. 63

SL Sector Logic. 102, 103, 202

SM Standard Model. 1–3, 18, 21, 23–26, 30, 32–34, 40, 123, 137, 199, 205

SPS Super Proton Synchrotron. 35, 36

SS same sign. 78

sTGC small-strip Thin-Gap Chambers. 6, 19, 50–56, 99, 100, 102, 105, 109, 112, 138, 145, 146, 200

SUSY Supersymmetry. 3, 34, 40

T

TDAQ Trigger and Data Acquisition. 48, 72, 73, 101, 201

TDR Technical Design Report. 45, 46, 50, 53, 97, 145, 200

TGC Thin-Gap Chambers. 5, 19, 45, 46, 50, 51, 55, 97–99, 103, 112, 138, 146, 180, 202

TMVA ROOT Toolkit for Multivariate Data Analysis. 68

TN True Negative. 164

TP True Positive. 164

TP Trigger Processor. 1, 21

TPR True Positive Rate. 164

TRN True Negative Rate. 164

TRT Transition Radiation Tracker. 42, 43, 62, 63

V

VM Virtual Machine. 74, 176

W

WLCG Worldwide LHC Computing Grid. 72, 73, 201

Bibliography

[1] Y. Fukuda et al., *Evidence for oscillation of atmospheric neutrinos*, Physical review letters **81** (1998) 1562 (cit. on pp. 2, 33).

[2] Q. R. Ahmad et al., *Direct evidence for neutrino flavor transformation from neutral-current interactions in the Sudbury Neutrino Observatory*, Physical review letters **89** (2002) 011301 (cit. on pp. 2, 33).

[3] A. Bilal, *Introduction to Supersymmetry*, 2001, URL: https://arxiv.org/abs/hep-th/0101055 (cit. on pp. 3, 34).

[4] C. CSÁKI, *THE MINIMAL SUPERSYMMETRIC STANDARD MODEL*, Modern Physics Letters A **11** (1996) 599, URL: https://doi.org/10.1142%2Fs021773239600062x (cit. on pp. 3, 34).

[5] U. Ellwanger, C. Hugonie and A. M. Teixeira, *The Next-to-Minimal Supersymmetric Standard Model*, Physics Reports **496** (2010) 1, URL: https://doi.org/10.1016%2Fj.physrep.2010.07.001 (cit. on pp. 3, 34).

[6] C. Rovelli, *Loop Quantum Gravity*, Living Reviews in Relativity **1** (1998), URL: https://doi.org/10.12942%2Flrr-1998-1 (cit. on pp. 3, 34).

[7] J. Pequenao, 'Computer generated image of the ATLAS Muons subsystem', 2008, URL: https://cds.cern.ch/record/1095929 (cit. on pp. 4, 46).

[8] ATLAS Collaboration, *ATLAS Muon Spectrometer: Technical Design Report*, ATLAS-TDR-10; CERN-LHCC-97-022, CERN, 1997, URL: https://cds.cern.ch/record/331068 (cit. on pp. 5, 35, 45, 46).

[9] T. A. collaboration, *Operation of the ATLAS trigger system in Run 2*, Journal of Instrumentation **15** (2020) P10004, URL: https://doi.org/10.1088%2F1748-0221%2F15%2F10%2Fp10004 (cit. on pp. 6, 47, 48).

[10] A. Collaboration, *Athena*, 2019, URL: https://doi.org/10.5281/zenodo.2641997 (cit. on pp. 6, 73, 105).

[11] W. A., *oct_sim*, https://github.com/annmwang/oct_sim, 2021 (cit. on pp. 7, 105).

[12] R. Brun and F. Rademakers, *ROOT: An object oriented data analysis framework*, Nucl. Instrum. Meth. A **389** (1997) 81, ed. by M. Werlen and D. Perret-Gallix (cit. on pp. 7, 74, 105).

[13] S. Ioffe and C. Szegedy, *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*, ArXiv **abs/1502.03167** (2015) (cit. on pp. 9, 107).

[14] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, *Dropout: A Simple Way to Prevent Neural Networks from Overfitting*, Journal of Machine Learning Research **15** (2014) 1929, URL: http://jmlr.org/papers/v15/srivastava14a.html (cit. on pp. 9, 107).

[15] D. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, 2014 (cit. on pp. 10, 93, 107).

[16] S. Caron, L. Hendriks and R. Verheyen, *Rare and Different: Anomaly Scores from a combination of likelihood and out-of-distribution models to detect new physics at the LHC*, SciPost Physics (2022) (cit. on pp. 14, 121).

[17] V. M. Boehm and U. Seljak, *Probabilistic Autoencoder*, Transactions of Machine Learning Research (2022), URL: https://openreview.net/forum?id=AEoYjvjKVA (cit. on pp. 14, 121, 122).

[18] G. Papamakarios, E. T. Nalisnick, D. J. Rezende, S. Mohamed and B. Lakshminarayanan, *Normalizing Flows for Probabilistic Modeling and Inference*, J. Mach. Learn. Res. **22** (2021) 57:1 (cit. on pp. 14, 114, 119, 120).

[19] P. S. Laplace, *Memoir on the Probability of the Causes of Events*, Statistical Science **1** (1986) 364, ISSN: 08834237, URL: http://www.jstor.org/stable/2245476 (visited on 27/09/2022) (cit. on pp. 14, 122).

[20] G. Kasieczka et al., *The LHC Olympics 2020 a community challenge for anomaly detection in high energy physics*, Reports on Progress in Physics **84** (2021) (cit. on pp. 15, 123–125).

[21] G. Choudalakis, 'On hypothesis testing, trials factor, hypertests and the BumpHunter', *PHYSTAT 2011*, 2011, arXiv: 1101.0390 [physics.data-an] (cit. on pp. 16, 127).

[22] G. Van Rossum and F. L. Drake, *Python 3 Reference Manual*, Scotts Valley, CA: CreateSpace, 2009, ISBN: 1441412697 (cit. on pp. 16, 126, 127).

[23] T. Kluyver et al., 'Jupyter Notebooks – a publishing format for reproducible computational workflows', *Positioning and Power in Academic Publishing: Players, Agents and Agendas*, ed. by F. Loizides and B. Schmidt, IOS Press, 2016 87 (cit. on pp. 16, 127).

[24] J. Lin, *Divergence measures based on the Shannon entropy*, IEEE Trans. Inf. Theory **37** (1991) 145 (cit. on pp. 16, 130, 168).

[25] F. Halzen and A. Martin, *Quark & Leptons: An Introductory Course In Modern Particle Physics*, Wiley student edition, Wiley India Pvt. Limited, 2008, ISBN: 9788126516568 (cit. on p. 23).

[26] R. L. Workman et al., *Review of Particle Physics*, PTEP **2022** (2022) 083C01 (cit. on pp. 23, 25).

[27] C. N. Yang and R. L. Mills, *Conservation of Isotopic Spin and Isotopic Gauge Invariance*, Phys. Rev. **96** (1 1954) 191, URL: https://link.aps.org/doi/10.1103/PhysRev.96.191 (cit. on pp. 23, 28, 29).

[28] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes and R. P. Hudson, *Experimental Test of Parity Conservation in β Decay*, Phys. Rev. **105** (1957) 1413 (cit. on p. 23).

[29] D. J. Gross and F. Wilczek, *Ultraviolet Behavior of Non-Abelian Gauge Theories*, Phys. Rev. Lett. **30** (26 1973) 1343, URL: https://link.aps.org/doi/10.1103/PhysRevLett.30.1343 (cit. on p. 23).

[30] S. L. Glashow, *Partial-symmetries of weak interactions*, Nuclear Physics **22** (1961) 579, ISSN: 0029-5582, URL: https://www.sciencedirect.com/science/article/pii/0029558261904692 (cit. on pp. 23, 29, 30).

[31] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (21 1967) 1264, URL: https://link.aps.org/doi/10.1103/PhysRevLett.19.1264 (cit. on pp. 23, 29).

[32] A. Salam, *Weak and Electromagnetic Interactions*, Conf. Proc. C **680519** (1968) 367 (cit. on pp. 23, 29).

[33] C. Sandbox Studio, *The Standard Model of particle physics*, Symmetry Magazine, 2015 (cit. on p. 24).

[34] E. Noether, *Invariante Variationsprobleme*, Nachr. d. König. Gesellsch. d. Wiss. zu Göttingen, Math-phys. Klasse, Seite 235-157 (1918), eprint: www.physics.ucla.edu/\~\cwp/articles/noether.trans/german/emmy235.html, URL: http://www.physics.ucla.edu/~cwp/articles/noether.trans/german/emmy235.html (cit. on p. 27).

[35] P. Higgs, *Broken symmetries, massless particles and gauge fields*, Physics Letters **12** (1964) 132, ISSN: 0031-9163, URL: https://www.sciencedirect.com/science/article/pii/0031916364911369 (cit. on p. 30).

[36] J. Ellis, *Higgs Physics*, (2015) 117, 52 pages, 45 figures, Lectures presented at the ESHEP 2013 School of High-Energy Physics, to appear as part of the proceedings in a CERN Yellow Report, arXiv: 1312.5672, URL: https://cds.cern.ch/record/1638469 (cit. on p. 31).

[37] R. Aaij et al., *Test of lepton universality in beauty-quark decays*, Nature Phys. **18** (2022) 277, arXiv: 2103.11769 [hep-ex] (cit. on p. 33).

[38] D. Croon, T. E. Gonzalo, L. Graf, N. Košnik and G. White, *GUT Physics in the Era of the LHC*, Frontiers in Physics **7** (2019), URL: https://doi.org/10.3389%2Ffphy.2019.00076 (cit. on p. 34).

[39] O. S. Brüning et al., *LHC Design Report*, CERN Yellow Reports: Monographs, Geneva: CERN, 2004, URL: https://cds.cern.ch/record/782076 (cit. on pp. 35, 38, 39).

[40] M. Benedikt, P. Collier, V. Mertens, J. Poole and K. Schindl, *LHC Design Report*, CERN Yellow Reports: Monographs, Geneva: CERN, 2004, URL: https://cds.cern.ch/record/823808 (cit. on p. 35).

[41] G. Aad et al., *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003. 437 p, Also published by CERN Geneva in 2010, URL: https://cds.cern.ch/record/1129811 (cit. on pp. 35, 57, 98).

[42] ATLAS Collaboration, *ATLAS New Small Wheel: Technical Design Report*, ATLAS-TDR-020; CERN-LHCC-2013-006, 2013, URL: https://cds.cern.ch/record/1552862 (cit. on pp. 35, 50–55, 97, 99, 100, 102, 103, 109, 112, 145, 147).

[43] E. Mobs, *The CERN accelerator complex - August 2018. Complexe des accélérateurs du CERN - Août 2018*, (2018), General Photo, URL: https://cds.cern.ch/record/2636343 (cit. on p. 36).

[44] A. Ball, *Technical challenges of the Large Hadron Collider experiments (ATLAS and CMS). Technical challenges of the Large Hadron Collider experiments (ATLAS and CMS)*, Philos. Trans. R. Soc. Lond. A **373** (2014) 20140045. 10 p, URL: https://cds.cern.ch/record/2295117 (cit. on p. 37).

[45] A. Team, 'Diagram of an LHC dipole magnet. Schéma d'un aimant dipôle du LHC', 1999, URL: https://cds.cern.ch/record/40524 (cit. on p. 38).

[46] *ATLAS inner detector: Technical Design Report, 1*, Technical design report. ATLAS, Geneva: CERN, 1997, URL: https://cds.cern.ch/record/331063 (cit. on p. 42).

[47] A. Airapetian et al., *ATLAS detector and physics performance: Technical Design Report, 1*, Technical design report. ATLAS, Geneva: CERN, 1999, URL: https://cds.cern.ch/record/391176 (cit. on pp. 42, 43, 48, 57).

[48] J. Pequenao, 'Computer generated image of the ATLAS inner detector', 2008, URL: https://cds.cern.ch/record/1095926 (cit. on p. 43).

[49] Y. Takubo, *ATLAS IBL operational experience*, tech. rep., CERN, 2016, URL: https://cds.cern.ch/record/2235541 (cit. on p. 42).

[50] J. Pequenao, 'Computer generated image of the ATLAS Liquid Argon', 2008, URL: https://cds.cern.ch/record/1095928 (cit. on p. 44).

[51] J. Pequenao, 'Computer Generated image of the ATLAS calorimeter', 2008, URL: https://cds.cern.ch/record/1095927 (cit. on p. 44).

[52] *High Luminosity LHC Project*, https://hilumilhc.web.cern.ch/content/hl-lhc-project, Accessed: 2022-06-01 (cit. on p. 48).

[53] M. Brice and J. M. Ordan, *New Small Wheel (NSW) descent into ATLAS' "side A" of the experimental cavern*, (2021), General Photo, URL: https://cds.cern.ch/record/2775525 (cit. on p. 49).

[54] Y. Giomataris, P. Rebourgeard, J. P. Robert and G. Charpak, *MICROMEGAS: A High granularity position sensitive gaseous detector for high particle flux environments*, Nucl. Instrum. Meth. A **376** (1996) 29 (cit. on p. 52).

[55] D. Attié et al., *Current Status and Future Developments of Micromegas Detectors for Physics and Applications*, Applied Sciences **11** (2021), ISSN: 2076-3417, URL: https://www.mdpi.com/2076-3417/11/12/5362 (cit. on p. 55).

[56] J. Pequenao and P. Schaffner, 'How ATLAS detects particles: diagram of particle paths in the detector', 2013, URL: https://cds.cern.ch/record/1505342 (cit. on p. 58).

[57] ATLAS Collaboration, *Performance of the ATLAS track reconstruction algorithms in dense environments in LHC Run 2*, Eur. Phys. J. C **77** (2017) 673, arXiv: 1704.07983 [hep-ex] (cit. on pp. 57, 59).

[58] K. Grimm et al., *Primary vertex reconstruction at the ATLAS experiment*, tech. rep., CERN, 2017, URL: https://cds.cern.ch/record/2253428 (cit. on p. 58).

[59] M. Aaboud et al., *Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC*, Eur. Phys. J. C **77** (2017) 332, arXiv: 1611.10235 [physics.ins-det] (cit. on p. 58).

[60] ATLAS Collaboration, *Development of ATLAS Primary Vertex Reconstruction for LHC Run 3*, ATL-PHYS-PUB-2019-015, 2019, URL: https://cds.cern.ch/record/2670380 (cit. on p. 58).

[61] ATLAS Collaboration, *Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton–proton collision data at* $\sqrt{s} = 13\,TeV$, Eur. Phys. J. C **79** (2019) 639, arXiv: 1902.04655 [hep-ex] (cit. on pp. 59–62).

[62] T. G. Cornelissen et al., *The global* χ^2 *track fitter in ATLAS*, Journal of Physics: Conference Series **119** (2008) 032013, URL: https://doi.org/10.1088/1742-6596/119/3/032013 (cit. on pp. 60, 63).

[63] R. Frühwirth, *Application of Kalman filtering to track and vertex fitting*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **262** (1987) 444, ISSN: 0168-9002, URL: https://www.sciencedirect.com/science/article/pii/0168900287908874 (cit. on p. 60).

[64] R. Frühwirth, *A Gaussian-mixture approximation of the Bethe-Heitler model of electron energy loss by bremsstrahlung*, Computer Physics Communications **154** (2003) 131, ISSN: 0010-4655, URL: https://www.sciencedirect.com/science/article/pii/S0010465503002923 (cit. on p. 60).

[65] ATLAS Collaboration, *Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at* $\sqrt{s} = 13\,TeV$, Eur. Phys. J. C **81** (2020) 578, arXiv: 2012.00578 [hep-ex] (cit. on pp. 63, 64).

[66] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*, Eur. Phys. J. C **77** (2017) 490, arXiv: 1603.02934 [hep-ex] (cit. on p. 65).

[67] ATLAS Collaboration, *Jet reconstruction and performance using particle flow with the ATLAS Detector*, Eur. Phys. J. C **77** (2017) 466, arXiv: 1703.10485 [hep-ex] (cit. on p. 65).

[68] R. Atkin, *Review of jet reconstruction algorithms*, Journal of Physics: Conference Series **645** (2015) 012008, URL: https://doi.org/10.1088/1742-6596/645/1/012008 (cit. on p. 66).

[69] G. P. Salam, *Towards jetography*, The European Physical Journal C **67** (2010) 637, URL: https://doi.org/10.1140%2Fepjc%2Fs10052-010-1314-6 (cit. on p. 66).

[70] M. Cacciari, G. P. Salam and G. Soyez, *FastJet user manual*, Eur. Phys. J. C **72** (2012) 1896, arXiv: 1111.6097 [hep-ph] (cit. on pp. 67, 126).

[71] N. Dawe et al., *scikit-hep/pyjet: Legacy version 1.9.0*, version 1.9.0, 2022, URL: https://doi.org/10.5281/zenodo.6787002 (cit. on pp. 67, 126).

[72] ATLAS Collaboration, *Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run*, ATL-PHYS-PUB-2017-013, 2017, URL: https://cds.cern.ch/record/2273281 (cit. on pp. 67, 68).

[73] ATLAS Collaboration, *Performance of b-jet identification in the ATLAS experiment*, JINST **11** (2016) P04008, arXiv: 1512.01094 [hep-ex] (cit. on p. 68).

[74] ATLAS Collaboration, *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 **79** (2019) 970, arXiv: 1907.05120 [hep-ex] (cit. on p. 68).

[75] ATLAS Collaboration, *Secondary vertex finding for jet flavour identification with the ATLAS detector*, ATL-PHYS-PUB-2017-011, 2017, URL: https://cds.cern.ch/record/2270366 (cit. on p. 68).

[76] ATLAS Collaboration, *Topological b-hadron decay reconstruction and identification of b-jets with the JetFitter package in the ATLAS experiment at the LHC*, ATL-PHYS-PUB-2018-025, 2018, URL: https://cds.cern.ch/record/2645405 (cit. on p. 68).

[77] A. Hoecker et al., *TMVA - Toolkit for Multivariate Data Analysis*, 2007, arXiv: physics/0703039 [physics.data-an] (cit. on p. 68).

[78] ATLAS Collaboration, *Performance of top-quark and W-boson tagging with ATLAS in Run 2 of the LHC*, Eur. Phys. J. C **79** (2019) 375, arXiv: 1808.07858 [hep-ex] (cit. on p. 69).

[79] D. E. Soper and M. Spannowsky, *Finding top quarks with shower deconstruction*, Phys. Rev. D **87** (2013) 054012, arXiv: 1211.3140 [hep-ph] (cit. on p. 69).

[80] T. Plehn, M. Spannowsky, M. Takeuchi and D. Zerwas, *Stop Reconstruction with Tagged Tops*, JHEP **10** (2010) 078, arXiv: 1006.2833 [hep-ph] (cit. on p. 69).

[81] J. Pearkes, W. Fedorko, A. Lister and C. Gay, *Jet Constituents for Deep Neural Network Based Top Quark Tagging*, (2017), arXiv: 1704.02124 [hep-ex] (cit. on p. 69).

[82] J. Thaler and K. Van Tilburg, *Identifying Boosted Objects with N-subjettiness*, JHEP **03** (2011) 015, arXiv: 1011.2268 [hep-ph] (cit. on p. 70).

[83] G. Aad et al., *ATLAS Measurements of the Properties of Jets for Boosted Particle Searches*, Phys. Rev. D **86** (2012) 072006, arXiv: 1206.5369 [hep-ex] (cit. on p. 70).

[84] L. G. Almeida et al., *Substructure of high-p_T Jets at the LHC*, Phys. Rev. D **79** (2009) 074017, arXiv: 0807.0234 [hep-ph] (cit. on p. 70).

[85] P. T. Komiske, E. M. Metodiev and J. Thaler, *Energy flow polynomials: A complete linear basis for jet substructure*, JHEP **04** (2018) 013, arXiv: 1712.07124 [hep-ph] (cit. on p. 70).

[86] L. de Oliveira, M. Kagan, L. Mackey, B. Nachman and A. Schwartzman, *Jet-images — deep learning edition*, JHEP **07** (2016) 069, arXiv: 1511.05190 [hep-ph] (cit. on pp. 70, 127).

[87] F. Akesson and E. Moyse, 'Event data model in ATLAS', *14th International Conference on Computing in High-Energy and Nuclear Physics*, 2005 255 (cit. on p. 72).

[88] K. Bos et al., *LHC computing Grid: Technical Design Report. Version 1.06 (20 Jun 2005)*, tech. rep., 2005, URL: http://cds.cern.ch/record/840543 (cit. on p. 72).

[89] I. Bird et al., *Update of the Computing Models of the WLCG and the LHC Experiments*, tech. rep., 2014, URL: https://cds.cern.ch/record/1695401 (cit. on p. 72).

[90] A. Krasznahorkay, *Data Handling and Analysis in ATLAS*, Hungarian HEP group seminar, 2000 (cit. on p. 73).

[91] ISO, *ISO/IEC 14882:2020 Information technology — Programming languages — C++*, Geneva, Switzerland: International Organization for Standardization, 2020 1853, URL: https://www.iso.org/standard/79358.html (cit. on p. 73).

[92] P. Buncic, C. Aguado-Sanchez, J. Blomer and A. Harutyunyan, *CernVM: Minimal maintenance approach to virtualization*, Journal of Physics: Conference Series **331** (2011) 052004, URL: https://doi.org/10.1088/1742-6596/331/5/052004 (cit. on p. 74).

[93] J. Blomer, G. Ganis, S. Mosciatti and R. Popescu, *Towards a serverless CernVM-FS*, EPJ Web of Conferences **214** (2019) 09007 (cit. on p. 74).

[94] I. Antcheva et al., *ROOT – A C++ framework for petabyte data storage, statistical analysis and visualization*, Computer Physics Communications **180** (2009) 2499 (cit. on p. 74).

[95] M. Barisits et al., *Rucio: Scientific Data Management*, Computing and Software for Big Science **3** (2019) 11, ISSN: 2510-2044, URL: https://doi.org/10.1007/s41781-019-0026-3 (cit. on p. 74).

[96] I.-M. Dinu, *Selection of heavy flavour electron fakes in $t\bar{t}$*, tech. rep., CERN, 2021, URL: https://cds.cern.ch/record/2751665 (cit. on p. 75).

[97] G. Aad et al., *Electron and photon performance measurements with the ATLAS detector using the 2015-2017 LHC proton-proton collision data*, Journal of Instrumentation **14** (2019) P12006, ISSN: 1748-0221, URL: http://dx.doi.org/10.1088/1748-0221/14/12/P12006 (cit. on pp. 75, 77).

[98] *Identification of electrons using a deep neural network in the ATLAS experiment*, tech. rep., All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2022-022: CERN, 2022, URL: https://cds.cern.ch/record/2809283 (cit. on p. 75).

[99] ATLAS Egamma Group, *TagAndProbeFrame*, https://gitlab.cern.ch/ATLAS-EGamma/Software/ElectronID/TagAndProbeFrame (cit. on p. 76).

[100] *Optimisation of the ATLAS b-tagging performance for the 2016 LHC Run*, tech. rep., All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2016-012: CERN, 2016, URL: https://cds.cern.ch/record/2160731 (cit. on p. 76).

[101] A. Straessner, M. Schott and the Atlas Collaboration, *A new tool for measuring detector performance in ATLAS*, Journal of Physics: Conference Series **219** (2010) 032023, URL: https://doi.org/10.1088/1742-6596/219/3/032023 (cit. on p. 77).

[102] ATLAS Isolation and Fake Forum, *IFF Truth CLassifier*, https://gitlab.cern.ch/ATLAS-IFF/IFFTruthClassifier (cit. on pp. 77, 79).

[103] ATLAS Collaboration, *Study of jets produced in association with a W boson in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, Phys. Rev. D **85** (2012) 092002, arXiv: 1201.1276 [hep-ex] (cit. on p. 81).

[104] ATLAS Collaboration, *Measurement of the $t\bar{t}$ production cross-section in the lepton+jets channel at $\sqrt{s} = 13$ TeV with the ATLAS experiment*, Phys. Lett. B **810** (2020) 135797, arXiv: 2006.13076 [hep-ex] (cit. on p. 81).

[105] ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton–proton collision data*, ATLAS-CONF-2016-024, 2016, URL: https://cds.cern.ch/record/2157687 (cit. on p. 83).

[106] I. Goodfellow, Y. Bengio and A. Courville, *Deep Learning*, http://www.deeplearningbook.org, MIT Press, 2016 (cit. on pp. 87, 114, 115).

[107] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Berlin, Heidelberg: Springer-Verlag, 2006, ISBN: 0387310738 (cit. on p. 87).

[108] O. Chapelle, B. Schölkopf and A. Zien, *Semi-Supervised Learning (Adaptive Computation and Machine Learning)*, The MIT Press, 2006, ISBN: 0262033585 (cit. on p. 88).

[109] J. Bassey, L. Qian and X. Li, *A Survey of Complex-Valued Neural Networks*, 2021, URL: https://arxiv.org/abs/2101.12249 (cit. on p. 88).

[110] T. Cover and P. Hart, *Nearest neighbor pattern classification*, IEEE Transactions on Information Theory **13** (1967) 21 (cit. on pp. 90, 116).

[111] J. R. Quinlan, *Induction of Decision Trees*, **1** (1986), ISSN: 0885-6125, URL: https://doi.org/10.1023/A:1022643204877 (cit. on p. 90).

[112] D. Wolpert and W. Macready, *No free lunch theorems for optimization*, IEEE Transactions on Evolutionary Computation **1** (1997) 67 (cit. on p. 90).

[113] K. Hornik, M. Stinchcombe and H. White, *Multilayer Feedforward Networks Are Universal Approximators*, Neural Netw. **2** (1989) 359, ISSN: 0893-6080 (cit. on p. 91).

[114] I. Safran and O. Shamir, 'Depth-Width Tradeoffs in Approximating Natural Functions with Neural Networks', *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, ICML'17, Sydney, NSW, Australia: JMLR.org, 2017 2979 (cit. on p. 92).

[115] L. B. Rall, 'Automatic Differentiation: Techniques and Applications', *Lecture Notes in Computer Science*, 1981 (cit. on p. 92).

[116] H. B. Curry, *The method of steepest descent for non-linear minimization problems*, Quarterly of Applied Mathematics **2** (1944) 258 (cit. on p. 92).

[117] H. Robbins and S. Monro, *A Stochastic Approximation Method*, The Annals of Mathematical Statistics **22** (1951) 400, URL: https://doi.org/10.1214/aoms/1177729586 (cit. on p. 92).

[118] G. Orr and K. Müller, *Neural Networks: Tricks of the Trade*, Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2003, ISBN: 9783540494300, URL: https://books.google.ro/books?id=VCKqCAAAQBAJ (cit. on p. 92).

[119] G. Hinton, *rmsprop: Devide the gradient by a running average of its*, Series of lectures: Neural Networks for Machine Learning, Lecture 6e, 2012 (cit. on p. 93).

[120] ATLAS Collaboration, *ATLAS TDAQ System Phase-I Upgrade: Technical Design Report*, ATLAS-TDR-023; CERN-LHCC-2013-018, 2013, URL: https://cds.cern.ch/record/1602235 (cit. on p. 97).

[121] ATLAS Collaboration, *ATLAS Muon Spectrometer Phase-II Upgrade: Technical Design Report*, ATLAS-TDR-026; CERN-LHCC-2017-017, 2017, URL: https://cds.cern.ch/record/2285580 (cit. on p. 97).

[122] L. Guan, *Trigger Algorithms and Electronics for the ATLAS Muon New Small Wheel Upgrade*, tech. rep., CERN, 2016, URL: https://cds.cern.ch/record/2066974 (cit. on pp. 100–102).

[123] ATLAS Collaboration, *ATLAS TDAQ Phase-II Upgrade: Technical Design Report*, ATLAS-TDR-029; CERN-LHCC-2017-020, 2017, URL: https://cds.cern.ch/record/2285584 (cit. on pp. 103, 104).

[124] F. Fahim et al., *hls4ml: An Open-Source Codesign Workflow to Empower Scientific Low-Power Machine Learning Devices*, 2021, URL: https://arxiv.org/abs/2103.05579 (cit. on pp. 103, 104).

[125] R. Chalapathy and S. Chawla, *Deep Learning for Anomaly Detection: A Survey*, 2019, URL: https://arxiv.org/abs/1901.03407 (cit. on p. 114).

[126] D. Hawkins, *Identification of outliers*, Monographs on applied probability and statistics, Chapman and Hall, 1980, X, 188, ISBN: 041221900X (cit. on p. 114).

[127] M. A. Kramer, *Nonlinear principal component analysis using autoassociative neural networks*, Aiche Journal **37** (1991) 233 (cit. on p. 114).

[128] M. Rosenblatt, *Remarks on Some Nonparametric Estimates of a Density Function*, The Annals of Mathematical Statistics **27** (1956) 832, URL: https://doi.org/10.1214/aoms/1177728190 (cit. on pp. 114, 116).

[129] E. Parzen, *On Estimation of a Probability Density Function and Mode*, The Annals of Mathematical Statistics **33** (1962) 1065, URL: https://doi.org/10.1214/aoms/1177704472 (cit. on pp. 114, 116).

[130] Y. Bengio, A. C. Courville and P. Vincent, *Representation Learning: A Review and New Perspectives*, IEEE Transactions on Pattern Analysis and Machine Intelligence **35** (2013) 1798 (cit. on p. 115).

[131] I. T. Jolliffe and J. Cadima, *Principal component analysis: a review and recent developments*, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences **374** (2016) (cit. on p. 115).

[132] B. W. Silverman, *Density estimation for statistics and data analysis*, Chapman & Hall/CRC monographs on statistics and applied probability, London: Chapman and Hall, 1986, URL: https://cds.cern.ch/record/1070306 (cit. on p. 116).

[133] L. Breiman, W. S. Meisel and E. Purcell, *Variable Kernel Estimates of Multivariate Densities*, Technometrics **19** (1977) 135 (cit. on pp. 116, 129).

[134] F. Pedregosa et al., *Scikit-learn: Machine Learning in Python*, Journal of Machine Learning Research **12** (2011) 2825 (cit. on pp. 117, 129).

[135] E. H. Bareiss, *Multistep Integer-Preserving Gaussian Elimination*, (1966), URL: https://www.osti.gov/biblio/4474185 (cit. on p. 118).

[136] P. Bachmann, *Die analytische Zahlentheorie*, Zahlentheorie. Versuch einer Gesammtdarstellung dieser Wissenschaft in ihren Haupttheilen. 2. th, Teubner, 1894, URL: https://books.google.ro/books?id=0iMCAAAAYAAJ (cit. on p. 118).

[137] E. Landau, *Handbuch der Lehre von der Verteilung der Primzahlen. 1*, vol. 1, Teubner, 1909, URL: https://books.google.ro/books?id=UEfXzQEACAAJ (cit. on p. 118).

[138] V. Bogachev, A. V. Kolesnikov and K. Medvedev, *Triangular transformations of measures*, Sbornik Mathematics **196** (2004) 309 (cit. on p. 118).

[139] L. Dinh, D. Krueger and Y. Bengio, *NICE: Non-linear Independent Components Estimation*, CoRR **abs/1410.8516** (2015) (cit. on p. 120).

[140] D. P. Kingma, T. Salimans and M. Welling, *Improved Variational Inference with Inverse Autoregressive Flow*, ArXiv **abs/1606.04934** (2017) (cit. on p. 120).

[141] G. Papamakarios, I. Murray and T. Pavlakou, *Masked Autoregressive Flow for Density Estimation*, ArXiv **abs/1705.07057** (2017) (cit. on p. 120).

[142] T. K. Aarrestad et al., *The Dark Machines Anomaly Score Challenge: Benchmark Data and Model Independent Event Classification for the Large Hadron Collider*, SciPost Physics (2022) (cit. on p. 123).

[143] T. Sjöstrand et al., *An introduction to PYTHIA 8.2*, Comput. Phys. Commun. **191** (2015) 159, arXiv: 1410.3012 [hep-ph] (cit. on p. 124).

[144] M. Bähr et al., *Herwig++ physics and manual*, Eur. Phys. J. C **58** (2008) 639, arXiv: 0803.0883 [hep-ph] (cit. on p. 124).

[145] J. de Favereau et al., *DELPHES 3, A modular framework for fast simulation of a generic collider experiment*, JHEP **02** (2014) 057, arXiv: 1307.6346 [hep-ex] (cit. on p. 124).

[146] G. Kasieczka, B. Nachman and D. Shih, *R&D Dataset for LHC Olympics 2020 Anomaly Detection Challenge*, version v5, Zenodo, 2019, URL: https://doi.org/10.5281/zenodo.6466204 (cit. on p. 124).

[147] G. Kasieczka, B. Nachman and D. Shih, *Official Datasets for LHC Olympics 2020 Anomaly Detection Challenge*, version v6, Zenodo, 2019, URL: https://doi.org/10.5281/zenodo.4536624 (cit. on p. 124).

[148] B. M. Dillon, C. Han, H. M. Lee and M. Park, *KK graviton resonance and cascade decays in warped gravity*, Int. J. Mod. Phys. A **32** (2017) 1745006, arXiv: 1606.07171 [hep-ph] (cit. on pp. 125, 126).

[149] M. Cacciari, G. P. Salam and G. Soyez, *The anti-k_t jet clustering algorithm*, JHEP **04** (2008) 063, arXiv: 0802.1189 [hep-ph] (cit. on p. 126).

[150] D. I-M., *clustering-lhco*, https://github.com/imdinu/clustering-lhco, 2021 (cit. on p. 127).

[151] T. Hastie, R. Tibshirani and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer series in statistics, Springer, 2009, ISBN: 9780387848846, URL: https://books.google.ro/books?id=eBSgoAEACAAJ (cit. on p. 134).

[152] G. Kasieczka and D. Shih, *Robust Jet Classifiers through Distance Correlation*, Phys. Rev. Lett. **125** (2020) 122001, arXiv: 2001.05310 [hep-ph] (cit. on p. 135).

[153] G. Kasieczka, B. Nachman, D. Shih, O. Amram, I.-M. Dinu et al., *The LHC olympics 2020: a community challenge for anomaly detection in high energy physics*, Reports on Progress in Physics (2021) (cit. on p. 139).

[154] I.-M. Dinu, I.-S. Trandafir and C. Alexa, *A Machine Learning Based Muon Trigger Algorithm for an Assembly of Micromegas Detectors*, Romanian Journal of Physics **67** (2022) 401 (cit. on p. 139).

[155] I.-M. Dinu, 'Event-Level Anomaly Detection for Multijet BSM Searches with Probabilistic Autoencoders', *9th Large Hadron Collider Physics Conference*, 2021 340 (cit. on p. 139).

[156] I.-M. Dinu and I.-S. Trandafir, 'A Machine Learning Based Muon Trigger Algorithm for an Assembly of Micromegas Detectors', *20th International Balkan Workshop on Applied Physics and Materials Science*, oral presentation, 2022, URL: http://ibwap.ro/wp-content/uploads/2022/07/book-abstracts_IBWAP2022.pdf (cit. on p. 139).

[157] A. Chirosca, I.-M. Dinu and C. Alexa, 'First Prototype of a Machine Learning Trigger Algorithm on FPGA for Micromegas detectors', *11th International Conference of the Balkan Physical Union*, oral presentation, 2022, URL: https://indico.bpu11.info/event/1/contributions/272/ (cit. on p. 139).

[158] G. Aad et al., *Search for lepton-flavor violation in Z-boson decays with τ leptons with the ATLAS detector*, Physical review letters **127** (2021) 271801 (cit. on p. 139).

[159] G. Aad et al., *Search for new phenomena in final states with two leptons and one or no b-tagged jets at $\sqrt{s} = 13$ TeV using the ATLAS detector*, Physical review letters **127** (2021) 141801 (cit. on p. 139).

[160] G. Aad et al.,
Search for R-parity-violating supersymmetry in a final state containing leptons and many jets with the ATLAS experiment using $\sqrt{s} = 13$ TeV proton–proton collision data,
The European Physical Journal C **81** (2021) 1 (cit. on p. 139).

[161] G. Aad et al., *Search for exotic decays of the Higgs boson into long-lived particles in pp collisions at $\sqrt{s} = 13$ TeV using displaced vertices in the ATLAS inner detector*,
Journal of High Energy Physics **2021** (2021) 1 (cit. on p. 139).

[162] G. Aad et al., *Measurement of the production cross section of pairs of isolated photons in pp collisions at 13 TeV with the ATLAS detector*,
Journal of High Energy Physics **2021** (2021) 1 (cit. on p. 140).

[163] G. Aad et al., *Search for charginos and neutralinos in final states with two boosted hadronically decaying bosons and missing transverse momentum in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Physical Review D **104** (2021) 112010 (cit. on p. 140).

[164] G. Aad et al., *Measurement of the energy response of the ATLAS calorimeter to charged pions from $W^{\pm} \to \tau^{\pm}(\to \pi^{\pm}\nu_{\tau})\nu_{\tau}$ events in Run 2 data*,
The European Physical Journal C **82** (2022) 1 (cit. on p. 140).

[165] G. Aad et al.,
Search for heavy particles in the b-tagged dijet mass distribution with additional b-tagged jets in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS experiment,
Physical Review D **105** (2022) 012001 (cit. on p. 140).

[166] G. Aad et al., *Measurement of the nuclear modification factor for muons from charm and bottom hadrons in Pb+ Pb collisions at 5.02 TeV with the ATLAS detector*,
Physics Letters B **829** (2022) 137077 (cit. on p. 140).

[167] G. Aad et al., *Observation of electroweak production of two jets in association with an isolated photon and missing transverse momentum, and search for a Higgs boson decaying into invisible particles at 13 TeV with the ATLAS detector*,
The European Physical Journal C **82** (2022) 1 (cit. on p. 140).

[168] A. collaboration et al.,
Operation and performance of the ATLAS semiconductor tracker in LHC Run 2,
arXiv preprint arXiv:2109.02591 (2021) (cit. on p. 140).

[169] G. Aad et al., *Search for exotic decays of the Higgs boson into $b\bar{b}$ and missing transverse momentum in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*,
Journal of High Energy Physics **2022** (2022) 1 (cit. on p. 140).

[170] A. collaboration et al., *AtlFast3: the next generation of fast simulation in ATLAS*,
arXiv preprint arXiv:2109.02551 (2021) (cit. on p. 140).

[171] G. Aad et al., *Measurement of the energy asymmetry in $t\bar{t}j$ production at 13 TeV with the ATLAS experiment and interpretation in the SMEFT framework*,
The European Physical Journal C **82** (2022) 1 (cit. on p. 140).

[172] G. Aad et al.,
Search for Higgs bosons decaying into new spin-0 or spin-1 particles in four-lepton final states with the ATLAS detector with 139 fb- 1 of pp collision data at $\sqrt{s} = 13\,TeV$,
Journal of High Energy Physics **2022** (2022) 1 (cit. on p. 140).

[173] G. Aad et al., *Measurements of* $W^+W^- + \geq 1$ *jet production cross-sections in* $/pp$ *collisions at* $\sqrt{s} = 13\,TeV$ *with the ATLAS detector*,
Journal of High Energy Physics **2021** (2021) 1 (cit. on p. 140).

[174] G. Aad et al., *Measurement of Higgs boson decay into b-quarks in associated production with a top-quark pair in* pp *collisions at* $\sqrt{s} = 13\,TeV$ *with the ATLAS detector*, (2021) (cit. on p. 140).

[175] G. Aad et al., *Measurements of azimuthal anisotropies of jet production in Pb+ Pb collisions at* $\sqrt{s_{NN}} = 5.02\,TeV$ *with the ATLAS detector*, Physical Review C **105** (2022) 064903 (cit. on p. 141).

[176] G. Aad et al., *Search for flavour-changing neutral-current interactions of a top quark and a gluon in pp collisions at* $\sqrt{s} = 13\,TeV$ *with the ATLAS detector*,
The European Physical Journal C **82** (2022) 1 (cit. on p. 141).

[177] G. Aad et al., *Determination of the parton distribution functions of the proton using diverse ATLAS data from* pp *collisions at* $\sqrt{s} = 7, 8, 13\,TeV$,
The European Physical Journal C **82** (2022) 1 (cit. on p. 141).

[178] G. Aad et al., *Search for Higgs boson pair production in the two bottom quarks plus two photons final state in* pp *collisions at* $\sqrt{s} = 13\,TeV$ *with the ATLAS detector*, tech. rep.,
ATLAS-HDBS-2018-34-003, 2021 (cit. on p. 141).

[179] G. Aad et al., *Modelling and computational improvements to the simulation of single vector-boson plus jet processes for the ATLAS experiment*, tech. rep.,
ATLAS-PMGR-2021-01-003, 2021 (cit. on p. 141).

[180] G. Aad et al., *A search for an unexpected asymmetry in the production of* $e^+\mu^-$ *and* $e^-\mu^+$ *pairs in proton–proton collisions recorded by the ATLAS detector at* $\sqrt{s} = 13\,TeV$,
Physics Letters B (2022) 137106 (cit. on p. 141).

[181] G. Aad et al., *Search for Higgs boson decays into a pair of pseudoscalar particles in the* $bb\mu\mu$ *final state with the ATLAS detector in pp collisions at* $\sqrt{s} = 13\,TeV$,
Physical Review D **105** (2022) 012006 (cit. on p. 141).

[182] G. Aad et al., *Measurement of the* $t\bar{t}t\bar{t}$ *production cross section in pp collisions at* $\sqrt{s} = 13\,TeV$ *with the ATLAS detector*, Journal of High Energy Physics **2021** (2021) 1 (cit. on p. 141).

[183] G. Aad et al., *Search for long-lived charginos based on a disappearing-track signature using* 136fb^{-1} *of pp collisions at* $\sqrt{s} = 13\,TeV$ *with the ATLAS detector*, tech. rep.,
ATLAS-SUSY-2018-19-003, 2022 (cit. on p. 141).

[184] G. Aad et al., *Search for single production of a vectorlike T quark decaying into a Higgs boson and top quark with fully hadronic final states using the ATLAS detector*, Physical Review D **105** (2022) 092012 (cit. on p. 141).

[185] G. Aad et al., *Direct constraint on the Higgs-charm coupling from a search for Higgs boson decays into charm quarks with the ATLAS detector*, tech. rep., ATLAS-HIGG-2021-12-003, 2022 (cit. on p. 141).

[186] G. Aad et al., *Measurement of the c-jet mistagging efficiency in $t\bar{t}$ events using pp collision data at $\sqrt{s} = 13$ TeV collected with the ATLAS detector*, The European Physical Journal C **82** (2022) 1 (cit. on p. 141).

[187] G. Aad et al., *Search for type-III seesaw heavy leptons in leptonic final states in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, tech. rep., ATLAS-EXOT-2020-02-003, 2022 (cit. on p. 141).

[188] G. Aad et al., *Measurements of the Higgs boson inclusive and differential fiducial cross-sections in the diphoton decay channel with pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, tech. rep., ATLAS-HIGG-2019-13-003, 2022 (cit. on p. 142).

[189] G. Aad et al., *Observation of WWW production in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, tech. rep., ATLAS-STDM-2019-09-003, 2022 (cit. on p. 142).

[190] G. Aad et al., *Search for invisible Higgs-boson decays in events with vector-boson fusion signatures using* 139fb^{-1} *of proton-proton data recorded by the ATLAS experiment*, tech. rep., ATLAS-EXOT-2020-11-003, 2022 (cit. on p. 142).

[191] G. Aad et al., *Measurement of the polarisation of single top quarks and antiquarks produced in the t-channel at $\sqrt{s} = 13$ TeV and bounds on the tWB dipole operator from the ATLAS experiment*, tech. rep., ATLAS-TOPQ-2018-10-003, 2022 (cit. on p. 142).

[192] G. Aad et al., *Measurements of differential cross-sections in top-quark pair events with a high transverse momentum top quark and limits on beyond the Standard Model contributions to top-quark pair production with the ATLAS detector at $\sqrt{s} = 13$ TeV*, tech. rep., ATLAS-TOPQ-2019-23-003, 2022 (cit. on p. 142).

[193] G. Aad et al., *Search for neutral long-lived particles in pp collisions at $\sqrt{s} = 13$ TeV that decay into displaced hadronic jets in the ATLAS calorimeter*, Journal of High Energy Physics **2022** (2022) 1 (cit. on p. 142).

[194] G. Aad et al., *Search for events with a pair of displaced vertices from long-lived neutral particles decaying into hadronic jets in the ATLAS muon spectrometer in pp collisions at $\sqrt{s} = 13$ TeV*, tech. rep., ATLAS-EXOT-2019-24-003, 2022 (cit. on p. 142).

[195] G. Aad et al., *Search for associated production of a Z boson with an invisibly decaying Higgs boson or dark matter candidates at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Physics Letters B **829** (2022) 137066 (cit. on p. 142).

[196] G. Aad et al., *Constraints on Higgs boson production with large transverse momentum using $H \to b\bar{b}$ decays in the ATLAS detector*, Physical Review D **105** (2022) 092003 (cit. on p. 142).

[197] G. Aad et al., *Search for resonant pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state using pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Physical Review D **105** (2022) 092002 (cit. on p. 142).

[198] G. Aad et al., *Study of $B_c^+ \to J/\psi D_s^+$ and $B_c^+ \to J/\psi D_s^{*+}$ decays in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Journal of High Energy Physics **2022** (2022) 1 (cit. on p. 142).

[199] G. Aad et al., *Measurements of Higgs boson production cross-sections in the $H \to \tau^+\tau^-$ decay channel in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Journal of High Energy Physics **2022** (2022) 1 (cit. on p. 142).

[200] C. E. Shannon, *A mathematical theory of communication*, Bell Syst. Tech. J. **27** (2001) 623 (cit. on p. 167).

[201] S. Kullback and R. A. Leibler, *On Information and Sufficiency*, The Annals of Mathematical Statistics **22** (1951) 79, URL: https://doi.org/10.1214/aoms/1177729694 (cit. on p. 168).

[202] B. J. Holzer, *Beam optics and lattice design for particle accelerators*, en, 2013, URL: http://cds.cern.ch/record/1533020 (cit. on p. 172).

List of Figures

List of Tables