

Université de Montréal

**La découverte de nouvelle physique à l'aide de la
détection d'anomalies avec l'apprentissage automatique
au Grand collisionneur de hadrons**

par

Julien Leissner-Martin

Département de physique
Faculté des arts et des sciences

Mémoire présenté en vue de l'obtention du grade de
Maître ès sciences (M.Sc.)
en Physique

Le 31 décembre 2023

Université de Montréal

Faculté des arts et des sciences

Ce mémoire intitulé

**La découverte de nouvelle physique à l'aide de
la détection d'anomalies avec l'apprentissage
automatique au Grand collisionneur de hadrons**

présenté par

Julien Leissner-Martin

a été évalué par un jury composé des personnes suivantes :

Jean-François Carrier

(président-rapporteur)

Jean-François Arguin

(directeur de recherche)

Georges Azuelos

(membre du jury)

Résumé

La physique des particules est une branche de la science qui est actuellement régie sous un ensemble de lois nommé le *modèle standard* (MS). Il dicte notamment quelles particules existent et comment elles interagissent entre elles. Il permet de prédire toutes sortes de résultats qui sont constamment testés et confirmés par une multitude d'expériences, dont l'expérience ATLAS, au Grand Collisionneur de Hadrons (LHC). Toutefois, ce modèle hautement précis et juste ne peut décrire qu'environ 5% de la matière de l'Univers et s'avère donc incomplet. Les scientifiques passent au peigne fin plusieurs études pour y retrouver de la nouvelle physique, mais en vain.

Les théoriciens ne sont pas en reste non plus, et ont concocté plusieurs théories pouvant être vues comme des extensions du modèle standard. Malheureusement, plus de dix ans après la découverte du boson de Higgs au LHC qui venait confirmer la théorie actuelle du MS, aucun signe de ces extensions n'a pu être trouvé. Nous proposons dans ce mémoire d'utiliser l'intelligence artificielle pour aider à trouver certains indices de nouvelle physique.

Pour ce faire, nous entraînerons des modèles d'apprentissage automatique (*machine learning*) à reconnaître des signes de la nouvelle physique dans des données réelles ou simulées issues de collisions proton-proton au sein du détecteur ATLAS. Ce détecteur oeuvre au LHC, le plus grand collisionneur au monde, où nos données proviennent d'énergies de centre de masse de 13 TeV. Nous utiliserons les quadrivecteurs des particules contenues dans les jets boostés à grand rayon, des amas collimatés de particules présents dans ATLAS, qui pourraient contenir cette fameuse nouvelle physique. Dans ce mémoire, nous tenterons entre autres de retrouver des signaux de quarks top ainsi que de particules hypothétiques issues d'un modèle avec un secteur étendu du boson de Higgs.

Actuellement, nos modèles sont capables de bien distinguer le signal du bruit de fond. Toutefois, les résultats sont corrélés à la masse des jets et toute tentative pour contrecarrer cette corrélation diminue de beaucoup la discrimination du signal et du bruit de fond. De plus, nous devons améliorer le rejet du bruit de fond pour espérer retrouver de la nouvelle

physique dans les données d'ATLAS.

Mots-clés : physique des particules, LHC, Grand collisionneur de hadrons, ATLAS, CERN, intelligence artificielle, apprentissage automatique, réseau de neurones, auto-encodeur variationnel, anomalies, jet boosté, jet à grand rayon

Abstract

Particle physics is currently governed by a set of laws called the Standard Model. This model notably includes which particles exist and how they interact with one another. It also allows the prediction of many results which are constantly tested and confirmed by all kinds of experiments, like the ATLAS experiment at the Large Hadron Collider (LHC). However, this highly precise model can only describe 5% of the Universe, so it is incomplete. Scientists across the globe analyzed all kinds of data to find new physics, but to no avail.

Theorists also aren't resting, and have concocted many new theories which can be seen as Standard Model extensions. Unfortunately, more than ten years after the discovery of the Higgs boson at LHC that confirmed the last bits of the Standard Model, no signs of these extensions have been found. In light of this, we propose to use artificial intelligence to help us find signs of new physics.

To perform this task, we will train machine learning models to recognize signs of new physics inside real or simulated data originating from proton-proton collisions in the ATLAS detector. This detector operates at LHC, the biggest particle collider in the world, where our data will come from center-of-mass energies of 13 TeV. We will use four-vectors of particles contained within large radius and boosted jets, which are dense streams of particles present in ATLAS and where new physics might hide. In this dissertation, we will notably try to find signals of top quarks and hypothetical particles originating from a model with an extended Higgs boson sector.

Currently, our models are able to distinguish between signal and background noise. However, these results are heavily correlated to jet mass, and any attempt at diminishing this correlation yields worse discriminating power between signal and background. We will also need to improve the background rejection to hope find new physics in the ATLAS data.

Keywords : particle physics, LHC, ATLAS, CERN, artificial intelligence, deep learning, neural network, variational autoencoder, anomaly, boosted jet, large radius jet

Table des matières

Résumé	5
Abstract	7
Liste des tableaux	13
Liste des figures	15
Liste des sigles et des abréviations	19
Remerciements	23
Introduction	25
Chapitre 1. Modèle Standard	29
1.1. Les particules	29
1.2. Invariance de jauge	32
1.3. Les forces	33
1.3.1. La force électromagnétique	34
1.3.2. Interactions et conséquences connues de la QED	35
1.3.3. La force nucléaire faible	36
1.3.4. La théorie électrofaible et le mécanisme de Higgs	38
1.3.5. La force nucléaire forte	40
1.3.6. Les jets	41
1.4. La physique au-delà du modèle standard	44
1.4.1. Les lacunes du modèle standard	45
1.4.2. Les théories au-delà du modèle standard	47
1.5. Discussion	50
Chapitre 2. Le Grand collisionneur de hadrons (LHC) et l'expérience ATLAS	55

2.1.	Le Grand Collisionneur de Hadrons	56
2.1.1.	L'accélération des particules	57
2.1.2.	La luminosité	59
2.1.3.	Les expériences	61
2.2.	Le détecteur ATLAS	62
2.2.1.	Les systèmes de coordonnées du détecteur	62
2.2.2.	Le détecteur interne	64
2.2.3.	Les calorimètres	66
2.2.4.	Le spectromètre à muons	68
2.3.	Le système d'acquisition des données	69
2.3.1.	Déclencheur de premier niveau	70
2.3.2.	Déclencheur de haut niveau	70
2.4.	La reconstruction d'événements	71
2.4.1.	Les jets hadroniques	71
2.5.	La simulation d'événements	75
2.5.1.	Les sous-étapes de la simulation	76
2.6.	Conclusion	79
Chapitre 3.	L'apprentissage automatique	81
3.1.	Les réseaux de neurones	82
3.1.1.	Le flux d'information du réseau de neurones	82
3.2.	L'entraînement des réseaux de neurones	86
3.2.1.	La fonction de perte et la descente de gradient	86
3.2.2.	Le surentraînement et les données d'entraînement, de validation et de test	89
3.2.3.	L'entraînement supervisé et l'entraînement non supervisé	92
3.3.	L'auto-encodeur	93
3.3.1.	L'auto-encodeur variationnel (VAE)	94
Chapitre 4.	La détection d'anomalies au Grand Collisionneur de Hadrons	101
4.1.	Méthodes d'analyse	102
4.2.	Description de l'architecture du VAE	105
4.2.1.	Le VAE sans exposition aux données hors distribution	106

4.2.2. Le VAE avec exposition aux données hors distribution (OE-VAE)	107
4.3. Description des échantillons	110
4.3.1. Le bruit de fond	110
4.3.2. Les signaux	110
4.3.3. Les données hors distribution	112
4.4. Résultats du OE-VAE	119
4.4.1. Avec les données hors distribution $H^\pm \rightarrow jj$	120
4.4.2. Avec les données hors distribution $H^\pm \rightarrow tb$	126
4.5. Discussion	128
4.6. Voies d'amélioration possibles et futur du projet	129
Conclusion	131
Références bibliographiques	133

Liste des tableaux

- 3.1 Fonctions d'activation ainsi que leurs dérivées fréquemment utilisées dans les NN 88
- 4.1 Liste des hyperparamètres communs aux deux modèles entraînés et évalués. 120

Liste des figures

1.1	Particules élémentaires du modèle standard.	30
1.2	Vertex primaire de la force électromagnétique.	32
1.3	Diagrammes de Feynmann de premier ordre d'interactions connues de la QED... 36	36
1.4	Vertex de base de la force faible impliquant des fermions.	37
1.5	Vertex de base de la force faible n'impliquant que des bosons et des photons.	37
1.6	Les trois vertex de base de la QCD.	41
1.7	Magnitude de la constante de couplage de la force forte, α_s , en fonction de l'énergie. 42	42
1.8	Visualisation de la formation de cascades de quarks et de l'hadronisation.	43
1.9	Comparaison entre un jet non boosté et un jet boosté dans un collisionneur de particules.	44
1.10	Diagramme le plus simple d'un boson de Higgs entrecoupé d'une boucle de quarks top.	46
1.11	Limites sur les masses des particules supersymétriques hypothétiques données par l'expérience ATLAS.	51
1.12	Limites sur les masses des nouvelles particules hypothétiques données par l'expérience ATLAS.	52
1.13	Limites sur les temps de vie des nouvelles particules hypothétiques données par l'expérience ATLAS.	53
2.1	Vue à vol d'oiseau du complexe du CERN, situé à la frontière entre la Suisse et la France, à Genève.	56
2.2	Schéma du complexe d'accélérateurs au CERN.	59
2.3	Luminosité intégrée selon le mois de l'année, année après année (gauche) et luminosité intégrée cumulative du LHC sur sa deuxième période d'opération <i>Run 2</i> (droite).	61
2.4	Les systèmes de coordonnées au sein du détecteur ATLAS.	64
2.5	Schéma du détecteur interne, qui constitue la première partie du détecteur ATLAS. 66	66

2.6	Schéma des calorimètres, qui constituent la deuxième partie du détecteur ATLAS.	68
2.7	Schéma du spectromètre à muons, qui constitue la partie finale du détecteur ATLAS.	69
2.8	Formation des amas topologiques dans le calorimètre hadronique avant (FCAL) servant à la reconstruction des jets.	72
2.9	La reconstruction d'un même jet par différents algorithmes.	74
2.10	Fonctions de distribution des partons pour une énergie du proton de 2 GeV (gauche) et de 100 GeV (droite).	78
3.1	Schéma du flux d'information dans un réseau de neurones.	83
3.2	Fonctions d'activation fréquemment utilisées dans les NN.	84
3.3	Visualisation du sous-apprentissage, surapprentissage et du « bon » apprentissage.	90
3.4	Visualisation de la progression typique de la fonction de perte sur les données d'entraînement et les données de validation en fonction du nombre d'époques.	91
3.5	Architecture générale d'un auto-encodeur.	94
3.6	Comparaison des espaces latents représentés en deux dimensions d'un auto-encodeur traditionnel et d'un auto-encodeur variationnel.	96
3.7	Architecture standard d'un VAE.	97
4.1	Illustration du principe de recherche de bosses.	103
4.2	Diagrammes de Feynman représentant les processus des jets hors distribution générés à deux et quatre sous-jets.	113
4.3	Comparaison de la distribution des événements $H^\pm \rightarrow tb$ dans l'espace de phase 2D de masse et p_T et de la prédiction théorique pour la configuration de masse $\{m_H = 5000 \text{ GeV}, m_{H=\pm} = 1000 \text{ GeV}\}$.	115
4.4	Masse et p_T attendus des différentes configurations de masse sélectionnées des jets hors distribution générés. Les événements les moins massifs sont de 200 GeV en raison de la masse de l'état final de plus de 175 GeV.	116
4.5	Histogramme en deux dimensions de la masse et du p_T des jets conservés lors de la génération des événements hors distributions.	118
4.6	Résultats obtenus pour l'entraînement effectué sur les jets hors distributions $H^\pm \rightarrow jj$ et une évaluation sur le signal du Z' avec pic à 200 GeV selon la masse du top choisie.	122

4.7	Résultats obtenus pour l'entraînement effectué sur les jets hors distributions $H^\pm \rightarrow jj$ et une évaluation sur le signal du Z' avec pic à 200 GeV selon la masse du top choisie (suite).....	123
4.8	Résultats obtenus pour l'entraînement effectué sur les jets hors distributions $H^\pm \rightarrow jj$ et une évaluation sur le signal du Z' où la masse du top et du signal est de 500 GeV.....	124
4.9	Résultats obtenus pour l'entraînement effectué sur les jets hors distributions $H^\pm \rightarrow jj$ et une évaluation sur le signal du Z' où la masse du top et du signal est de 500 GeV (suite).....	125
4.10	Résultats obtenus pour l'entraînement effectué sur les jets hors distributions $H^\pm \rightarrow tb$ et le signal du Z' à 500 GeV en vertu de la masse du top sélectionnée. .	126
4.11	Résultats obtenus pour l'entraînement effectué sur les jets hors distributions $H^\pm \rightarrow tb$ et le signal du Z' à 500 GeV en vertu de la masse du top sélectionnée (suite).....	127

Liste des sigles et des abréviations

AA	Apprentissage automatique
AE	Auto-encodeur
ALICE	Une expérience d'un large collisionneur d'ions, de l'anglais <i>A Large Ion Collider Experiment</i>
ATLAS	Un appareil toroïdal du LHC, de l'anglais <i>A Toroidal LHC Apparatus</i>
CERN	Conseil Européen pour la Recherche Nucléaire
CMS	Solénoïde compact à muons, de l'anglais <i>Compact Muon Solenoid</i>
HLV	Variables de haut niveau, de l'anglais <i>High-Level Variables</i>
IA	Intelligence artificielle
KLD	Divergence de Kullback-Leibler, de l'anglais <i>Kullback–Leibler divergence</i>

LHC	Grand Collisionneur de Hadrons, de l'anglais <i>Large Hadron Collider</i>
MLP	Perceptron multicouche, de l'anglais <i>multilayer perceptron</i>
MS	Modèle Standard
MSE	Erreur quadratique moyenne, de l'anglais <i>Mean Square Error</i>
NN	Réseau de neurones, de l'anglais <i>Neural Network</i>
OE	Exposition aux données hors distribution, de l'anglais <i>Outlier Exposure</i>
OE-VAE	Auto-encodeur variationnel avec exposition aux données hors distribution, de l'anglais <i>Outlier Exposure Variational Autoencoder</i>
OoD	Hors distribution, de l'anglais <i>Out of Distribution</i>
PDF	Fonction de distribution des partons, de l'anglais <i>Parton Density Function</i>
QCD	Chromodynamique quantique, de l'anglais <i>Quantum Chromodynamics</i>
VAE	Auto-encodeur variationnel, de l'anglais <i>Variational Autoencoder</i>

vev

Valeur attendue du vide, de l'anglais *vacuum expectancy value*

Remerciements

Je tiens à commencer par remercier Jean-François, mon directeur de recherche. Depuis que j'ai commencé à travailler avec toi en 2019, tu as fait preuve d'ouverture pour mes idées par rapport au projet de recherche et à mes questions. Tu as toujours été à l'écoute, présent et bienveillant. Je souhaite aussi remercier tous les collègues que j'ai eus dans mon parcours, notamment Dominique et Nghia, sans qui le projet n'aurait pas pu progresser. Merci d'avoir été présents pour m'épauler et pour partager vos connaissances.

Je remercie ensuite mes amis pour toutes ces soirées magnifiques qui se répètent depuis plus de dix ans. J'espère qu'elles se continueront sur plusieurs décennies encore. Vous avez pris une place importante dans ma vie depuis tout ce temps et j'en suis reconnaissant.

Je tiens à remercier ma belle-famille. Vous êtes généreux et bons. Je constate toute la chance que j'ai eue de vous rencontrer. Je salue aussi mon neveu et ma nièce qui débordent de vie et qui trouvent une façon de nous faire rire et sourire à tout coup.

Évidemment, j'offre toute ma reconnaissance et tous mes remerciements pour ma famille, qui a toujours été un pilier incroyable sur lequel j'ai pu m'appuyer. Je tiens à remercier ma soeur, qui a été une mentor pour moi toute ma vie et sur qui j'ai si souvent pris exemple. Je remercie aussi son mari et mon beau-frère d'être présent lui aussi depuis plus de dix ans dans ma vie. Je ne compte plus les fous rires et les soirées passées avec vous deux. Finalement, merci à mes parents, qui m'ont tout donné et tout appris. Je ne serais pas l'ombre de la personne que je suis aujourd'hui sans vous. Je suis conscient du privilège que j'ai d'avoir toujours pu compter sur vous. Vous n'êtes pas étrangers à ma réussite scolaire.

Finalement, merci à Angélique, ma copine et partenaire de vie. Merci d'avoir relu mon mémoire pour me conseiller. Merci d'être présente et d'être la personne que tu es, merci de m'épauler dans tous mes projets et de toute la compréhension dont tu fais preuve. Merci de partager avec moi tous nos plans. Nous formons la plus belle des équipes, ce qui m'a tant

aidé à la réalisation de ce mémoire.

Je veux rendre un dernier hommage à mon grand-père, un homme généreux et altruiste qui a été la source de la passion pour les sciences dans la famille. Merci d'avoir toujours pensé aux autres avant toi-même et d'avoir été si bon. Tu vas nous manquer énormément.

Introduction

Depuis la découverte du boson de Higgs au LHC en 2012, qui avait déjà été théorisé en 1964, la physique des particules s'est plus que jamais tournée vers la découverte de la nouvelle physique. En effet, bien que la théorie actuelle régnant sur la physique des particules, le modèle standard, ait été hautement confirmée depuis des décennies, elle demeure incomplète. Par exemple, bien que ce modèle décrive avec justesse la force forte, la force faible et la force électromagnétique, il ne peut expliquer la gravité. De plus, il n'inclut que la matière dite ordinaire qui nous entoure – qui ne représente que 5% de l'Univers –, mais n'inclut notamment pas la matière sombre.

De nombreux modèles théoriques ont été développés au fil du temps pour combler les lacunes du modèle standard. Par exemple, la plus populaire, la supersymétrie, propose que chaque particule connue posséderait aussi un « superpartenaire », c'est-à-dire une particule analogue supersymétrique. Ainsi, chaque boson est associé à un fermion, dénommé *bosino* et chaque fermion est associé à un boson (*sfermion*). Par contre, après des années de recherche à trouver des signes confirmant l'un de ces modèles, aucun n'a pu être trouvé. Face à cette absence de résultats, l'intelligence artificielle devient une alternative novatrice qui pourrait accélérer, voire débloquent, la recherche de cette nouvelle physique. L'utilisation toujours grandissante de cet outil au LHC témoigne de sa puissance : par exemple, il peut être utilisé pour la reconstruction ou l'identification d'événements. L'avantage principal de l'intelligence artificielle dans ce contexte est qu'elle ne demande pas de confirmer des modèles théoriques déjà instaurés. Elle permet plutôt de faire une recherche large et globale dans les données du LHC, nous permettant de chercher partout plutôt que seulement dans les endroits pointés par la théorie. De cette manière, nous enlevons la dépendance aux théories pré-existantes, ce qui comporte de nombreux bienfaits. Tout d'abord, nous avons le potentiel de devancer les avancées théoriques et de trouver des signes de nouvelle physique jamais auparavant imaginée. Cela signifie simultanément que nous pouvons progresser même si tous les modèles déjà concoctés s'avèrent à être faux. Finalement, l'intelligence artificielle nous permet de potentiellement sauver du temps, puisque de nombreuses années

de recherche théoriques sur des nouveaux modèles peuvent être évitées.

Évidemment, ce mémoire propose d'utiliser l'intelligence artificielle pour trouver des signes de la nouvelle physique en faisant le moins d'hypothèses initiales possibles sur la forme qu'elle pourrait prendre. Pour ce faire, nous entraînerons un auto encodeur variationnel, soit un type de réseau de neurones, à détecter de la nouvelle physique. Évidemment, faire une recherche complètement indépendante de modèles existants et d'hypothèses serait impossible. Nous nous concentrerons alors sur les jets boostés, des amas collimatés de particules à haute énergie, qui sont pressentis pour être des candidats idéaux à la découverte de nouvelle physique. Notre méthode sera d'entraîner le réseau sur des jets issus du modèle standard dans le but qu'il en comprenne la structure et la sous-structure. Puis, lorsque le modèle sera appliqué sur des données anormales provenant de la nouvelle physique, il devrait être en mesure de les détecter puisqu'elles ne devraient pas posséder une sous-structure qu'il connaît. Notre but est donc qu'au terme de ce projet, nous puissions appliquer le modèle retenu sur les jets boostés contenus dans les données d'ATLAS, un des quatre détecteurs de particules du LHC et qui représente l'expérience sur laquelle nous travaillons. Puis, en analysant les données identifiées comme les plus anormales, nous devrions nous retrouver avec un échantillon beaucoup plus concentré en données issues de la nouvelle physique qu'en analysant l'ensemble des jets boostés détectés par ATLAS.

Toutefois, nous en sommes encore à la phase de recherche et de développement du projet. Pour l'instant, nous testons notre idée en entraînant des réseaux de neurones sur des données simulées de jets boostés provenant de gluons et de quarks légers du modèle standard. Puis, nous simulons un environnement de découverte de nouvelle physique en appliquant les modèles à divers signaux de test, qui cachent au sein de l'événement une nouvelle particule inconnue du réseau de neurones qu'il devrait être en mesure d'identifier.

Ce mémoire est divisé en quatre chapitres qui présentent divers aspects du projet. Le premier aborde les notions théoriques du modèle standard, où nous discuterons de ce qu'il décrit ainsi que de ses lacunes. Le deuxième chapitre présente l'appareillage du LHC et d'ATLAS, de la formation et de la reconstruction des jets dans ATLAS et de la simulation de données par logiciel. Le troisième est axé sur la description de l'apprentissage automatique (*machine learning*), la branche de l'intelligence artificielle utilisée dans ce mémoire. Finalement, le dernier chapitre se concentre sur notre projet de recherche, discute des motivations ayant poussé notre groupe à faire divers choix en plus de présenter les résultats obtenus par nos réseaux de neurones.

Notons finalement qu'un accent important sera mis à la section 4.3.3 sur la génération de particules dites hors distribution, cruciales à notre projet et discutées au chapitre 4. Cette génération a représenté une de mes contributions personnelles les plus importantes au projet. Parmi mes autres contributions notables figurent aussi l'étude des données hors distribution sur la performance du modèle (voir section 4.2.2) et l'optimisation de β et de λ , deux hyperparamètres du modèle (voir section 4.4). J'ai aussi activement participé à l'entraînement et l'évaluation de divers modèles dans le but d'évaluer l'impact de toutes sortes de changements que nous avons effectués au cours de ce projet, que ce soit des changements d'hyperparamètres, d'architecture du modèle, de changement d'ensembles de données ou de toute autre modification apportée au projet. J'ai aussi participé aux décisions à prendre pour optimiser nos modèles suite à ces études ainsi qu'à émettre des hypothèses pour comprendre de quelle manière certains paramètres impactaient les modèles.

Chapitre 1

Modèle Standard

Le but de ce mémoire est de trouver de la nouvelle physique dans le domaine de la physique des particules. Or, cette branche de la physique est déjà régie par un ensemble de lois extrêmement robuste connu sous le nom du Modèle Standard (MS). Cette théorie décrit les particules fondamentales ainsi que leurs forces et interactions les unes envers les autres. Ce chapitre présentera d'abord le MS tel qu'il est actuellement connu avant de montrer son incomplétude et ses lacunes pour expliquer certains phénomènes observés mais non inclus ou inexacts dans le MS. Puis, un bref survol des principales théories permettant de compléter partiellement le modèle standard ainsi que l'état actuel des recherches pour celles-ci sera effectué.

1.1. Les particules

Ce projet de recherche tente d'analyser les données du LHC pour y trouver des anomalies (c'est-à-dire des phénomènes difficilement explicables avec la théorie acceptée du MS) et ainsi de la nouvelle physique. Pour y parvenir, il faut bien comprendre la nature et le comportement des particules connues qui se retrouvent dans ces données. De plus, la nouvelle physique a de très bonnes chances de se manifester sous la forme de nouvelles particules encore inconnues pour nous. Pour cette raison, les particules sont au centre de ce projet, et une définition ainsi qu'un survol des particules connues s'impose. Celles-ci sont illustrées à la figure 1.1.

Tout d'abord, les particules sont classées via leur type, que ce soit un quark, un lepton, un boson de jauge ou un boson scalaire. Les *fermions*, qui englobent les quarks ainsi que les leptons, sont ensuite catégorisés en trois *générations* situées dans les trois premières colonnes de la figure 1.1. Les fermions de troisième génération sont beaucoup plus massifs que ceux de deuxième génération qui sont à leur tour beaucoup plus massifs que ceux de la

Standard Model of Elementary Particles

three generations of matter (fermions)			interactions / force carriers (bosons)	
	I	II	III	
mass	$\approx 2.2 \text{ MeV}/c^2$	$\approx 1.28 \text{ GeV}/c^2$	$\approx 173.1 \text{ GeV}/c^2$	0
charge	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	0
spin	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1
	u up	c charm	t top	g gluon
	d down	s strange	b bottom	γ photon
	e electron	μ muon	τ tau	Z Z boson
	ν_e electron neutrino	ν_μ muon neutrino	ν_τ tau neutrino	W W boson
	$\approx 4.7 \text{ MeV}/c^2$	$\approx 96 \text{ MeV}/c^2$	$\approx 4.18 \text{ GeV}/c^2$	$\approx 91.19 \text{ GeV}/c^2$
	$-\frac{1}{3}$	$-\frac{1}{3}$	$-\frac{1}{3}$	0
	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1
	H higgs			
	$\approx 0.511 \text{ MeV}/c^2$	$\approx 105.66 \text{ MeV}/c^2$	$\approx 1.7768 \text{ GeV}/c^2$	$\approx 80.360 \text{ GeV}/c^2$
	-1	-1	-1	± 1
	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	0
				1

Fig. 1.1. Particules élémentaires du modèle standard. Les particules sont regroupées colonne par colonne par la génération de fermions, puis la quatrième colonne représente les bosons de jauge, avec le boson de Higgs (le seul boson scalaire) dans la cinquième colonne. Les différentes couleurs représentent les différents types de particules, soient les quarks, les leptons, les bosons de jauge le boson de Higgs. La masse, la charge électronique ainsi que le spin sont indiquées pour chaque particule [62].

première, à l'exception potentielle des neutrinos.¹ Seule la première génération de fermions est stable et ainsi, les autres types de fermions se désintègrent presque instantanément vers des fermions de générations inférieures jusqu'à atteindre une particule stable de première génération. Finalement, chaque fermion est associé à une antiparticule de même masse, de même spin mais de certains nombres quantiques inversés, ce qui résulte entre autres en une charge électrique inversée.

Dans le secteur des quarks, il existe deux particules par génération. Les particules de la première rangée, dont font partie les quarks *up* (*u*), *charm* (*c*) et *top* (*t*), possèdent une charge électrique de $\frac{2}{3}$, tandis que ceux de la deuxième rangée, soient le *down* (*d*), le *strange* (*s*) et le *bottom* (*b*) ont une charge de $-\frac{1}{3}$. Les quarks ont la particularité

¹ En effet, les neutrinos sont si légers et interagissent si peu avec la matière ordinaire que leur masse n'est pas précisément connue à ce jour.

de ne posséder pas seulement une charge électrique, mais aussi une charge de couleur, qui se traduit en différents états quantiques atteignables par les quarks. Chaque quark possède une couleur parmi les trois possibles, qui sont le rouge, le bleu et le vert tandis que chaque antiquark possède une anticouleur, soient l'antirouge, l'antibleu et l'antivert. Cette couleur leur permet entre autres d'interagir avec les gluons (voir plus tard dans ce texte) et donc d'interagir avec la force nucléaire forte. Ainsi, en raison de leur charge électrique non nulle et de leur couleur, les quarks peuvent interagir avec la force forte, la force électromagnétique et la force faible. Finalement, les quarks ne peuvent jamais être libres et doivent plutôt former des états liés pour créer d'autres particules de couleur neutre nommées *hadrons*. Une couleur neutre est atteinte par une combinaison des trois couleurs, des trois anticouleurs ou d'une couleur et son anticouleur associée. C'est ce qu'on appelle le principe de confinement des quarks. Ainsi, ils peuvent se lier deux par deux en paires quark-antiquark, formant alors des *mésons*, comme par exemple les pions ($\pi^+ = u\bar{d}$, $\pi^0 = u\bar{u}/d\bar{d}$, $\pi^- = \bar{u}d$). Les quarks sont aussi les constituants des *baryons*, qui regroupent trois quarks ou antiquarks comme par exemple le très connu proton ($p = uud$).

Du côté des leptons, les neutrinos électronique (ν_e), muonique (ν_μ) et tauique (ν_τ) possèdent une charge électrique nulle et n'interagissent que via la force faible. Les neutrinos n'interagissent donc que très peu avec la matière. Par exemple, une année-lumière de plomb n'arrêtera qu'environ 50% des neutrinos la traversant [117]. De leur côté, l'électron (e^-), le muon (μ^-) et le tau (τ^-) ont une charge électrique de -1 . Ces trois derniers interagissent via les forces électromagnétique et nucléaire faible sous échange de photons ou de bosons W ou Z .

Enfin, la quatrième colonne de la figure 1.1 représente les bosons de jauge, qui ont un spin de 1. Ils sont au coeur des interactions entre les particules, puisque c'est via l'échange d'un boson de jauge qu'elles interagissent. Les différents bosons proposent des interactions pour différents types de charges. Tout d'abord, le gluon est de charge électrique et de masse nulles, se dérive en huit états de couleurs différentes et *couple*, c'est-à-dire interagit, avec toute particule possédant une couleur, mais pas les autres. C'est donc dire qu'il n'interagit qu'avec lui-même et avec les quarks. Ces interactions sont reconnues comme étant la *force nucléaire forte* (voir section 1.3.5).

Ensuite, le photon est responsable de la *force électromagnétique* (voir section 1.3.1), qui interagit avec toutes les particules chargées électriquement. Ce dernier est toutefois électriquement neutre et sans masse. Finalement, le boson Z , électriquement neutre, et les bosons W^\pm , de charge ± 1 , possèdent une masse. Leur échange mène à l'interaction via la *force faible*, qui est la force la moins puissante des trois. Ces trois bosons sont appelés

bosons intermédiaires et interagissent avec toutes les particules connues du modèle standard.

En dernier lieu, le boson de Higgs fait classe à part et possède un spin de 0. Contrairement aux autres bosons, il possède une parité positive ($P = 1$), c'est-à-dire que son comportement est identique sous inversion de l'espace. Le boson de Higgs représente des ondulations du champ scalaire de Higgs. C'est l'interaction entre les bosons et fermions intermédiaires et ce champ scalaire qui leur permet d'acquérir une masse, incluant le boson de Higgs lui-même (voir section 1.3.4). Le boson de Higgs est célèbre pour avoir été théorisé en 1964, mais observé expérimentalement au LHC seulement en 2012.

1.2. Invariance de jauge

La théorie du modèle standard est basée sur les notions mathématiques de symétrie et d'invariance de jauge [6]. Les interactions entre les particules sont régies par un Lagrangien qui, similairement aux théories de mécanique classique, décrit le mouvement des particules. Le Lagrangien des différentes forces décrit entre autres les *vertex primaires* ou *vertex de base*. Il s'agit de petits schémas représentant des interactions possibles entre les différentes particules pouvant interagir avec cette force, voir par exemple les figures 1.2 et 1.6. Ces schémas peuvent alors être assemblés tels des blocs Lego pour créer ce qu'on appelle un *diagramme de Feynmann* [124]. Tout diagramme pouvant être construit avec au moins deux vertex primaires des différentes forces représente un phénomène possible de la nature. Nous verrons ces vertex plus en détail et leur lien avec le Lagrangien tout au long de la section 1.3.

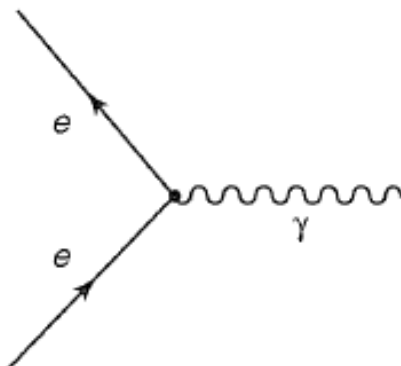


Fig. 1.2. Vertex primaire de la force électromagnétique. Ce vertex permet l'interaction entre deux fermions chargés et un photon. Cette figure démontre le cas où les deux fermions sont des électrons, mais il peut s'agir de n'importe quel fermion chargé [77, Chapitre 2].

Toutefois, le Lagrangien des particules soumises aux différentes forces fondamentales est invariant sous transformation de jauge. C'est de cette invariance que naît la symétrie du système, qui est elle-même représentée par un *groupe de symétrie*. Comme nous le verrons tout au long de ce chapitre, chaque force est décrite par un groupe de symétrie qui varie d'une force à l'autre. Ces différents groupes définissent toutes les transformations dont le Lagrangien est invariant.

Intéressons-nous maintenant à la définition d'une transformation correspondant à une symétrie : pour un champ $\phi(x)$ dépendant de l'espace subissant la transformation locale suivante :

$$\phi(x) \longrightarrow e^{iq\alpha(x)}\phi(x) ; \quad (1.2.1)$$

où α est une quantité infinitésimale dont la valeur dépend de l'espace et q une charge électrique. Alors, le Lagrangien est transformé de la manière suivante :

$$\mathcal{L}(x) \longrightarrow \mathcal{L}(x) + \alpha(x)\partial_\mu J^\mu(x) ; \quad (1.2.2)$$

avec $J^\mu(x)$ un courant. Or, selon le théorème de Noether [127, 16], toute invariance du Lagrangien sous transformation correspond à une quantité conservée. Alors, toute symétrie respectant l'équation 1.2.2 sera associée à une quantité conservée correspondant au courant $J^\mu(x)$. Cette condition mène à la naissance des bosons de jauge, des particules qui seront discutées tout au long du chapitre et qui transportent cette invariance tout en étant les générateurs des groupes de symétrie.

Finalement, les quantités conservées du MS sont décrites par diverses charges. Ainsi, par exemple, la force forte conserve la charge de couleur, la force électromagnétique conserve la charge électrique et la force faible conserve la charge d'isospin faible. La théorie électrofaible, qui unit les forces électromagnétique et faible, conserve quant à elle l'hypercharge. En combinant les symétries de jauge reliées aux trois forces fondamentales du MS, la symétrie de jauge du MS est donnée par le groupe $SU(3) \times SU(2)_L \times U(1)_Y$, où l'indice L représente les particules gauchères (voir section 1.3.3) et Y représente l'hypercharge (voir section 1.3.4).

Ainsi, les trois forces fondamentales du MS se décrivent avec des symétries de jauge. La prochaine section analyse plus en profondeur les symétries impliquées dans ces forces et donne le Lagrangien associé à chacune d'elles.

1.3. Les forces

La section précédente présente les particules du MS et les forces avec lesquelles elles sont susceptibles d'interagir, en plus des bosons de jauge responsables de ces forces. Cette section

présente les trois forces fondamentales du MS plus en détail pour tenter de comprendre leur importance au LHC.

Avant de se plonger dans les détails de chaque force, expliquons d’abord quelques points communs entre elles. Par exemple, les différentes forces possèdent toutes une *constante de couplage* qui régit leur importance : plus grande est une constante de couplage, plus l’amplitude de cette force sera importante. À basse énergie, soit à environ 1 GeV, les ratios des différentes constantes de couplage du MS et de la gravité sont donnés par l’équation 1.3.1 [96] :

$$\alpha_s : \alpha_{em} : \alpha_W : \alpha_{grav} = 1 : \frac{1}{137} : 10^{-6} : 10^{-39} ; \quad (1.3.1)$$

où α_s est la constante de couplage de la force forte, α_{em} de la force électromagnétique, α_W de la force faible et finalement α_{grav} de la force gravitationnelle. Nous voyons à quel point les différentes forces sont à des échelles différentes et à quel point la gravité est tout à fait négligeable. Il est à noter que ces constantes varient en fonction de l’énergie. Or, il semblerait expérimentalement que ces différentes constantes convergeraient vers environ la même valeur à l’échelle d’énergie de Planck, soit autour de 10^{19} GeV. Ce seuil représente aussi l’énergie à laquelle la force gravitationnelle cesse d’être négligeable et où la théorie des champs quantiques (lire : le MS) devient incomplète puisqu’elle n’inclut pas la gravité. À cette énergie, les quatre forces fondamentales influencent les particules approximativement de la même façon. Toutefois, cette énergie est si grande qu’elle ne peut être sondée, et ainsi, leurs constantes de couplage n’ont pas pu être mesurées.

1.3.1. La force électromagnétique

La force électromagnétique, qui est communiquée via les photons, est théorisée sous la forme de l’électrodynamique quantique (QED) et a été développée dès les années 1920. La QED peut être vue comme étant une théorie de la perturbation du vide. Son succès a accéléré le développement de théories semblables pour les autres forces du MS, qui se sont fortement inspirées de celle-ci pour développer leur formalisme [77, Chapitre 2].

La QED propose un vertex de base qui fait interagir deux fermions chargés ainsi qu’un photon. Celui-ci est présenté à la figure 1.2 et permet de décrire toutes les autres interactions issues de la QED.

Notons que ce vertex de base ne peut exister seul dans la nature, puisqu’il pourrait aussi bien représenter un électron qui émet un photon. C’est impossible, car dans le référentiel de l’électron (où sa quantité de mouvement est nulle), il n’y a pas d’énergie disponible pour produire un photon. Il faut donc toujours le combiner à d’autres vertex primaires d’une des

forces du MS. Ce vertex primaire est décrit dans le Lagrangien de la QED aux équations 1.3.2 :

$$\mathcal{L}_{QED} = \bar{\psi} (i\gamma^\mu \partial_\mu - m) \psi - q\bar{\psi}\gamma^\mu\psi A_\mu - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} , \quad (1.3.2a)$$

$$F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu ; \quad (1.3.2b)$$

ici, les indices répétés sont sommés en suivant la convention de la sommation d'Einstein [123]. Ici, ψ représente le spineur de Dirac, $\bar{\psi}$ son adjoint de Dirac, A_μ le photon, m la masse du fermion impliqué, γ^μ les matrices de Dirac et q la charge de la particule impliquée.

De plus, le premier terme représente le terme cinétique, le second l'interaction entre le photon et les fermions (outre le neutrino) et le terme final représente les champs électriques et magnétiques des équations de Maxwell. Cette équation s'ensuit du fait que la QED est décrite par le groupe de symétrie $U(1)$, le groupe des transformations de phase unitaires. Celui-ci est aussi simplement le groupe des matrices 1×1 unitaires, c'est-à-dire l'ensemble des nombres complexes de norme 1 défini sous multiplication. Le générateur du groupe est représenté par le photon, un boson de jauge sans masse, tandis que la quantité conservée par le groupe est la charge électromagnétique. Le photon peut donc interagir avec toute particule chargée.

1.3.2. Interactions et conséquences connues de la QED

Tel que mentionné précédemment, la QED permet d'expliquer un grand nombre de phénomènes et non seulement le vertex primaire montré à la section 1.3.1. Parmi elles, on compte notamment l'absorption ou l'émission d'un photon, où un électron (ou positron) perd une partie de son énergie pour émettre un photon ou l'augmente en absorbant un. On y retrouve aussi l'annihilation ou la production de paires fermion-antifermion ainsi que la diffusion Compton. Ces processus sont représentés à la figure 1.3 et permettent de comprendre comment assembler les vertex primaires comme des blocs Lego : tant que chaque point de jonction entre les différentes particules respecte un vertex primaire, le diagramme de Feynmann est valide.

Enfin, la QED est aussi à l'origine du phénomène appelé *bremsstrahlung* (rayonnement continu de freinage) [126]. D'abord, selon les équations de Maxwell, toute charge électrique qui est accélérée émet un rayonnement sous forme de photons. Le *bremsstrahlung* se produit lorsqu'une charge (le plus souvent un électron) est déviée par les champs électrique et magnétique d'un noyau atomique. Cette interaction est très fréquente au LHC, puisque les particules qui y sont produites peuvent facilement entrer en contact avec les noyaux atomiques présents dans les atomes des détecteurs. De plus, à très grandes énergies comme au

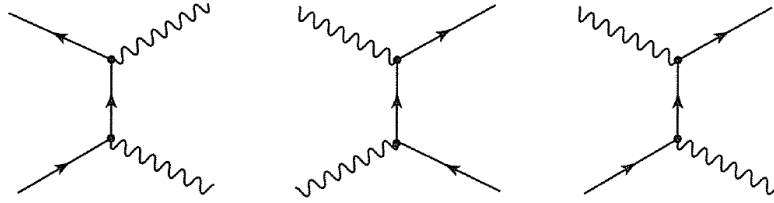


Fig. 1.3. Diagrammes de Feynmann de premier ordre d’interactions connues de la QED. À gauche, l’annihilation de paires est le processus où une paire de fermion-antifermion s’annihile et produit deux photons en retour. Au centre, le processus inverse se produit, où deux photons s’annihilent pour produire une paire fermion-antifermion. Finalement, à droite, la diffusion de Compton survient lorsqu’un photon interagit avec un électron, ce qui modifie l’impulsion des particules impliquées. De plus, elle peut survenir pour un électron lié à un noyau électronique, éjectant l’électron de l’atome et déviant la trajectoire du photon. La diffusion Compton est célèbre puisqu’elle est une preuve que la lumière n’est pas simplement ondulatoire, mais possède également une composante particulaire [77].

LHC, les photons issus du bremsstrahlung peuvent à leur tour créer une paire de fermion-antifermion qui pourra à son tour produire des photons via bremsstrahlung, et ainsi de suite. C’est ce qu’on appelle une *gerbe électromagnétique* ou *cascade électromagnétique*. Ces gerbes sont utilisées pour détecter la présence de particules d’intérêt dans des détecteurs de particules fonctionnant sous le principe de la calorimétrie (voir section 2.2.3), comme par exemple l’électron ou le photon.

1.3.3. La force nucléaire faible

La force nucléaire faible, produite via échange de bosons W et Z , interagit avec tous les fermions du modèle standard en leur permettant d’échanger de l’énergie, de la masse et de la charge électrique. D’ailleurs, la force faible est la seule capable de changer la saveur des quarks. En effet, elle permet changer un quark de charge électrique $+2/3$ en un autre quark de charge $-1/3$, ou vice versa (et peut aussi changer un lepton en son neutrino associé). Notons que ce changement de charge est possible puisque les bosons W^\pm sont chargés, donc la charge totale reste conservée. C’est d’ailleurs pourquoi les interactions sont séparées selon deux classes, soient les courants neutres et les courants chargés. Ces deux courants possèdent des vertex de base qui sont démontrés à la figure 1.4.

Les bosons de jauge W et Z peuvent aussi interagir entre eux. Enfin, puisque le W est chargé, il peut aussi interagir avec le photon, créant d’autres vertex primaires. Ceux-ci sont illustrés à la figure 1.5, mais ils sont tellement rares qu’ils ne portent pas une grande importance dans la phénoménologie de la force faible [77, Chapitre 2].

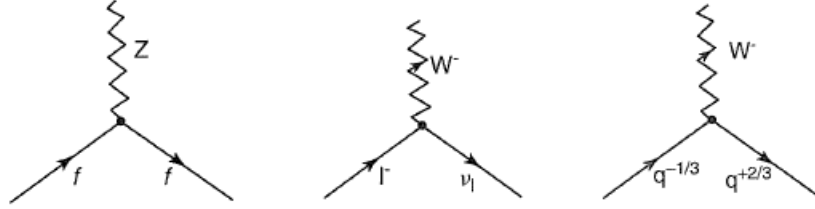


Fig. 1.4. Vertex de base de la force faible impliquant des fermions. À gauche, nous voyons le courant neutre qui échange un boson Z entre deux particules. Puis, les deux vertex de droite nous montrent le courant chargé, capable de changer un lepton en son neutrino associé ou même de changer la saveur des quarks.



Fig. 1.5. Vertex de base de la force faible n'impliquant que des bosons et des photons. Les trois vertex de gauche montrent les auto-interactions entre les bosons de la force faible, tandis que les trois autres de droite montrent le couplage du W avec le photon. Bien que ces auto-interactions apportent de nombreux nouveaux vertex primaires, ils n'apportent toutefois pas une grande contribution à la phénoménologie de la théorie en vertu de leur rareté [77, Chapitre 2].

Bien que la force faible permette le changement de génération des quarks, ceci reste un phénomène rare. En effet, comme il est coutume de le voir en mécanique quantique, les états propres des quarks sont composés d'une combinaison linéaire des différents états physiques et observables des quarks. Ainsi, si les états propres sont nommés d' , s' et b' et les états physiques d , s et b , nous retrouvons la relation suivante :

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}, \quad (1.3.3)$$

où la matrice centrale 3×3 est la *matrice de Cabibbo–Kobayashi–Maskawa*. Ses éléments sont des nombres complexes et décrivent la probabilité de changer d'un quark de saveur j vers une saveur i : elle est donnée par $|V_{ij}|^2$. Bien entendu, cette matrice est unitaire. Les modules des différentes entrées de cette matrice ont été trouvés expérimentalement [105] :

$$\begin{pmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{pmatrix} \approx \begin{pmatrix} 0.97428 & 0.2253 & 0.00347 \\ 0.2252 & 0.97345 & 0.0410 \\ 0.00862 & 0.0403 & 0.999152 \end{pmatrix}. \quad (1.3.4)$$

Les incertitudes ont été omises puisqu'on ne s'intéresse qu'aux ordres de grandeur des éléments. Comme il est possible de le voir, la matrice ressemble numériquement à une

matrice identité, et ainsi, tout changement de génération est rare, d'autant plus que la probabilité d'un tel phénomène est donnée par le carré des amplitudes de la matrice. De plus, tout changement de génération impliquant la troisième génération de quarks est encore plus rare, puisque les éléments hors diagonale de la troisième colonne et de la troisième rangée sont les plus petits. Enfin, notons que cette matrice est conventionnellement appliquée sur les quarks d , s et b , mais elle est équivalente et possède les mêmes valeurs numériques si nous l'appliquons plutôt sur les états physiques des quarks u , c et t .

Finalement, l'interaction faible peut être décrite par le groupe de symétrie $SU(2)$, le groupe des matrices unitaires sous rotation en deux dimensions, où les bosons W^\pm et Z représentent les générateurs du groupe. D'ailleurs, plutôt que de conserver la « charge » de la couleur comme la force forte, elle conserve plutôt l'isospin faible (I_W). La force faible est aussi la seule des forces fondamentales qui brise les symétries P et CP, qui consistent en l'application des opérateurs de parité (P) et de charge-parité (CP). Tandis que la parité inverse l'espace en changeant les coordonnées spatiales (x, y, z) en $(-x, -y, -z)$, l'opérateur de charge inverse la charge d'une particule, essentiellement passant d'une particule à une antiparticule ou vice-versa. Le bris de symétrie sous parité a été compris en introduisant un principe de *chiralité* qui change de valeur sous parité et où la force faible n'est appliquée que sur les particules de chiralité gauche.

1.3.4. La théorie électrofaible et le mécanisme de Higgs

La théorie électrofaible a été développée dans les années 1960. Elle unit la force électromagnétique ainsi que la force faible dans le sens qu'elle peut décrire ces deux interactions comme étant la conséquence d'un seul et même phénomène. Cette théorie est basée sur le groupe de symétrie $SU(2)_L \times U(1)_Y$. Rappelons-nous qu'aux sections 1.3.1 et 1.3.3, la force électromagnétique était décrite par le groupe $U(1)$ tandis que la force faible l'était par le groupe $SU(2)$, et ainsi, la théorie électrofaible est basée sur le produit direct des groupes des deux forces qui la composent. Ce dernier possède quatre générateurs, un associé à $U(1)_Y$ et les trois autres associés à $SU(2)_L$. Historiquement, ils ont été représentés par quatre bosons de jauge *sans masse*, soient le boson B_μ pour $U(1)_Y$ et les bosons W_μ^i , $i \in \{1, 2, 3\}$ pour $SU(2)_L$. Notons que ceux-ci ne représentent pas des bosons physiques, mais serviront plutôt à construire ces états via des combinaisons linéaires, tels que les quarks l'ont fait à l'équation 1.3.3. Nous verrons ces combinaisons à l'équation 1.3.6. Finalement, les W_μ^i ne couplent qu'aux fermions gauchers, ce qui est indiqué par l'indice L , voulant dire *left* (gauche). La charge liée au groupe $SU(2)_L$ est l'isospin faible, I_W , déjà introduit à la section 1.3.3, tandis que celle du groupe $U(1)_Y$ est l'hypercharge Y , qui est définie par $Y = 2(Q_{em} + I_W^3)$, où

Q_{em} représente la charge électromagnétique et I_W^3 représente la troisième composante de l'isospin faible.

Toutefois, nous savons que les bosons physiques W^\pm et Z^0 possèdent une masse tandis que le photon n'en possède pas. Or, cette théorie prédit quatre états propres sans masse, ce qui n'est pas cohérent. C'est ici qu'intervient le mécanisme de Higgs, où le champ scalaire de Higgs donne une masse aux particules. En effet, ce champ donne une valeur énergétique minimale au vide, notée par v et appelée *vev* (*vacuum expectation value*), traduite par *valeur attendue du vide*. C'est par l'interaction des particules avec ce champ qu'elles trouvent leur masse, et ainsi, nous pouvons sensiblement voir ce champ comme un réservoir d'énergie du vide dans lequel les particules interagissent pour y puiser leur masse. Toutes les particules du MS interagissent avec celui-ci, sauf le photon et le gluon, qui sont les deux seules particules sans masse du MS. L'interaction entre ce champ et les particules porte le nom de *potentiel de Yukawa*. Notons finalement que la particule du boson de Higgs représente des excitations de ce champ, traduites par des ondes de celui-ci. Cette théorie prédit finalement la masse des bosons de jauge physiques via les équations 1.3.5 [130] :

$$M_W^\pm = \frac{gv}{2} , \quad (1.3.5a)$$

$$M_Z = \frac{v\sqrt{g^2 + (g')^2}}{2} , \quad (1.3.5b)$$

$$M_\gamma = 0 ; \quad (1.3.5c)$$

où v est le *vev*, et g et g' sont les constantes de couplages des groupes $SU(2)_L$ et $U(1)_Y$, respectivement. Finalement, la théorie prédit aussi quels sont les états physiques des quatre bosons de jauge introduits qui sont donnés aux équations 1.3.6 :

$$W_\mu^\pm = \frac{W_\mu^1 \mp iW_\mu^2}{\sqrt{2}} , \quad (1.3.6a)$$

$$A_\mu = B_\mu \cos \theta_W + W_\mu^3 \sin \theta_W , \quad (1.3.6b)$$

$$Z_\mu = -B_\mu \sin \theta_W + W_\mu^3 \cos \theta_W ; \quad (1.3.6c)$$

où l'angle faible de mixage θ_W a été introduit [130]. Notons que A_μ représente le photon et Z_μ le boson Z^0 . Finalement, le Lagrangien de la théorie électrofaible donne aussi lieu à des interactions entre les bosons intermédiaires. Il donne aussi naissance aux vertex à trois bosons et quatre bosons vus à la figure 1.5. Ce Lagrangien est donné aux équations 1.3.7 :

$$\mathcal{L}_{\text{théorie électrofaible}} = \mathcal{L}_{\text{leptons}} + \mathcal{L}_{\text{Electroweak}} + \mathcal{L}_{\text{Higgs}} . \quad (1.3.7a)$$

$$\mathcal{L}_{\text{leptons}} = \sum_i \bar{\psi}_i (i\gamma^\mu \partial_\mu - m_i) \psi_i , \quad (1.3.7b)$$

$$\mathcal{L}_{\text{Electroweak}} = \frac{g}{2} \sum_{\phi_L} \bar{\phi}_L \left(\sigma_{1,2,3} W_\mu^{1,2,3} \right) \phi_L - \frac{g'}{2} \sum_i \bar{\psi}_i (Y_W B_\mu) \psi_i , \quad (1.3.7c)$$

$$\mathcal{L}_{\text{Higgs}} = (D_\mu \Psi)^\dagger (D^\mu \Psi) - m^2 \Psi^\dagger \Psi + \lambda (\Psi^\dagger \Psi)^2 , \quad (1.3.7d)$$

$$D_\mu = \partial_\mu + \frac{ig\sigma^i W_\mu^i}{2} + \frac{ig'Y B_\mu}{2} . \quad (1.3.7e)$$

Encore une fois, les ψ_i représentent les champs des leptons, les σ les matrices de Pauli, Ψ le champ scalaire du Higgs et m la masse observable du boson de Higgs. Finalement, notons que la découverte expérimentale du boson de Higgs, annoncée le 4 juillet 2012 par les expériences ATLAS et CMS, a confirmé la théorie électrofaible [56, 46].

1.3.5. La force nucléaire forte

La force nucléaire forte est régie par la chromodynamique quantique (QCD). Celle-ci interagit avec la *couleur* des particules et n'agit donc que sur les quarks et les gluons (ou tout état lié composé de ces particules, comme les mésons et les baryons). Cette force est reconnue notamment pour assurer la cohésion du noyau atomique. En effet, elle surpasse la force électromagnétique (en raison de sa plus forte constante de couplage dans le noyau) qui voudrait séparer les protons chargés positivement. Par contre, elle est moins connue en raison de sa portée très minimale de l'ordre de 10^{-15} m. Au-delà de cette distance, sa magnitude est tellement augmentée qu'elle crée des nouvelles particules en raison du principe de liberté asymptotique (voir section 1.3.6). C'est donc paradoxalement qu'en ayant une si grande magnitude aux grandes distances qu'elle n'est jamais permise d'opérer à celles-ci, car l'apparition de nouvelles particules recrée un système où les distances sont toutes petites.

Puisque les gluons possèdent une couleur, la force forte peut parvenir à changer la couleur des quarks. En effet, la couleur est une quantité conservée et un échange de gluons permet donc ce changement sans briser les lois de la physique. C'est toutefois la seule interaction pouvant le faire puisque le gluon est le seul boson de jauge coloré. Cela s'explique du fait que la force forte est décrite par le groupe de symétrie $SU(3)$. En effet, les trois couleurs représentent les trois composantes de $SU(3)$ et les huit gluons sont les générateurs du groupe ; ceci assure la conservation de la couleur. Ceci donne lieu à trois vertex primaires présentés à la figure 1.6 qui sont décrits par le Lagrangien de la QCD aux équations 1.3.8 :

$$\mathcal{L}_{QCD} = \sum_q \bar{\psi}_{q,a} \left(i\gamma^\mu \partial_\mu \delta_{ab} - g_s \gamma^\mu t_{ab}^C \mathcal{A}_\mu^C - m_q \delta_{ab} \right) \psi_{q,b} - \frac{1}{4} \mathcal{F}_{\mu\nu}^C \mathcal{F}_C^{\mu\nu} , \quad (1.3.8a)$$

$$\mathcal{F}_C^{\mu\nu} = \partial^\mu F_C^\nu - \partial^\nu F_C^\mu - gf^{C\beta\gamma} F_\beta^\mu F_\gamma^\nu ; \quad (1.3.8b)$$

où les indices q et a représentent les saveurs de quark ainsi que les trois couleurs, respectivement, m_q la masse du quark, \mathcal{A}_μ^C les champs des gluons, C les huit gluons, ψ les champs des quarks, γ^μ les matrices de Dirac et finalement t_{ab}^C les huit matrices génératrices du

groupe $SU(3)$. Notons que la notation d'Einstein [123] est encore utilisée. Dans l'équation 1.3.8a, les premier et troisième termes à l'intérieur des parenthèses représentent les termes cinétiques du Lagrangien tandis que celui du centre représente l'interaction entre deux quarks et un gluon (voir figure 1.6 à gauche). Finalement, le dernier terme de l'équation 1.3.8a régit la cinématique des gluons (deux premiers termes de l'équation 1.3.8b) et leurs auto-interactions (dernier terme de l'équation 1.3.8b), dont les vertex primaires sont donnés au centre et à droite de la figure 1.6.

Notons finalement que le Lagrangien complet du modèle standard est donné en additionnant le Lagrangien de la force électrofaible (équation 1.3.7) et celui de la force forte (équation 1.3.8).

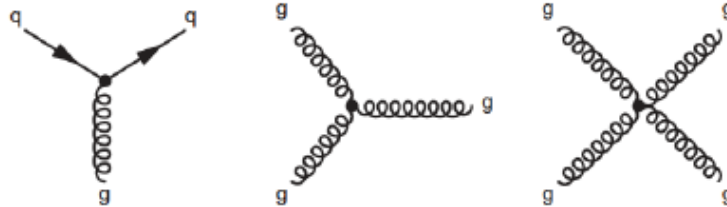


Fig. 1.6. Les trois vertex de base de la QCD. À gauche, nous voyons l'interaction quark-quark-gluon, puis nous apercevons les interactions à trois gluons (centre) et à quatre gluons (droite) [77, Chapitre 2]. Encore une fois, tout assemblage de ces vertex donne lieu à un phénomène physique possible.

1.3.6. Les jets

Les jets sont des flots collimatés de particules issus de la production de quarks, d'anti-quarks et/ou de gluons énergétiques dans des expériences de physique des particules à hautes énergies. Toutefois, le principe derrière la formation des jets est issu de la QCD et demande une certaine explication. En effet, la force forte possède la spécificité de produire de très petites interactions à courte portée (de l'ordre de 10^{-15} m), par exemple à l'intérieur des noyaux atomiques. Ceci s'explique par la quantité α_s , la constante de couplage de la force forte, qui est dissimulée dans le Lagrangien de la QCD aux équations 1.3.8 par la relation $\alpha_s = g_s^2/4\pi$. En effet, la première chose à savoir par rapport à cette constante de couplage, c'est qu'elle n'est pas constante. Elle diminue à hautes énergies, ce qui se traduit habituellement par des petites distances. Ce comportement est connu sous le nom de *liberté asymptotique* et est montré à la figure 1.7.

Une des conséquences de ce comportement est qu'en tentant de séparer deux quarks, la force forte devient si forte qu'il en devient énergétiquement plus avantageux, c'est-à-dire que l'énergie totale du système sera plus petite, s'il y a production de paires de

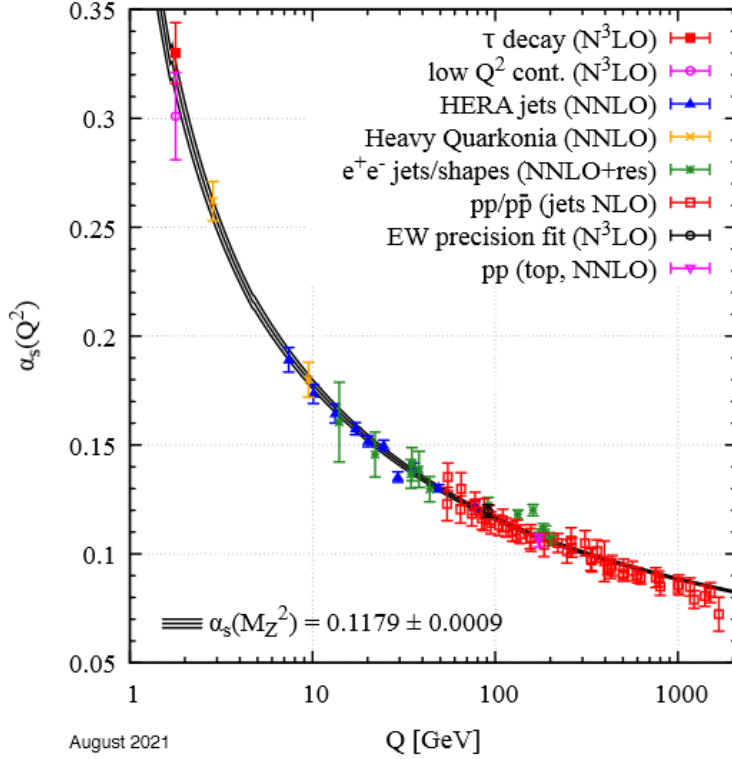


Fig. 1.7. Magnitude de la constante de couplage de la force forte, α_s , en fonction de l'échelle d'énergie Q (GeV). Cette figure sert aussi à comprendre la liberté asymptotique : les quarks sont libres, c'est-à-dire qu'il ne subissent que peu d'interaction de la force forte si la distance entre eux est petite (ce qui correspond à des hautes énergies). À l'inverse, les particules éloignées subiront une force forte importante. Sur cette figure, les lignes pleines représentent l'intervalle de confiance prédit par la théorie tandis que les points représentent diverses mesures prises expérimentalement avec leurs incertitudes [130, Chapitre 9].

quark-antiquark. Ce phénomène est illustré à la gauche de la figure 1.8. C'est aussi ce qui explique le principe de confinement des quarks mentionné à la section 1.1 qui stipule que les quarks doivent toujours être liés pour former des états de couleur neutre. En effet, supposons que nous voulions créer un quark libre. Il faudrait alors le subtiliser d'un état lié, mais en raison de l'augmentation de la force forte au fur et à mesure que nous retirerions le quark de son état lié précédent, il deviendrait alors plus avantageux de créer une paire de quark-antiquark. De plus, tout nouvel état lié créé devra être de couleur neutre.

À hautes énergies, cette création de paires peut être répétée successivement. En effet, les quarks produits doivent être émis en direction opposée pour respecter le principe de la conservation de la quantité de mouvement, ce qui favorise la création d'une nouvelle paire, et ainsi de suite. Les quarks s'agglomèrent enfin pour constituer des hadrons lorsqu'il n'y a plus assez d'énergie pour séparer les nouvelles paires créées. C'est ce qu'on appelle

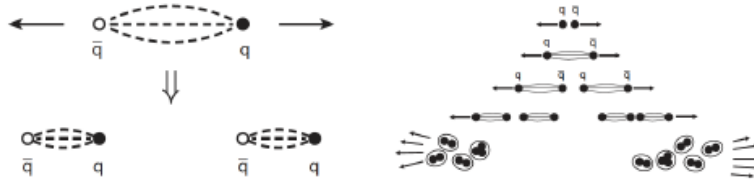


Fig. 1.8. Visualisation de la formation de cascades de quarks et de l’hadronisation. En raison de la liberté asymptotique, il est impossible de séparer deux quarks formant un état lié en quarks, résultant alors en la production de deux paires quark-antiquark (gauche). À hautes énergies, ce processus peut se répéter plusieurs fois pour créer une cascade de quarks (droite), appelés *jets* si ce flot de particules est suffisamment rapproché dans l’espace [99]. Une fois l’énergie suffisamment basse, les quarks s’agglomèrent en hadrons, ce que l’on appelle l’hadronisation.

l’*hadronisation* [118], qui peut être observée dans la dernière partie au bas de la figure 1.8. Si ces particules sont toutes produites dans un flot collimaté de particules, nous sommes dorénavant en présence des fameux *jets*. Ce phénomène se produit notamment si la particule fragmentée d’origine possède une quantité de mouvement importante, typiquement de l’ordre de la dizaine de GeV ou plus.

Les jets surviennent lorsqu’une particule à la fois très massive et très énergétique se désintègre en au moins deux quarks ou gluons qui initient chacun un jet. Toutefois, si ces jets sont tous rapprochés dans l’espace et alignés selon une direction semblable, ils peuvent alors être considérés comme étant un seul grand jet. Plus précisément, on parle de jets dits *boostés* lorsque la particule initiale était suffisamment énergétique pour que les constituants (particules) des jets puissent être regroupés dans un seul cône étroit. Ainsi, nous sommes en présence d’un grand jet possédant des sous-jets, plutôt que plusieurs jets distincts, tel qu’illustré à la droite de la figure 1.9. Les exemples connus de particules pouvant résulter en la formation de jets boostés incluent le W , le Z , le Higgs et le top.

La gauche de la figure 1.9 montre un jet qui n’est pas boosté malgré la présence d’un top qui se désintègre ultimement en plusieurs jets. En effet, le top initial n’a pas une quantité de mouvement suffisante (n’est pas assez énergétique) pour que les jets produits soient rapprochés. Il ne ferait alors pas de sens de reconstruire les trois jets illustrés en un seul, puisque le rayon de ce jet hypothétique serait énorme.

Au LHC et dans le détecteur ATLAS, de nombreux algorithmes existent pour reconstruire les jets (voir section 2.4.1). En effet, ceux-ci utilisent l’information des produits de désintégration recueillis par les différents détecteurs pour déterminer la présence ou non d’un jet, et, le cas échéant, le reconstruire jusqu’à son origine. Pour les jets boostés, la *masse invariante* des constituants de tous les jets doit sommer à la masse invariante de la particule massive s’étant désintégrée. Nous pouvons donc parfois inférer sur la particule à



Fig. 1.9. Comparaison entre un jet non boosté et un jet boosté dans un collisionneur de particules. Dans ce cas particulier, la collision initiale entre deux protons produit un quark top. À gauche, le top est trop peu énergétique et les jets ultimement créés sont trop éparés (jet non boosté). À droite, le top est suffisamment énergétique pour pouvoir reconstruire tous les jets en un seul grand jet boosté [119].

l'origine de ce jet si la masse du jet est très proche de celle d'une particule connue. Par contre, il existe des jets qui sont trop éparés et dont la reconstruction « manquera » certains constituants qui étaient un peu trop éloignés du cône de reconstruction. Ceci arrive quand un jet flirte avec la définition de jet boosté, quand son rayon est juste assez petit pour reconstruire les sous-jets en un seul grand jet. Dans ce cas-ci, le jet est « incomplet » et la masse des particules le composant est inférieure à la masse réelle de la particule d'origine. Ce phénomène sera important lors de la génération de certains ensembles de données simulées que j'ai effectuées dans ce projet de recherche (voir section 4.3.3). La reconstruction générale des jets est discutée en plus grands détails à la section 2.4.1.

Finalement, notons que ce projet de mémoire se concentrera sur l'étude des jets boostés, puisqu'il est possible que de la nouvelle physique s'y retrouve [120]. En effet, il suffit d'avoir une nouvelle particule massive qui se désintègre en particules pouvant ultimement créer des jets pour qu'elle puisse former un jet boosté, qui est alors issu de la nouvelle physique. Elle serait aussi relativement simple à découvrir si nous avions les données de tous les jets issus de celle-ci puisqu'ils auraient une masse très proche de cette particule. C'est d'ailleurs en évaluant la masse des jets anormaux que nous espérons trouver de la nouvelle physique (voir section 4.1).

1.4. La physique au-delà du modèle standard

Ce chapitre a abordé les principales facettes du modèle standard, une théorie robuste qui a été prouvée et confirmée expérimentalement avec précision au fil du temps [13]. Toutefois, nous savons que ce modèle, bien qu'excellent pour ce qu'il décrit, possède quelques lacunes et faits inexplicés. La preuve la plus éloquent de ce manque à gagner revient sans doute au fait que le MS ne décrit que 5% de l'énergie totale de l'Univers, tandis que le reste demeure inexplicé. Cette section est destinée à présenter les problèmes actuels du modèle standard,

de faire un bref survol de quelques théories qui pourraient combler certaines brèches du MS et finalement les résultats expérimentaux de ces théories.

1.4.1. Les lacunes du modèle standard

Pour débiter, la communauté scientifique s'entend pour statuer qu'il existe fort probablement au moins deux autres sources d'énergie autres que celle de la matière ordinaire comprise dans le MS. En effet, des études astrophysiques et cosmologiques ont repéré des sources d'énergie ou de masse inexpliquées par le modèle standard.

Tout d'abord, c'est en observant le mouvement d'étoiles d'amas de galaxies que des scientifiques élaborèrent la possibilité qu'une partie de la masse de cet amas était « cachée » vers les années 1930. En effet, selon des mesures de masse associées à la luminosité de ces galaxies, il leur manquerait énormément de masse pour justifier la vitesse élevée des étoiles qui les composent [65]. Depuis, de nombreuses autres observations astrophysiques ne peuvent être expliquées par les théories gravitationnelles connues à moins que l'on soit en présence d'une très grande quantité de matière invisible dans ces galaxies. Parmi elles figurent surtout la rotation des galaxies, un phénomène étudié entre autres par la physicienne Vera Rubin lors des années 1960 [129]. Toutes ces observations pointent donc en direction de l'existence de la matière sombre qui serait invisible pour nous puisqu'elle n'interagirait pas avec la force électromagnétique – et ainsi, elle ne produit, n'absorbe ou n'émet aucune lumière ou toute autre radiation électromagnétique détectable. C'est d'ailleurs son absence de lumière qui lui a valu son nom de matière *sombre*, ou *dark matter*. La densité de la matière sombre est estimée à 27% dans l'Univers [65, 100].

L'observation de l'accélération de l'expansion de l'Univers a aussi mené à la théorisation de l'énergie sombre. En effet, les galaxies s'éloignent les unes des autres, ce qui peut être expliqué par le Big Bang, par exemple. Toutefois, en raison de la force gravitationnelle qui tend à rapprocher les galaxies les unes des autres, leur éloignement devrait être ralenti au fil du temps, puisqu'une force tend à les rapprocher. Toutefois, nous observons plutôt une accélération de cet éloignement. Une force doit donc contribuer à la répulsion des galaxies, mais il n'existe pour le moment aucune explication dans le modèle standard ou dans les théories gravitationnelles pour expliquer ceci, d'où l'introduction de l'énergie sombre comme explication potentielle de ce phénomène. Tandis que la matière sombre exerce une force pour rapprocher les galaxies, l'énergie sombre les repousse les unes des autres. Celle-ci est largement inconnue [98] et représenterait les 68% restants de l'énergie de l'Univers [65, 100, 98]. Notons que la matière sombre et l'énergie sombre n'ont jamais directement été observées : leur existence est supposée via toutes les observations indirectes qui ont été

mentionnées.

Il existe aussi d'autres problèmes avec le modèle standard actuel. Non seulement ne décrit-il que 5% de l'Univers, des considérations théoriques sont toujours incomprises en son sein. En effet, lors des calculs théoriques de la masse du boson de Higgs, le modèle standard stipule qu'il faut tenir compte des interactions en forme de boucle comme celle présentée à la figure 1.10. Il devient alors possible d'ajouter un nombre quelconque d'interactions de la sorte à n'importe quel diagramme de Feynmann impliquant le Higgs. En ajoutant ces contributions aux calculs nécessaires pour prédire la masse du Higgs, celle-ci est de l'ordre de grandeur de la masse Planck (autour de 10^{19} GeV). Toutefois, sa masse observée expérimentalement est d'environ 125 GeV [49, 59]. Ce phénomène connu possède le nom de *problème de hiérarchie*.

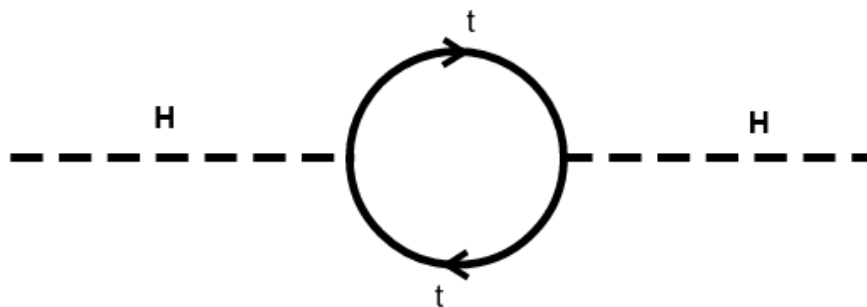


Fig. 1.10. Diagramme le plus simple d'un boson de Higgs entrecoupé d'une boucle de quarks top. En ajoutant des boucles de la sorte dans les différentes interactions qui doivent être tenues en compte pour mesurer la masse du Higgs, nous arrivons à la conclusion théorique qu'elle doit être 17 ordres de grandeur plus grande qu'elle ne l'est vraiment.

Une autre inexactitude du modèle standard est qu'il prédit que les neutrinos sont sans masse. Par contre, sans connaître leur masse exacte, il a été démontré que tous les neutrinos en possèdent une très petite, mais non nulle. En effet, de nombreuses expériences sur l'oscillation des neutrinos ont démontré que ces derniers peuvent changer de saveur. Par exemple, ce changement a été observé entre les neutrinos émis par le soleil et frappant la Terre où la saveur des neutrinos pouvait changer durant la période entre leur émission du soleil et leur analyse sur Terre. De plus, les expériences SNO [115], au Canada et T2K (*Tokai to Kamioka*) [61], au Japon, ont aussi prouvé cette oscillation [4, 10]. Cette expérience produit un faisceau quasiment pur de neutrinos muoniques à Tokai qui est ensuite envoyé à Kamioka (deux villes japonaises situées à 295 km de distance l'une de l'autre), à 1700 mètres sous le niveau de la mer. Toutefois, une fois arrivés à Kamioka, le faisceau présente beaucoup de neutrinos électroniques, ce qui démontre l'oscillation de neutrinos $\nu_\mu \rightarrow \nu_e$. Ces oscillations prouvent que la masse du neutrino est non nulle, ce qui

en a bouleversé leur compréhension et, plus largement, du modèle standard. Depuis, toutes sortes d'expériences tentent de trouver la masse des différents neutrinos et tenter de mieux comprendre l'origine et les implications de cette découverte.

Ensuite, un problème de taille du MS est qu'il ne décrit pas la gravité. De plus, la mécanique quantique, qui est à la base du MS ainsi que la théorie générale de la relativité, qui décrit la gravité, sont incompatibles dans des conditions extrêmes, par exemple à très hautes énergies. Ainsi, le modèle standard ne régit pas toutes les interactions possibles entre les particules, ce qui le rend très incomplet. Les physiciens ont déjà commencé à tenter de rallier le modèle standard et la théorie de la relativité d'Einstein, qui représente la théorie connue et précise de la gravité. Cette « théorie du tout » ou « théorie d'unification » en est encore à ses balbutiements, et rien de solide n'a été trouvé en ce sens. Le développement d'une telle théorie serait assurément l'un des plus grands accomplissements scientifiques de l'histoire.

Finalement, il existe un petit bémol sur le MS actuel qui n'est pas un problème en soi, mais qui découle peut-être d'un manque de compréhension plus profond sur la nature propre de ses différentes parties. En effet, ce dernier comporte 19 paramètres libres, comme par exemple les masses des différents fermions et les constantes de couplage des trois forces décrites par le MS, mais aussi quelques autres qui ont été abordés dans ce chapitre, comme l'angle électrofaible θ_W , la valeur attendue du vide (vev) v et les éléments de la matrice CKM, où ces derniers sont construits à partir de 4 paramètres libres. Aucune théorie ne peut prédire ces paramètres et ils sont indépendants les uns des autres : il faut donc passer par l'expérience pour déterminer leurs valeurs. Idéalement, il y aurait beaucoup moins de ces paramètres libres, préférablement qu'un seul, sur lequel tout le MS serait basé. Bien que ce problème ne sera peut-être jamais réglé, force est d'admettre qu'il rend le MS un peu moins robuste.

Tous ces arguments ont donné beaucoup de motivation aux physiciens pour tenter de compléter et de parfaire le MS. Ainsi, de nombreuses théories ont été concoctées et de nombreuses expériences ont été réalisées pour essayer de trouver des signes de la nouvelle physique.

1.4.2. Les théories au-delà du modèle standard

Le nombre de théories au-delà du modèle standard qui ont été écrites est ahurissant. La forme, la complexité et les buts de ces théories divergent grandement, mais elles tentent toutes d'aborder au moins un des problèmes de la section précédente. Quelques-unes de

ces théories se distinguent particulièrement du lot puisqu'elles semblent plus complètes et plausibles et ont généré un engouement suffisant pour que celles-ci soient recherchées dans les grandes expériences scientifiques du monde. Nous nous pencherons sur quelques-uns de ces modèles à l'aide d'un survol rapide de leurs prédictions ainsi que sur l'état de leur recherche dans le monde réel.

La théorie la plus populaire à avoir vu le jour est sans doute la supersymétrie (SUSY). Selon celle-ci, chaque particule du modèle standard aurait un équivalent supersymétrique, c'est-à-dire une nouvelle particule qui possède les mêmes propriétés que sa particule associée, mais avec un spin qui diffère de $1/2$ et avec une masse différente [33]. Se rappelant que les fermions ont des spins demi-entiers et que les bosons ont des spins entiers, cela signifie que les bosons sont associés à des fermions et vice-versa. Par exemple, il existerait le stop, le sneutrino et le sup, les particules scalaires (spin 0) associées au top, au neutrino et au quark up. Ce modèle impressionnant qui ouvre la porte à des dizaines de nouvelles particules trouve sa vertu en étant capable de régler la divergence de la masse du boson de Higgs. En effet, dans ce modèle, le Higgs couple aussi au stop, et peut donc interagir avec lui avec un diagramme en boucle, de la même façon que le Higgs le fait avec le top à la figure 1.10. Ces corrections à la masse du Higgs aux ordres supérieurs contrecarreraient donc celles apportées par le top et ainsi, le problème de hiérarchie de la section précédente serait réglé [15]. Un autre avantage marqué de ce nouveau modèle est qu'il permet de faire converger les constantes de couplage des forces du MS à l'échelle de Planck, ce qui pourrait être un début vers l'unification des forces du MS [69]. Finalement, ces particules supersymétriques sont des candidates de ce qui forme la matière sombre [130, Chapitre 88]. Ce modèle solide sur papier a fait l'objet d'immenses études au LHC, mais malheureusement, elles n'ont abouti sur aucun élément concret venant confirmer ou même suggérer son existence. Par contre, ces expériences ont réussi à imposer des limites de plus en plus contraignantes sur les masses de ces particules. Celles-ci sont présentées à la figure 1.11.

Il existe aussi la catégorie de modèles des WIMP (*Weakly Interacting Massive Particles*, qui se traduit en *particules massives interagissant faiblement*). Tel que leur nom le suggère, ces particules auraient une masse non nulle mais n'interagiraient presque pas avec la matière « ordinaire », sauf par gravité. Ainsi, ces modèles hypothétiques possèdent des candidats parfaits pour la matière sombre. Si ces particules existent, elles seraient toutefois très compliquées à observer puisqu'elles n'interagissent presque pas avec la matière ordinaire. Ainsi, les recherches expérimentales se concentrent sur quelques avenues pour les détecter. Par exemple, des groupes se spécialisent pour tenter de détecter des produits de désintégration des WIMP qui s'annihilent entre eux en provenance des galaxies. Une des difficultés de cette recherche est que l'on retrouve aussi ce genre de produits de désintégration dans les

processus issus de la matière ordinaire, c'est-à-dire des rayons gamma, des neutrinos et des rayons cosmiques. Nous sommes donc en présence d'un important bruit de fond difficile à réduire. Il existe aussi des expériences tentant de détecter les WIMP directement en utilisant d'immenses détecteurs de plusieurs tonnes pour augmenter les chances que les WIMP interagissent avec eux. Le détecteur PICO [9], notamment, recherche cette matière. Pour minimiser le nombre d'interactions avec la matière ordinaire, ce détecteur situé en Ontario est placé à plus de 2 kilomètres sous terre. D'autres expériences d'envergure se penchent aussi sur la détection directe des WIMP, dont Xenon 1T [11] et CDMS [8]. Un autre espoir de ce modèle est de voir apparaître des WIMP dans des grands collisionneurs comme le LHC, puis de les détecter. Cependant, comme avec la supersymétrie, aucune preuve de la validité de cette théorie n'a pu être formulée.

Un autre modèle qui est moins médiatisé dans le monde scientifique consiste en celui des axions, qui seraient des bosons neutres électriquement de masse très petite, mais non nulle (de l'ordre des μeV à quelques eV). Ils seraient eux aussi des constituants de la matière sombre et trouveraient leur origine théorique au fait qu'ils pourraient être la solution de la violation CP en chromodynamique quantique [121]. Une voie théorique pour observer ces particules existe déjà. En effet, ces bosons pourraient se transformer en photons sous l'influence de champs magnétiques très élevés [130, Chapitre 91]. Puis, l'effet Primakoff [108] permettrait à ces photons très énergétiques de produire des mésons, et c'est cette production qui serait analysée pour trouver ces bosons. Plusieurs expériences pour tenter de découvrir ces particules se retrouvent au CERN, comme par exemple CAST [89, 26] et OSQAR [17, 41]. Il existe aussi l'expérience ADMX [84] située au centre de physique expérimentale de l'Université de Washington. Sans surprise pour tout lecteur avisé, aucun résultat de ce modèle n'a été démontré expérimentalement.

Finalement, l'hypothèse se basant sur les particules de Majorana expliquerait la masse non nulle des neutrinos et leurs oscillations de saveur [5]. Cette nouvelle théorie prédit des particules de Majorana, des fermions qui sont aussi leur propre antiparticule et donc électriquement neutres. Parmi les suspects les plus probables de ces nouvelles particules, on retrouve le neutrino de Majorana [3] ainsi que le neutralino [47], une des particules issues de la supersymétrie, mais qui n'est pas le partenaire direct du neutrino, ce dernier étant le sneutrino.

Finalement, bien qu'il existe de nombreux autres modèles qui peuvent tous pallier à certaines lacunes du MS, aucun n'a pu être observé. De plus, des signes de nouvelle physique tardent à se manifester. Plusieurs pensent que cette nouvelle physique n'apparaît ou n'est

délectable qu'à haute énergie, et ainsi, l'un des défis les plus importants des grandes expériences mondiales, dont le LHC, est d'augmenter l'énergie disponible de leurs installations. Les figures 1.12 et 1.13 résument les différentes particules hypothétiques recherchées expérimentalement au LHC par l'expérience ATLAS et y montrent les limites sur leur masse et leur temps de vie. Nous pouvons voir le nombre important de ces particules et la quantité d'efforts qui ont été déployés pour produire tous ces résultats dépendant des modèles théoriques déjà concoctés.

1.5. Discussion

L'existence de plusieurs modèles théoriques pouvant aller au-delà du modèle standard mais dont aucun ne démontre de résultats intéressants au niveau expérimental suggère que de chercher de la nouvelle physique sous forme de modèles pré-existants n'est peut-être pas optimal. En effet, il faut de nombreuses années pour élaborer les modèles, puis de nombreuses autres pour les rechercher dans des analyses de données complexes. De plus, de tous ces modèles, il en existe certainement plusieurs faux. Une grande quantité d'efforts qui y sont mis sont donc pour ainsi dire gaspillés.

Il est décidément plus novateur de rechercher de la nouvelle physique sous forme d'*anomalies* dans les données actuelles des grandes expériences, dont le LHC et ATLAS. Ainsi, au lieu de rechercher des modèles spécifiques qui n'existent potentiellement pas, nous proposerons dans la suite de ce mémoire des techniques novatrices pour rechercher de la nouvelle physique se manifestant sous forme d'anomalies dans les données d'ATLAS. Celles-ci sont définies par des données qui sont les plus difficilement explicables par le modèle standard. Cela se traduit par des données différentes des autres. Pour ce faire, nous ferons appel à l'apprentissage machine (*machine learning*). L'avantage ici est que toute anomalie peut être détectée, peu importe son explication théorique et sans la supposer.

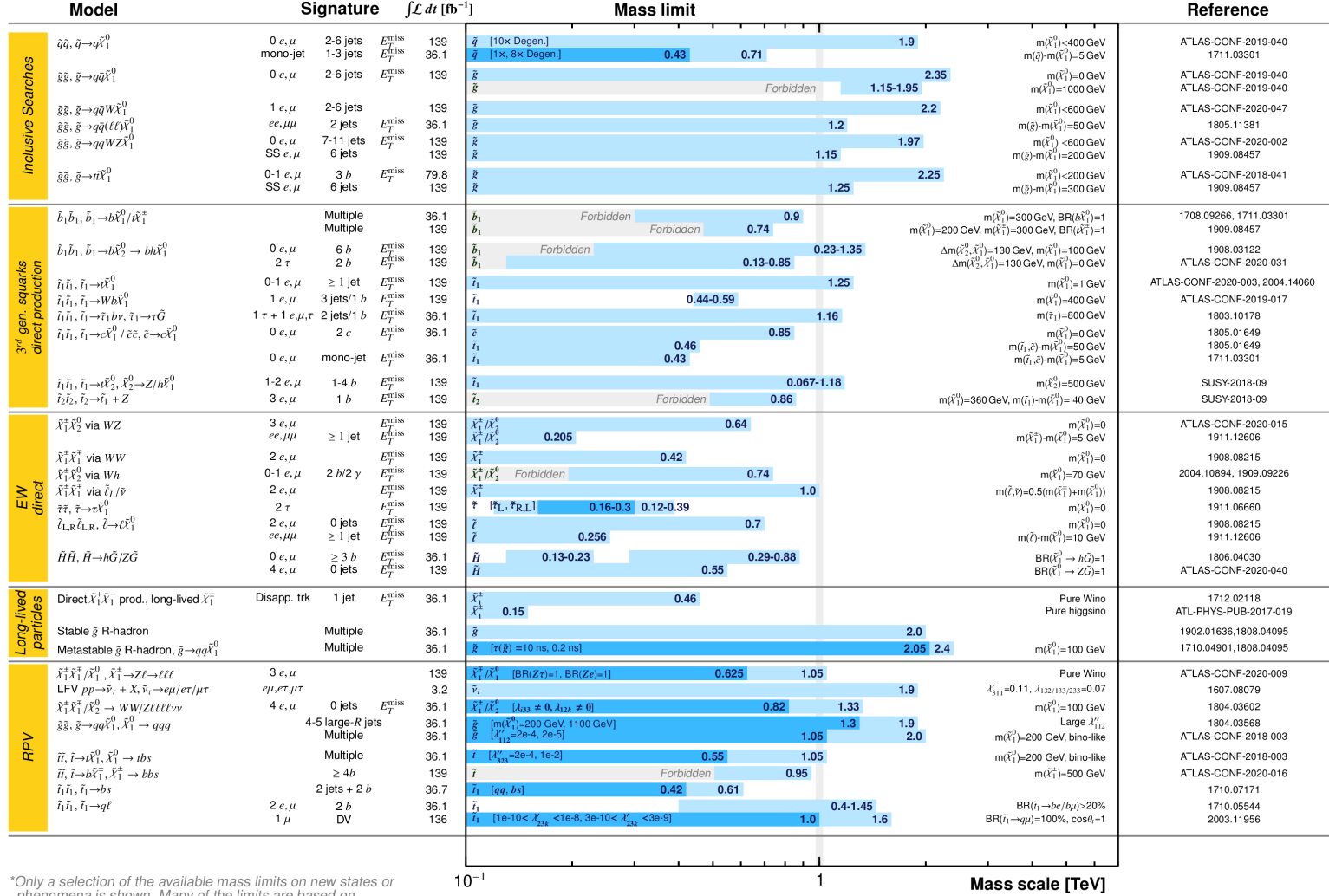


Fig. 1.11. Limites sur les masses des particules supersymétriques hypothétiques données par l'expérience ATLAS. Les valeurs sont données à un degré de confiance de 95% [14].

ATLAS Long-lived Particle Searches* - 95% CL Exclusion

Status: May 2020

ATLAS Preliminary

$$\int \mathcal{L} dt = (18.4 - 136) \text{ fb}^{-1}$$

$$\sqrt{s} = 8, 13 \text{ TeV}$$

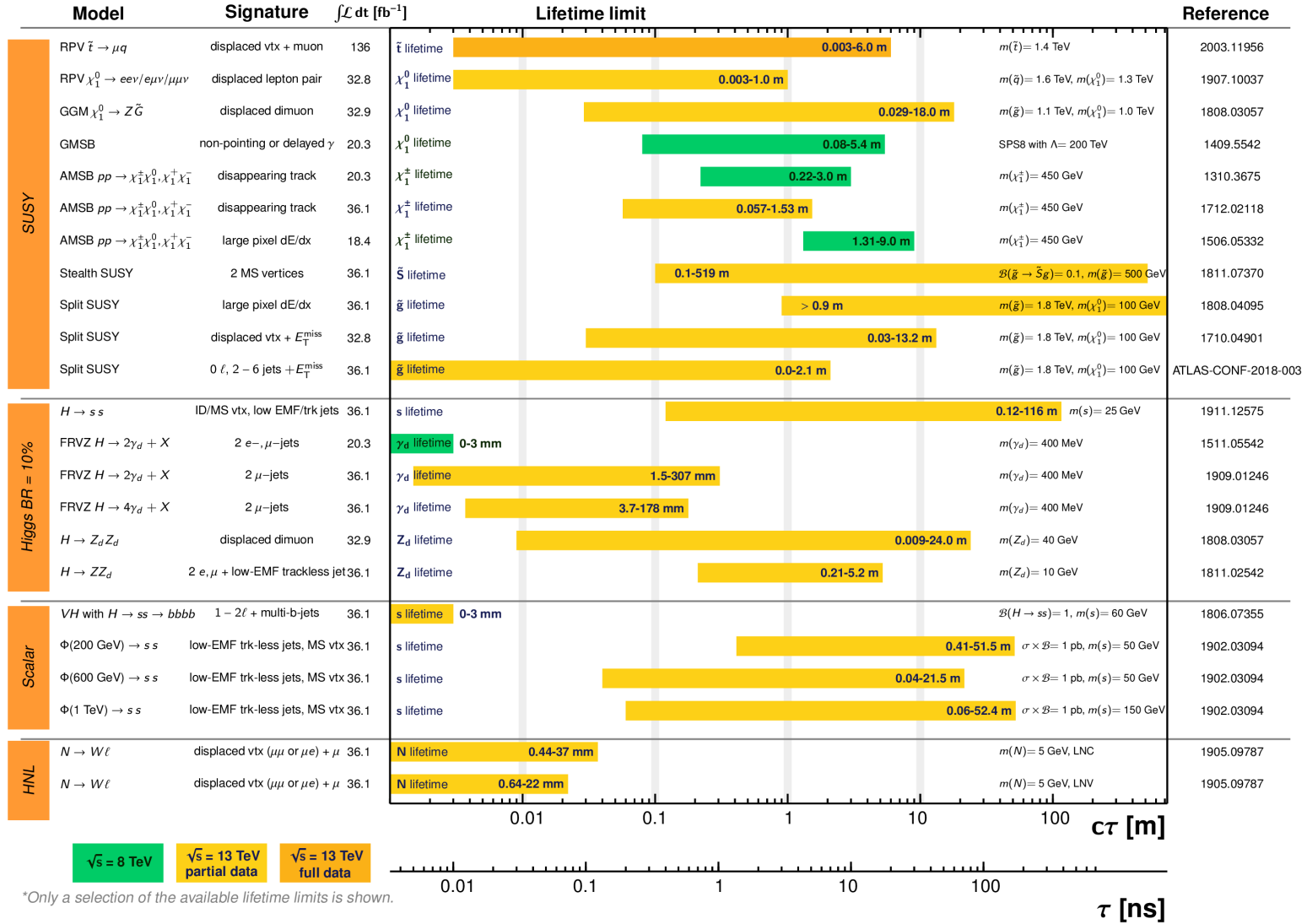


Fig. 1.13. Limites sur les temps de vie des nouvelles particules hypothétiques données par l'expérience ATLAS. Les valeurs sont données à un degré de confiance de 95% [12].

Chapitre 2

Le Grand collisionneur de hadrons (LHC) et l'expérience ATLAS

La physique des particules a pris un essor considérable au tournant des années 1970 notamment en raison du peaufinement de l'outil expérimental le plus important du domaine : le collisionneur de particules. Celui-ci utilise des champs électriques et magnétiques pour accélérer des particules chargées électriquement et les faire collisionner entre elles ou sur une cible. Ces collisions mènent à la création souvent brève de plusieurs autres particules qui passent ensuite au travers de détecteurs et qui produisent des données pouvant ensuite être analysées. C'est la possibilité d'avoir un très grand nombre de collisions à haute énergie focalisées en un seul point d'interaction qui a contribué à révolutionner le monde de la physique des particules. Cet outil permet de confirmer des théories existantes de la physique ou d'en découvrir d'autres en observant les résultats des collisions. Historiquement, c'est environ la moitié des avancées qui ont d'abord été théorisées, puis confirmées par l'expérience et l'autre moitié d'abord trouvées expérimentalement et théorisées ensuite.

De nombreux accélérateurs ont été développés au fil du temps et plusieurs ont mené à des découvertes importantes. Par exemple, le méson J/ψ , un méson constitué de quarks $c\bar{c}$, a été découvert indépendamment au *Stanford Linear Accelerator* [116] et le *Brookhaven National Laboratory* [93] en 1974 [83]. Cette découverte a révolutionné le monde de la physique des particules au point d'être référée à la *révolution de novembre du J/ψ* [83]. Nous avons aussi déjà vu qu'en 2012, le boson de Higgs a été découvert au LHC par les expériences ATLAS et CMS (voir section 1.3.4). En réalité, ce sont toutes les découvertes de particules qui sont passées par les accélérateurs de particules, ou presque [128]. On doit aussi aux collisionneurs la confirmation générale de la théorie du modèle standard, ainsi que la mesure très précise de plusieurs paramètres de celui-ci via d'innombrables expériences.

De tous les accélérateurs, c'est toutefois le *Grand collisionneur de hadrons*, abrégé LHC (*Large Hadron Collider*), qui est l'accélérateur le plus puissant. Inauguré en 2009, il possède désormais une énergie de collision de 13.6 TeV pour faire collisionner ensemble deux faisceaux de protons. Le LHC possède en son sein le détecteur *ATLAS*, dont les données et les simulations de données ont été utilisées dans le cadre de ce mémoire.

Le LHC est situé au complexe de l'*Organisation européenne pour la recherche nucléaire* (CERN) qui est montré à la figure 2.1. Le CERN est composé d'une collaboration internationale de plus de 70 pays et de plus de 600 instituts et universités qui contribuent à l'avancée de plusieurs projets phares de la science ou de la technologie. [40] Outre le LHC, le CERN est aussi à l'origine de l'invention d'« Internet » (*World Wide Web*) [32] ainsi que de multiples avancées en traitement, gestion, stockage et diffusion de données [31]. Ce chapitre s'intéresse au fonctionnement du LHC, mais aussi d'un de ses quatre détecteurs, ATLAS. Ensuite, nous analyserons comment les événements sont reconstruits grâce aux informations fournies par ATLAS.



Fig. 2.1. Vue à vol d'oiseau du complexe du CERN, situé à la frontière entre la Suisse et la France, à Genève. L'anneau principal du LHC de 27 km de circonférence est mis de l'avant dans cette figure. [24]

2.1. Le Grand Collisionneur de Hadrons

Le LHC est un immense accélérateur circulaire enfoui environ 100 mètres sous terre et d'une circonférence de 27 km situé à la frontière entre la Suisse et la France [37]. Le LHC fait principalement collisionner des faisceaux de protons, mais utilise aussi parfois des

noyaux de plomb [37].

Le LHC reçoit périodiquement des améliorations technologiques. En plus des maintenances de base, il se voit dédier des périodes prolongées d'arrêt lors desquelles l'accélérateur ainsi que ses détecteurs sont mis à jour avec des technologies plus puissantes. Le LHC a repris du service en juillet 2022 pour sa troisième phase d'opération (*Run 3*), après son deuxième long arrêt (*long shutdown 2*) qui s'est étalé de 2018 à 2022 [38].

2.1.1. L'accélération des particules

Comme les collisions proton-proton représentent la vaste majorité des collisions du LHC, tout ce qui suit ne traitera que de l'accélération des protons, ignorant l'accélération du plomb nécessaire aux collisions plomb-plomb.

Bien que les collisions se produisent dans l'anneau principal de 27 km de circonférence, les particules sont issues d'une série d'accélération au sein de plusieurs accélérateurs secondaires avant d'entrer dans cette section finale de leur parcours. Ce complexe d'accélération est présenté dans cette section et est illustré à la figure 2.2.

Tout d'abord, des atomes H^- , composés d'un proton et de deux électrons, sont envoyés dans le *Linac4* [25], un accélérateur linéaire. Ils y atteignent une énergie de 160 MeV avant d'être transmis au *Booster du Synchrotron à protons* [34]. C'est lors de cette transition que les électrons sont arrachés de leur noyau, créant alors un faisceau uniquement composé de protons. Ce booster est constitué de quatre anneaux synchrotron superposés. Dans ce booster, les protons sont amenés à une énergie de 2 GeV. Il n'est en réalité qu'une étape de préparation pour le *Synchrotron à protons* (SP) [42], qui est le prochain accélérateur de la chaîne. En effet, le booster permet d'augmenter l'énergie du faisceau avant son injection dans le SP, ce qui permet au SP d'accepter 100 fois plus de protons que s'il les recevait directement du Linac4.

Le SP a déjà été l'accélérateur principal du CERN, l'analogue du LHC actuel, et le plus puissant au monde à son inauguration en novembre 1959. Aujourd'hui, il sert surtout à produire un faisceau de 26 GeV pour alimenter le *Supersynchrotron à protons* (SPS) [36]. Le SPS est un accélérateur circulaire d'une circonférence de 7 km qui accélère le faisceau jusqu'à une énergie de 450 GeV. En 1981, il a servi de collisionneur proton-antiproton et a mené à la découverte de deux particules importantes du modèle standard, les bosons W et Z [36].

C'est finalement après le SPS que le faisceau est redirigé dans le LHC [29]. Ce dernier accélère les particules à l'aide de cavités radiofréquence (RF) [45], disposés tout au long de la trajectoire des protons et qui produisent dans leur intérieur un champ électromagnétique oscillant. Elles sont conçues pour assurer la résonance du champ pour en obtenir l'intensité la plus élevée. Compte tenu de l'aspect oscillatoire des cavités RF, il est important de synchroniser l'envoi de particules avec le champ. Ainsi, les protons sont envoyés en déphasage avec le champ en petits groupes appelés *paquets*, ce qui maximise le taux d'injection de particules dans les cavités. Notons que le déphasage se fait de façon naturelle : les particules qui ont une énergie ou une synchronisation inadéquate se voient ajuster leur trajectoire pour être en déphasage complet avec la cavité [45].

Finalement, les cavités RF ne servent qu'à augmenter l'énergie des particules, mais ne sont pas capables de les courber pour qu'elles puissent conserver un mouvement circulaire. Cette courbure se fait à l'aide de très puissants champs magnétiques produits par 1232 aimants supraconducteurs. Malheureusement, la production de champs d'une grande magnitude est limitée par des contraintes technologiques qui les restreignent à des valeurs de quelques teslas. De plus, atteindre ces valeurs représente un défi difficile et coûteux puisque les aimants doivent être refroidis en permanence à $1.9 K$ (kelvins) pour assurer leur supraconductivité [43, 28]. Finalement, si jamais le système de refroidissement fait défaut, l'aimant peut rapidement causer des dommages importants [44], imposant l'installation de plusieurs systèmes de sécurité robustes [122].

Ces aimants sont la cause principale de la limitation de l'énergie au LHC. S'il n'en était pas de cette contrainte, les cavités RF pourraient fonctionner presque indéfiniment, injectant de l'énergie à chaque paquet lors de leur passage tandis que les pertes d'énergie seraient minimales. Par contre, l'impulsion des particules du CERN est régie par l'équation 2.1.1, où p représente la quantité de mouvement des particules en GeV/c , ou en GeV dans le système d'unités naturelles, B le champ magnétique dans l'accélérateur en Teslas et R son rayon en mètres.

$$p = 0.3BR \tag{2.1.1}$$

Comme le rayon du CERN est fixé, nous voyons que l'impulsion maximale ne dépend que de l'intensité maximale du champ. C'est aussi pourquoi le LHC a une si grande circonférence : ceci limite le champ nécessaire à la courbure des particules. Finalement, d'autres aimants quadripolaires sont utilisés pour focaliser les faisceaux juste avant leur collision pour maximiser le nombre de collisions par croisement de paquet. En effet, plus le faisceau est mince, plus les probabilités de collision sont importantes qu'expliqué à la section 2.1.2.

The CERN accelerator complex Complexe des accélérateurs du CERN

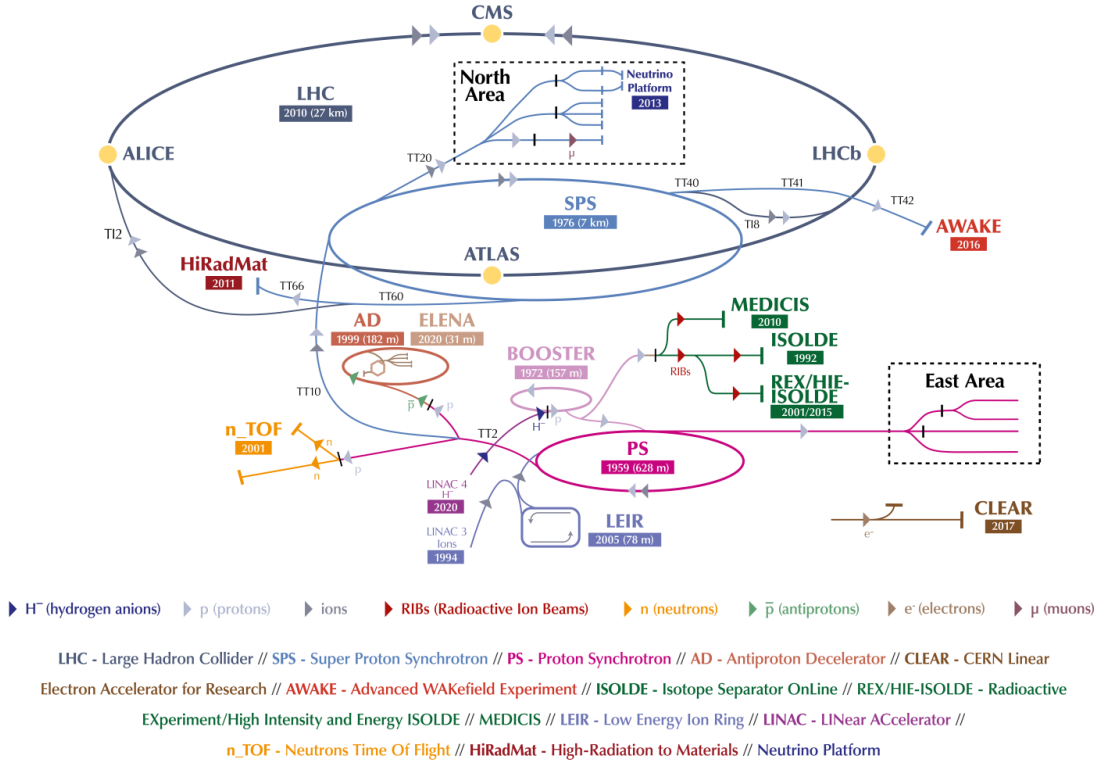


Fig. 2.2. Schéma du complexe d'accélérateurs au CERN. Outre la chaîne d'accélérateurs décrite dans la section 2.1.1, on y aperçoit aussi d'autres chaînes utilisées pour collisionner les ions de plomb dans le LHC, où les ions passent d'abord par le *Linac3* puis l'*Anneau d'ions de basse énergie* (LEIR) avant d'être envoyés au PS, au SPS puis au LHC. Cette figure démontre aussi toutes sortes d'autres expériences que les quatre principales du LHC, dont ISOLDE, MEDICIS, le décélérateur d'antiprotons et plusieurs autres [35].

2.1.2. La luminosité

La luminosité est un concept très important dans tout accélérateur. Elle permet de décrire le nombre de collisions par seconde par unité d'aire. Dans le cadre de collisions en un point de focalisation, qui possède une aire de collisions fixe, on peut la voir comme la fréquence de collision des particules accélérées. Dans un accélérateur de particules faisant collisionner deux faisceaux de façon frontale, la *luminosité instantanée* L est donnée à l'équation 2.1.2 [130, Chapitre 31] :

$$L = \frac{f_{coll} N_1 N_2}{4\pi\sigma_x\sigma_y}, \quad (2.1.2)$$

où f_{coll} représente la fréquence de collision des paquets (s^{-1}), N_1 et N_2 représentent le nombre de protons par paquet et le produit $\sigma_x\sigma_y$ représente l'aire du faisceau au point d'interaction. Notons que si nous avons n_b paquets par faisceau, alors $f_{coll} = n_b f_0$, où f_0 représente la fréquence de révolution des paquets. Nous voyons ainsi l'importance de focaliser les faisceaux pour créer le plus petit point d'interaction possible : plus $\sigma_x\sigma_y$ est petit, plus la luminosité augmente.

Le LHC est conçu pour obtenir une luminosité de $10^{34}/\text{cm}^{-2}\text{s}^{-1}$. La fréquence de collision est de 25 ns^{-1} et il y a 1.2×10^{11} protons par paquet et un total de 2808 paquets par faisceau. Finalement, σ_x et σ_y sont de l'ordre du μm .

Pour trouver la quantité totale d'événements en un laps de temps, par exemple en un an d'exploitation d'un accélérateur, il suffit d'intégrer la *luminosité instantanée* pour trouver la *luminosité intégrée totale* \mathcal{L} [73] :

$$\mathcal{L} = \int L dt . \quad (2.1.3)$$

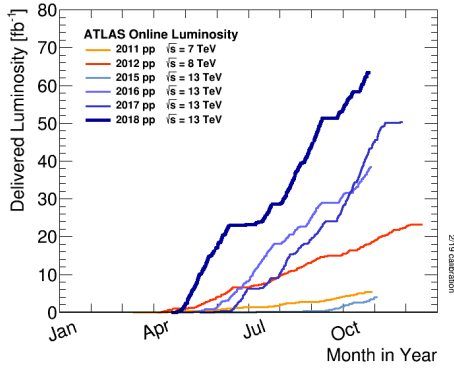
Enfin, pour quantifier le nombre d'événements d'un processus en particulier produits par seconde R_p , il faut multiplier la luminosité instantanée totale par σ_p , qui représente la *section efficace (cross-section)* de ce processus. Cette quantité σ_p possède des unités d'aire, varie grandement d'un processus à l'autre et peut tout simplement être interprétée comme la probabilité que ce processus ne se produise [73] :

$$R_p = \sigma_p L . \quad (2.1.4)$$

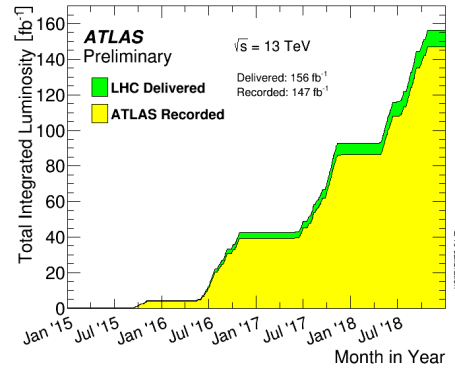
Ensuite, il va de soi que son nombre total d'événements pendant un certain laps de temps soit aussi donné par :

$$N_{p,tot} = \int R_p dt . \quad (2.1.5)$$

Tel que mentionné précédemment, l'avantage principal des accélérateurs est qu'ils permettent d'atteindre de très hautes luminosités, ce qui est crucial pour observer suffisamment d'événements de tous les processus rares de la physique (possédant un très petit σ_p) en suivant les équations 2.1.4 et 2.1.5. Par exemple, la création d'un boson de Higgs n'apparaît que dans un événement par milliard au LHC [30], donc il faut au bas mot plusieurs centaines, milliers ou millions de milliards d'événements totaux au LHC pour étudier ce phénomène (qui de plus, est très difficile à observer puisqu'il peut être absent de l'état final de la collision et possède un temps de vie de l'ordre de 10^{-22} secondes [27]). De plus, pour tenter d'observer de la physique non découverte qui est presque assurément encore beaucoup plus rare, et dont on ne connaît pas la forme, il faut encore plus d'événements.



(a) Luminosité intégrée année par année au LHC.



(b) Luminosité intégrée cumulative sur la période de fonctionnement *Run 2* du LHC.

Fig. 2.3. Luminosité intégrée selon le mois de l'année, année après année (gauche) et luminosité intégrée cumulative du LHC sur sa deuxième période d'opération *Run 2* (droite). La figure de gauche permet d'observer l'augmentation de la luminosité du LHC au fil des années, tandis que la figure de droite permet de mieux visualiser à quel point l'ensemble des données disponibles augmente dans le temps. Finalement, nous observons sur la figure de droite que seulement une petite portion des données n'a pu être enregistrée par l'expérience ATLAS [39].

En somme, la luminosité totale (et par ricochet la luminosité instantanée) forme l'une des plus importantes considérations pour un collisionneur. Il doit probablement se produire au moins des millions de milliards d'événements pour observer de la nouvelle physique. C'est pourquoi le LHC augmente continuellement sa luminosité instantanée avec toutes sortes de progrès technologiques. La figure 2.3 montre l'évolution de la luminosité intégrée du LHC au cours de son histoire.

2.1.3. Les expériences

Le LHC possède quatre détecteurs en son sein, tous situés à des endroits différents de l'anneau principal de 27 km. Ils forment aussi les quatre expériences principales du CERN, qui comporte plusieurs autres expériences d'envergure moindre. Outre *A Toroidal LHC Apparatus* (ATLAS) [54], il y a aussi le *Compact Muon Solenoid* (CMS) [58], le *Large Hadron Collider beauty* (LHCb) [60] et *A Large Ion Collider Experiment* (ALICE) [50].

ATLAS et CMS forment les deux plus importantes expériences du LHC, ayant des équipes de 6000 et de 4000 personnes, respectivement [52, 57]. Ils ont tous deux pour but de sonder les données du LHC dans leur intégralité et possèdent plusieurs objectifs, dont l'étude du boson de Higgs, de la matière sombre et de la nouvelle physique.

De leur côté, ALICE et LHCb sont des expériences plus spécialisées ayant des buts plus spécifiques. Tandis que le LHCb fait l'étude du quark b (*bottom*), ALICE se penche sur les collisions d'ions lourds dans le but d'étudier la force forte et le plasma quark-gluon.

Le détecteur ATLAS est celui qui a été utilisé dans ce projet de recherche et c'est donc lui qui sera décrit et analysé plus en détail dans les sections à venir.

2.2. Le détecteur ATLAS

Le détecteur ATLAS [51] est situé tout autour du point d'interaction des faisceaux et possède une forme cylindrique tel que vu à la figure 2.4. Mesurant 46 m de long et 25 m de diamètre pour 7000 tonnes, il est le plus grand détecteur d'un collisionneur de particules au monde. Plus de 1 milliard de collisions par seconde se font au point d'interaction situé au centre du détecteur. ATLAS est formé de plusieurs couches concentriques successives de détecteurs jouant tous un rôle particulier dans la prise de données. En effet, les différentes particules interagissent différemment avec la matière et sont donc traitées selon l'une des deux méthodes de *trajectométrie* et de *calorimétrie*. La trajectométrie tente d'interagir le moins possible avec les particules, ne recueillant que des petites traces d'énergie qu'elles auront laissé dans le détecteur. Celles-ci sont le signe de leur passage et sont utilisées pour tenter de reconstruire la trajectoire de ces particules. Comme le détecteur est soumis à un champ magnétique courbant les particules chargées, connaître leur trajectoire et leur courbure permet de déduire leur charge et leur quantité de mouvement. De son côté, la calorimétrie tente de récupérer toute l'énergie de la particule dans le détecteur pour la mesurer, mais détruisant ultimement la particule. Ainsi, les trajectomètres seront placés avant les calorimètres dans tout détecteur de collisionneur, dont ATLAS.

2.2.1. Les systèmes de coordonnées du détecteur

Pour localiser les particules et mesurer les distances entre celles-ci, deux systèmes de coordonnées sont employés dans le détecteur ATLAS. Le premier utilise des variables $(\hat{x}, \hat{y}, \hat{z})$ représentant le système de coordonnées cartésien habituel. Dans ce système, l'axe \hat{z} est orienté dans la direction de l'axe du faisceau, \hat{y} pointe vers le ciel et \hat{x} est dirigé vers le centre de l'anneau principal du LHC, tandis que l'origine est située au point d'interaction des protons. Ces coordonnées sont illustrées à la figure 2.4.

Puisque les deux faisceaux sont orientés selon l'axe \hat{z} , la quantité de mouvement totale de la collision doit être orientée selon ce même axe.¹ De cette manière, la somme des quantités de mouvement dans le plan (\hat{x}, \hat{y}) doit toujours être nulle pour chaque collision. Cette quantité importante est appelée quantité de mouvement transverse (ou impulsion transverse) et est dénotée p_T . Pour une particule donnée, on retrouve donc que :

$$p_T = \sqrt{p_x^2 + p_y^2} . \quad (2.2.1)$$

Toutefois, il est fort utile d'aussi introduire d'autres variables souvent plus pratiques en utilisant à notre avantage l'aspect cylindrique du détecteur et l'aspect relativiste des particules. Nous utiliserons $\hat{\phi}$, qui représente l'angle dans le plan (\hat{x}, \hat{y}) . Par contre, plutôt que de compléter notre système selon les simples coordonnées cylindriques habituelles $(\hat{r}, \hat{\phi}, \hat{z})$, nous introduisons les variables de rapidité y et de pseudo-rapidité η comme suit :

$$y = \frac{1}{2} \log \left(\frac{E + p_z}{E - p_z} \right) , \quad (2.2.2)$$

$$\eta = - \log \left[\tan \left(\frac{\theta}{2} \right) \right] . \quad (2.2.3)$$

Celles-ci sont utiles en milieu relativiste car les différences en y et en η sont invariantes sous une transformation de Lorentz.² Notons que ce n'est pas le cas de θ , qui représente l'angle dans le plan (\hat{z}, \hat{y}) et c'est pourquoi cette quantité n'est pas utilisée telle quelle dans le système de coordonnées. Ainsi, pour déterminer l'emplacement d'une particule, il est donc possible d'utiliser les trois variables (η, ϕ, r) , où r représente la distance de l'origine. Ce système est lui aussi montré à la figure 2.4.

Ces coordonnées, sous forme cartésienne ou simili-cylindriques, sont cruciales puisqu'elles aident à reconstruire les événements, que ce soit en créant des amas topologiques (*topoclusters*) de particules qui sont utilisées pour reconstruire les jets (voir section 2.4.1) ou en suivant la trajectoire des particules, pour ne nommer que ceux-là. Elles sont au coeur d'un nombre incalculable de variables et d'algorithmes utilisés dans ATLAS.

¹ N'oublions pas que les protons sont composés essentiellement de quarks et de gluons. Puisque chaque collision ne fait intervenir qu'une seule particule de chaque proton et que la quantité de mouvement de ces particules est inconnue dans un cadre relativiste, il en survient que l'impulsion totale n'est pas nulle en \hat{z} , voir section 2.5.1.

² La pseudo-rapidité n'est qu'une excellente approximation de la rapidité à haute énergie. Ainsi, c'est simplement dans des milieux relativistes que les intervalles de pseudo-rapidité sont un invariant de Lorentz, ce qui est le cas au LHC. L'avantage de cette variable est qu'elle n'est qu'angulaire.

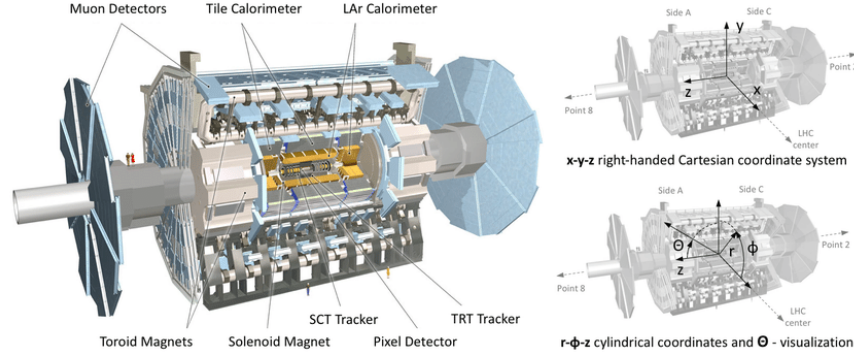


Fig. 2.4. Schéma représentant les deux systèmes de coordonnées au sein du détecteur ATLAS. Le système cartésien comporte l'axe \hat{z} pointant dans la direction du faisceau, l'axe \hat{x} pointant vers le centre du LHC et l'axe \hat{y} pointant vers le haut. Le système pseudo-cylindrique montre l'angle η qui dépend de l'angle θ , qui lui-même représente l'angle entre l'axe \hat{z} et \hat{y} , tel que démontré à l'équation 2.2.3. Finalement, ce système de coordonnées montre ϕ , qui représente l'angle dans le plan (\hat{x}, \hat{y}) et r , qui représente la distance de la particule par rapport au point d'interaction des protons [92].

2.2.2. Le détecteur interne

Le détecteur interne constitue le premier contact avec l'appareillage d'ATLAS qu'ont les particules après une collision. Il est composé de trois parties distinctes qui ont toutes des rôles particuliers. C'est dans l'entière du détecteur interne que le champ magnétique de 2 T est envoyé pour courber les particules chargées dans le but d'en mesurer plus précisément leur énergie et leur quantité de mouvement.

Tout d'abord, le détecteur à pixels forme la première étape du détecteur interne. Il n'est situé qu'à 3.3 cm du point d'interaction des faisceaux et est composé de 4 couches de détecteurs en silicium à pixels. Son but premier est de déterminer l'origine et la quantité de mouvement des particules. Ainsi, son rôle ainsi que sa proximité du point d'interaction en requièrent une précision extrême. C'est pourquoi le détecteur est excessivement compact et contient plus de 92 millions de pixels d'une taille de $50\mu\text{m} \times 250\mu\text{m}$ (pixels les plus proches du point d'interaction) à $50\mu\text{m} \times 400\mu\text{m}$ (pixels éloignés) chacun couvrant au total 1.9m^2 autour du point d'interaction. La position des traces laissées par les particules dans ces détecteurs sont précises de l'ordre de $10\ \mu\text{m}$. Ensuite, le trajectomètre semi-conducteur entoure le détecteur à pixels. Il consiste en 4 couches cylindriques de capteurs à micro-bandes en silicium d'une taille de $80\mu\text{m} \times 12\text{cm}$ chacune. En traversant les 4 couches, les particules laissent 4 indices supplémentaires de leur passage qui sont toutes analysées et reconstruites pour mieux comprendre la trajectoire complète des particules.

Ces deux premiers paliers fonctionnent avec une technologie similaire. En effet, en passant au travers des détecteurs, les particules chargées ionisent la matière du détecteur, c'est-à-dire qu'elles libèrent des électrons de leur noyau. On soumet aussi les pixels ou les micro-bandes de silicium à des voltages importants de l'ordre de la centaine de volts qui permettent d'amplifier le signal. Ainsi, nous assurons que les électrons frappés se rendent jusqu'à l'électrode où ils seront détectés.

Finalement, le détecteur à radiation de transition complète la structure du détecteur interne. Plus de 300,000 petits tubes en forme de paille de 4 mm de diamètre et de 39 ou 144 cm de long sont disposés de façon cylindrique, soient des pailles parallèles au faisceau (baril) ainsi que deux disques (embouts) perpendiculaires au faisceau. Des petits fils de 0.03 mm de diamètre de tungstène plaqués en or sont au centre de chaque tube pour faciliter le transport des électrons vers l'électrode à l'extrémité du tube. Cette étape permet de poursuivre la trajectoire des particules déjà détectées dans les parties antérieures du détecteur interne mais aussi d'en savoir plus sur le type de particule dont il s'agit en utilisant l'information de radiation de transition. En effet, lorsqu'une particule change de milieu, par exemple dans l'espace entre deux tubes, elle produit un rayonnement de transition qui consiste en un photon d'énergie typiquement associée aux rayons X. Ce rayonnement produit parfois des signaux plus importants que le passage de la particule elle-même dans les tubes. Ce phénomène est utilisé pour identifier les électrons et les distinguer des pions π , un type de mésons, via leurs ratio énergie/masse respectifs. En effet, puisque le pion π est près de 265 fois plus massif que l'électron, il possède une probabilité beaucoup plus faible d'engendrer un rayonnement de transition à une énergie donnée [20].

Ici aussi, la détection des particules se fait via l'ionisation du matériau du détecteur, amplifiée via l'imposition d'un haut voltage puis finalement recueillie sur des électrodes. Par contre, cette fois, le matériau à ioniser est du gaz, ce qui crée une différence cruciale. En effet, le voltage permet cette fois-ci d'augmenter l'énergie des électrons éjectés pour qu'ils ionisent à leur tour d'autres électrons, et ainsi de suite, créant une cascade d'électrons éjectés. Ceci permet en outre d'augmenter le signal d'un facteur 10^5 .

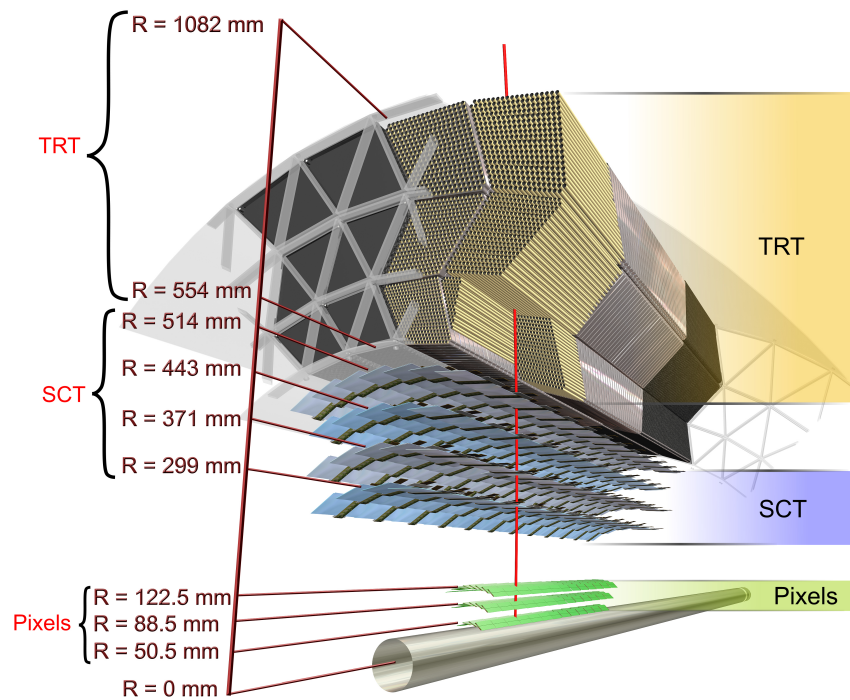


Fig. 2.5. Schéma du détecteur interne. On y aperçoit le faisceau de particules tout au bas de la figure, suivi du détecteur à pixels, du trajectomètre semi-conducteur et enfin du trajectomètre à radiation. [102]

2.2.3. Les calorimètres

La calorimétrie est une technique consistant à créer des gerbes de particules puis de les mesurer. En effet, le principe de base veut qu'une particule hautement énergétique incidente dans le calorimètre interagisse avec celui-ci, puis crée d'autres particules d'énergie moindre, celles-ci interagissant ensuite encore pour créer des particules d'énergie encore moindre, et ainsi de suite. On parle aussi d'effet cascade ou, en anglais, de *shower* (douche). Une fois les particules rendues trop peu énergétiques, elles sont arrêtées et absorbées dans le calorimètre. Ainsi, le calorimètre devrait en principe arrêter toute particule produite, sauf les muons (voir section 2.2.4), les neutrinos (voir section 1.1) et peut-être d'autres particules encore inconnues à l'humain comme la matière sombre.

Pour ce faire, les calorimètres sont composés de couches successives de matériau passif et de matériau actif. Les couches passives favorisent l'interaction des particules incidentes avec le calorimètre et produisent les gerbes. Les couches actives mesurent l'énergie des particules

en les absorbant et en minimisant toute autre interaction avec celles-ci. Cela permet aussi d'identifier les particules en analysant leurs dépôts d'énergie, que ce soit via leur quantité ou leur géométrie.

Il existe deux grands types de gerbes : les gerbes électromagnétiques et les gerbes hadroniques. Tandis que les premières sont causées majoritairement par la production de paires et du rayonnement bremsstrahlung (voir section 1.3.2), les deuxièmes sont en grande partie constituées de hadrons. Il faut noter que les cascades électromagnétiques sont beaucoup plus « propres » que les cascades hadroniques puisque la chromodynamique quantique (QCD) engendre beaucoup plus de fluctuations. D'ailleurs, les calculs théoriques reliés à la QCD sont en ce moment impossibles à réaliser en vertu de la constante de couplage α_s qui explose aux petites énergies. Nous devons donc nous contenter d'ajuster des modèles de gerbes hadroniques avec les données des expériences, sans pouvoir calculer explicitement les effets de la QCD. Les deux types de cascades nécessitent des détecteurs distincts, nommés le calorimètre électromagnétique et le calorimètre hadronique, qui sont tous deux représentés à la figure 2.6. Tandis que le calorimètre électromagnétique permettra de détecter les photons, les électrons et les positrons via la force électromagnétique, le calorimètre hadronique permettra de détecter les hadrons neutres et chargés.

Le calorimètre électromagnétique comporte un baril et deux embouts qui sont respectivement dénotés par *LAr electromagnetic barrel* et *EMEC* sur la figure 2.6. Tel que son nom le suggère, le baril est composé d'argon liquide qui constitue son matériau actif, tandis que son matériau passif est le plomb. Similairement à un trajectomètre, les particules absorbées dans la partie active sont converties en signal électrique qui sera ensuite utilisé pour détecter leur présence [55].

Finalement, le calorimètre hadronique est aussi composé d'un baril (*Tile barrel*), mais aussi d'un baril étendu (*Tile extended barrel*), de deux embouts (HEC) et de deux parties avant (*FCal*), où les dénominations entre parenthèses représentent leur appellation visible à la figure 2.6. Dans les barils, le matériau passif est l'acier tandis que des scintillateurs en tuile de plastique forment le matériau actif. Des tubes photo-multiplicateurs sont capables de capter l'énergie des cascades en recueillant et en mesurant la lumière produite par celles-ci. Enfin, le HEC et le FCal utilisent de l'argon liquide comme matériau actif, le premier sous forme de plaques et le deuxième sous forme de tubes. Ils utilisent aussi des plaques de cuivre et un mélange de cuivre-tungstène comme matériau passif, respectivement. Les parties avant (FCal) servent à détecter les particules qui ont un petit angle de 1 à 5 degrés par rapport

au faisceau et doivent donc être capables de soutenir une radiation très intense, justifiant l'utilisation de tubes plutôt que de plaques puisqu'ils résistent mieux à la radiation [23].

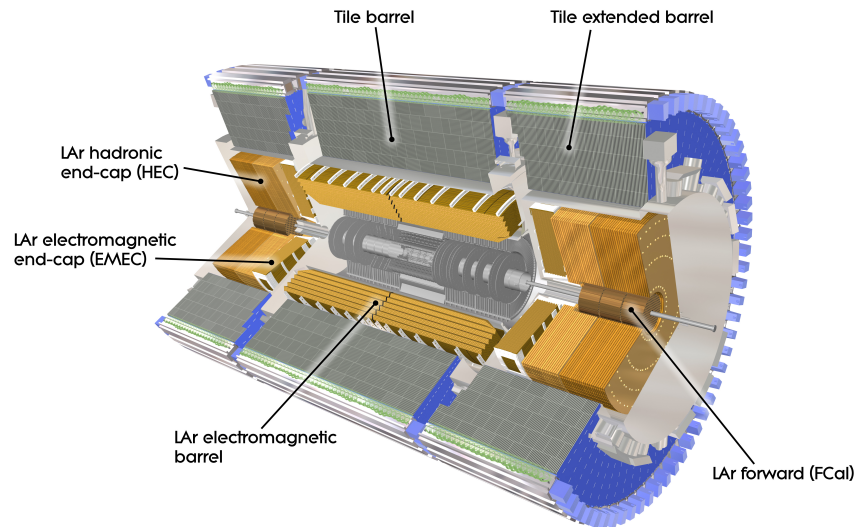


Fig. 2.6. Schéma des calorimètres d'ATLAS. Le baril d'argon liquide (*LAr electromagnetic barrel*) ainsi que les deux disques EMEC constituent le calorimètre électromagnétique. D'autre part, le calorimètre hadronique inclut le baril *Tile barrel* et le baril *Tile extended barrel* en plus des deux disques HEC et de la partie avant FCal [101].

2.2.4. Le spectromètre à muons

Le spectromètre à muons est la dernière couche du détecteur ATLAS. Comme son nom l'indique, il est spécialisé pour détecter les muons puisque ceux-ci sont autrement très difficiles à observer. En effet, puisque les muons sont chargés négativement (antimuons chargés positivement), ils vont laisser une trace dans le détecteur interne. Par contre, les calorimètres auront de grandes difficultés à détecter leur passage. En effet, introduisons tout d'abord le pouvoir d'arrêt d'un matériau sur une particule qui est une mesure de la quantité d'énergie moyenne perdue par ionisation lorsque cette particule traverse le matériau. Typiquement, cette quantité varie selon l'énergie de la particule traversant le matériau. Malheureusement, le pouvoir d'arrêt du cuivre (qui est le matériau principal des calorimètres, voir section 2.2.3) sur les muons est à son minimum aux énergies typiques des muons au LHC, soit de l'ordre du GeV. Il faut donc une autre stratégie pour parvenir à les identifier, et c'est pourquoi les muons justifient à eux seuls cette dernière couche de détecteurs.

Le spectromètre est en réalité un trajectomètre à muons. Il se base sur les mêmes principes d'ionisation des atomes du détecteur pour retracer la particules sur plusieurs mètres. Cette fois-ci, les muons ionisent un mélange d'argon et de dioxyde de carbone dans les *tubes à dérive contrôlée*, un amas de petits tubes entassés les uns sur les autres. L'anode au centre du tube transportant les électrons et les ions (voir section 2.2.2) est encore en tungstène-rhénium et plaqué en or. Ceci permet de calculer la trajectoire des muons au travers du détecteur, qui possède de trois à huit couches de ces tubes à dérive.

Le spectromètre est également soumis à un champ magnétique toroïdal créé par un système de trois aimants en forme de cylindre, comprenant les deux embouts et le « contour » du cylindre, le baril (*Barrel toroid*). Tout comme le détecteur interne, le but est de mesurer l'impulsion des particules, en l'occurrence des muons, en calculant leur courbure. Cette mesure est indépendante de celle du détecteur interne et pourra nous servir pour associer les traces du détecteur interne à celles du spectromètre.

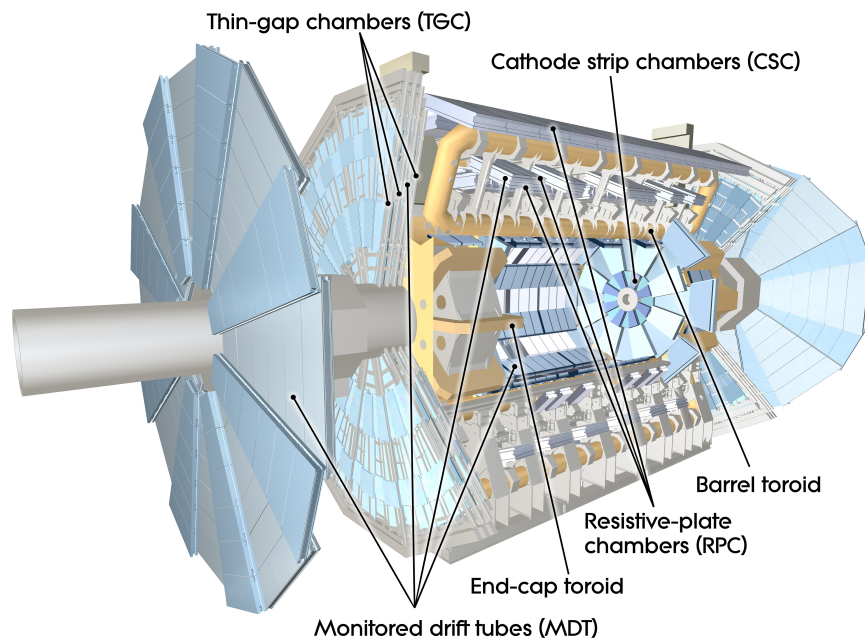


Fig. 2.7. Schéma du spectromètre à muons. On y voit les aimants toroïdes engendrant le champ magnétique dans le spectromètre, les tubes à dérive (MDT) ainsi que les autres systèmes de déclenchement de l'enregistrement des événements (RPC, CSC, TGC) [103].

2.3. Le système d'acquisition des données

Comme le LHC produit un milliard de collisions par seconde dans ATLAS, qui produit chacune plusieurs centaines ou milliers de particules, il est impossible de conserver toute l'information recueillie par le détecteur. En effet, celle-ci représente plus de 60 TB/s qui doit

non seulement être récoltée, sauvegardée et acheminée à des serveurs, mais aussi stockée. Il est impossible de maintenir un taux si élevé, et la vaste majorité des événements doit être éliminée. Ainsi, seulement 1000 événements par seconde seront conservés au LHC, ce qui représente une collision par million. Celles-ci doivent avoir été identifiées comme potentiellement intéressantes à des fins d'analyse par le *système d'acquisition des données* (SAD). Ce dernier doit être absolument robuste, puisque son tri rejette de façon permanente 99.9999% de ce qui est produit au sein d'ATLAS.

Heureusement, des événements inintéressants représentent une énorme proportion des événements produits. De façon générale, les événements importants à analyser sont plus rares, puisqu'ils requièrent généralement des objets à très haute énergie, notamment des objets boostés (grande quantité de mouvement) ou très massifs, ou les deux. Les événements laissés de côté sont donc souvent déjà très communs et amplement analysés.

Le SAD est composé de deux déclencheurs, soient le déclencheur du premier niveau (*Level 1 Trigger*) et le déclencheur de haut niveau (*High Level Trigger*). Ensemble, ils prennent environ 200 microsecondes (μs) pour trier chaque événement comme étant à conserver ou à supprimer.

2.3.1. Déclencheur de premier niveau

Cette première partie du SAD est faite de composantes électroniques (*hardware*) hyper-spécialisées situées à même le détecteur. Ces composantes sont capables de sonder l'information recueillie dans l'espoir de trouver des objets physiques reconnus pour être intéressants à des fins d'analyse, soient des leptons, c'est-à-dire des électrons, des muons (μ) à haute énergie, des taus (τ), des jets à haut p_T .

Cette étape permet de réduire la demande de stockage d'un facteur de 10^4 , puisque le déclencheur de haut niveau ne peut accepter que 100,000 événements par seconde au maximum. Pour chaque événement, la décision est prise en $2.5\mu\text{s}$ [53].

2.3.2. Déclencheur de haut niveau

Ce dernier niveau du SAD vise à polir le tri effectué par le premier niveau avec des calculs beaucoup plus longs à effectuer. En effet, un superordinateur composé d'un parc d'ordinateurs applique toutes sortes d'algorithmes à l'aide de logiciels hautement sophistiqués aux données qui englobent cette fois-ci toute l'information captée, incluant le détecteur interne. Cette fois-ci, les données sont classées en $200\mu\text{s}$, soit 80 fois plus que le temps requis au premier niveau tandis qu'il y a 10,000 fois moins d'événements à traiter au

total. Il est donc plus facile de comprendre pourquoi il était nécessaire de faire un premier tri, sinon il y aurait tout simplement trop d'information à traiter en temps réel pour ce déclencheur.

Le déclencheur de haut niveau sélectionnera finalement environ 1000 événements par seconde à conserver qui seront tout d'abord reconstruits et puis stockés dans la banque de données du CERN. Les analystes peuvent ensuite consulter cette banque et peuvent utiliser divers critères des niveaux de déclenchement pour ne sélectionner qu'une partie de tous les événements sauvegardés afin de peaufiner leur analyse. Par exemple, ils peuvent n'analyser que des événements comportant un muon ainsi que trois électrons.

2.4. La reconstruction d'événements

En plus de récupérer et mesurer les dépôts d'énergie des particules traversant le détecteur ATLAS, il faut aussi tenter d'identifier la nature des particules des collisions en plus de reconstruire l'événement dans son ensemble. Par exemple, il peut s'agir de déterminer quelles sont les particules initialement créées par la collision proton-proton ou s'il y a eu la présence d'un boson de Higgs, d'une particule à très haute énergie ou n'importe quelle autre particule d'intérêt à un certain moment dans le détecteur. Pour ce faire, divers algorithmes sont appliqués pour reconstruire diverses particules, comme par exemple les électrons, les photons ou les muons. Nous tentons aussi de détecter l'énergie manquante due à des particules invisibles dans le détecteur. Puisque notre projet de recherche se concentre sur les jets, je décrirai dans cette section comment ils sont reconstruits au sein du détecteur ATLAS.

2.4.1. Les jets hadroniques

L'idée générale de la reconstruction des jets, déjà présentés comme des flots collimatés de particules à la section 1.3.6, est de tenter d'identifier, de reconstruire et de regrouper les constituants des jets – définis comme les particules stables composant un jet. La reconstruction offre aussi de trouver les quadrivecteurs de ces constituants, tandis que le regroupement permet de savoir lesquels font partie d'un même jet. Pour ce faire, nous tentons d'associer des amas topologiques (*topological clusters* ou *topoclusters*), contenus dans l'information des cellules des calorimètres, aux constituants des jets. Dans un monde idéal, chaque amas topologique représente un constituant du jet.

Les premières cellules composant un amas sont d'abord sélectionnées comme étant celles les plus énergétiques, soient celles qui possèdent une plus grande énergie que le bruit de fond, qui est défini comme étant la somme quadratique du bruit relié à l'électronique et

du *pile-up*.³ Plus précisément, nous demandons un écart statistique de 4σ ou plus par rapport à ce bruit. Ensuite, les cellules voisines sont itérativement ajoutées à l'amas si elles dépassent d'au moins 2σ la valeur du bruit de fond. Puis, une fois qu'il n'existe plus de cellules respectant cette condition, un anneau final est ajouté aux amas, qui ont maintenant pris leur forme finale [131, 71]. Cet algorithme est schématisé à la figure 2.8, qui montre étape par étape la formation d'un amas topologique dans le calorimètre avant (FCal).

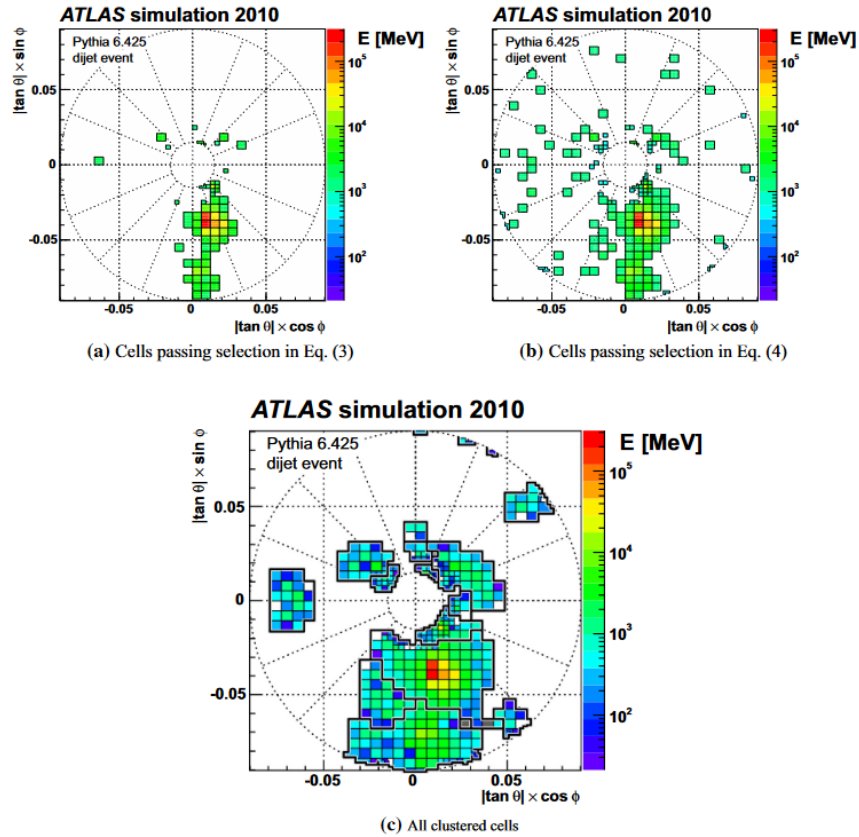


Fig. 2.8. Formation des amas topologiques dans le calorimètre hadronique avant (FCAL) servant à la reconstruction des jets. On y aperçoit les étapes de la formation de ces amas, où les cellules plus énergétiques sont sélectionnées (en haut à gauche), auxquelles s'ajoutent d'autres cellules avoisinantes qui possèdent une énergie suffisante et des coordonnées spatiales suffisamment rapprochées de l'amas (en haut à droite). Finalement, l'image du bas représente l'étape où un anneau est formé autour des cellules sélectionnées précédemment pour terminer la formation de l'amas [2].

³ Le *pile-up* représente le bruit relié à un autre événement que celui engendré par la collision proton-proton initiale. En effet, il peut y avoir plus d'une collision par croisement de paquets au point d'interaction, et ces événements seront alors superposés dans le détecteur. Toutes sortes de techniques et d'algorithmes existent pour réduire l'importance du *pile-up* en séparant les informations recueillies dans le détecteur selon leur collision d'origine.

L'algorithme du flot des particules (*particle flow algorithm*) est ensuite appliqué. Cette technique permet d'enlever les chevauchements entre les mesures de quantité de mouvement et d'énergie faites dans le détecteur internes et les calorimètres en soustrayant les valeurs d'énergie cellule par cellule. Ultiment, cet algorithme fournit une liste des traces et des amas topologiques qui contiendra à la fois les amas non modifiés et un nouvel ensemble d'amas modifié obtenu avec la technique de soustraction d'énergie des cellules [71]. D'abord, l'algorithme tente d'associer chaque trace du détecteur interne avec un seul amas topologique des calorimètres. Ensuite, l'algorithme tente d'estimer la quantité d'énergie attendue dans le calorimètre en se basant sur la trace ainsi que la position de l'amas. Puisqu'il est commun d'avoir une particule qui dépose de l'énergie dans de multiples amas, l'algorithme évalue la probabilité que cela se soit produit, puis décide s'il est nécessaire d'ajouter plus d'amas topologiques pour mieux capter toute l'énergie de la cascade hadronique. Finalement, l'énergie attendue dans le calorimètre est soustraite cellule par cellule à l'ensemble des amas correspondants. En dernier lieu, si l'énergie restante dans le système est conforme avec les fluctuations attendues de l'une particule de la cascade, les restes de la soustraction de l'amas sont supprimés. Dans ce cas, il est compris que l'énergie restante origine des fluctuations de la cascade et que l'amas a été produit par une seule particule. Dans le cas contraire, les restes de la soustractions sont conservés, et il est probable que de multiples particules ont déposé de l'énergie dans cette région. Ainsi, l'ensemble des traces et des amas topologiques restants tente de représenter une reconstruction qui ne compte pas deux fois l'énergie de particules entre les traces et les amas.

La prochaine étape est d'assembler ces amas (particules) pour en former un jet. Il existe une panoplie de méthodes pour accomplir cette tâche, mais nous n'analyserons que celles issues de la grande famille des algorithmes de regroupement (*clustering algorithms*), puisqu'elle est robuste et propose de bons résultats dans une variété d'événements.⁴

Les algorithmes de regroupement les plus utilisés sont les algorithmes de k_T , de Cambridge et Aachen et d'anti- k_T . Ils définissent d'abord une distance d_{ij} et une séparation ΔR_{ij} entre deux particules i et j en plus d'une distance d_{iB} calculée entre la particule i et le faisceau comme suit :

$$\begin{aligned}
 d_{ij} &= \min \left(p_{Ti}^{2p}, p_{Tj}^{2p} \right) \frac{\Delta R_{ij}^2}{R^2} \quad , \\
 \Delta R_{ij}^2 &= (y_i - y_j)^2 + (\phi_i - \phi_j)^2 \quad , \\
 d_{iB} &= p_{Ti}^{2p} \quad .
 \end{aligned}
 \tag{2.4.1}$$

⁴ Une autre grande famille existe, soit les algorithmes de cône, mais elle ne permet pas d'avoir de bons résultats dans diverses situations, notamment lorsqu'une particule produit deux jets distincts [111].

Ici, p_T représente la quantité de mouvement transverse d'une particule, y sa rapidité et ϕ sa position azimutale. La valeur de la variable R est définie manuellement dans l'algorithme avant de le performer et représente un cône de rayon R dans le plan $(\eta \times \phi)$.

Ces trois algorithmes utilisent les équations 2.4.1 en variant la valeur de p , qui prend la valeur de 1 pour l'algorithme de k_T inclusif, 0 pour l'algorithme de Cambridge et Aachen et -1 pour l'algorithme d'anti- k_T . L'algorithme de Cambridge et Aachen est utile pour étudier les jets de sous-structure complexes comportant deux ou trois sous-jets puisqu'il reconstruit les jets de manière plus irrégulière [67, 111]. De son côté, l'algorithme anti- k_T sélectionne des particules plus énergétiques et ses jets sont donc plus réguliers et plus circulaires. Des exemples de reconstruction via ces différents algorithmes sont montrés à la figure 2.9.

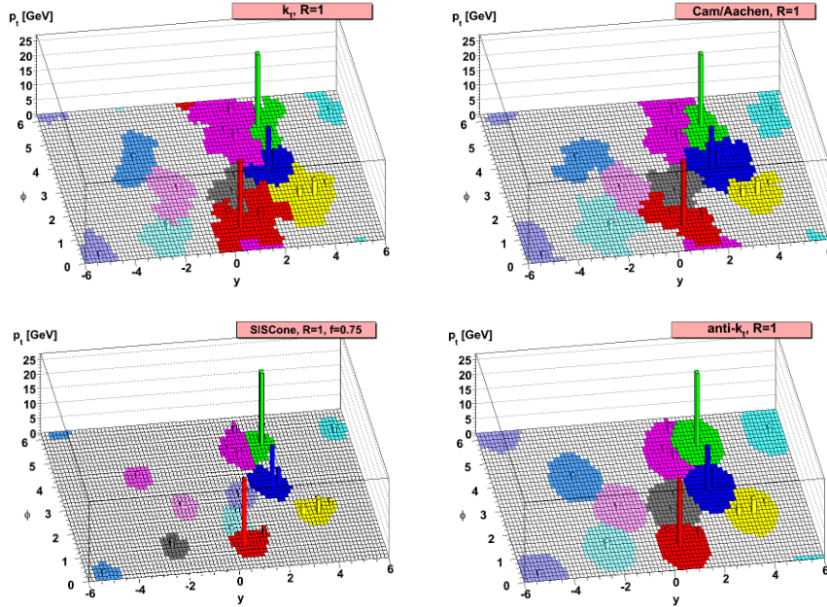


Fig. 2.9. La reconstruction d'un même jet par différents algorithmes. L'algorithme anti- k_T (en bas à droite) reconstruit le jet de façon régulière et circulaire, tandis que celui de Cambridge et Aachen (en haut à droite) le reconstruit de façon plus irrégulière. L'algorithme de k_T inclusif (en haut à gauche) offre une reconstruction dont la régularité se situe entre les deux autres algorithmes [22].

Or, dans notre projet, nous concentrons nos recherches sur les jets boostés comportant souvent plusieurs sous-jets et possédant un grand rayon $R = 1.0$. Au sein d'ATLAS, il est coutume de reconstruire des jets avec un rayon de 0.4, ou des jets à grand rayon qui possèdent un rayon de $R = 1.0$. Nous avons utilisé l'algorithme d'anti- k_T pour reconstruire ces jets à grand rayon à partir de l'information issue de l'algorithme du flot de particules. En

augmentant le rayon de cette manière, il faut s’assurer d’ajouter une étape supplémentaire d’élagage (*trimming*). Elle se résume à recombinaison des jets formés en utilisant un rayon plus petit ; ce faisant, nous identifions des sous-jets qui sont ensuite conservés et assemblés s’ils dépassent un certain niveau d’énergie [111]. L’élagage réduit la contamination des événements de plusieurs sources, notamment du bruit de fond et du *pile-up* puisqu’il permet de s’assurer que chaque petit sous-jet est très énergétique et donc plus probablement issu de la collision de partons initiale [91].

Finalement, l’énergie des jets est calibrée en utilisant des processus bien connus tels que $Z + \text{jet}$, $\gamma + \text{jet}$ et $\gamma + \text{deux jets}$. Une correction liée au *pile-up* est aussi appliquée en se basant sur des simulations Monte Carlo [1, 131].

2.5. La simulation d’événements

En plus de l’enregistrement des événements produits au LHC dans le détecteur ATLAS, il existe plusieurs méthodes pour simuler des événements simplement à partir de logiciels et sans avoir à collisionner des particules. Celles-ci peuvent servir à augmenter le nombre d’événements à étudier de certains processus rares qui doivent leur rareté à leur petite section efficace (voir section 2.1.2). Cela se produit la plupart du temps à leur configuration atypique, comme par exemple via la présence de particules très énergétiques ou encore de par la création d’une ou plusieurs particules rares. Les simulations permettent aussi de s’assurer de produire des événements qui sont strictement à l’intérieur ou à l’extérieur du modèle standard. Finalement, celles-ci simulent la collision initiale des constituants des protons (appelés *partons*), puis toutes les désintégrations successives suivies par chaque particule produite, ainsi que les interactions avec le détecteur ATLAS, que ce soit leur production de gerbes hadroniques (voir section 2.2.3), leur résolution ou n’importe quel autre phénomène. Les simulations s’occupent aussi de reconstruire les jets, tel que vu à la section 2.4.1, et créent aussi plusieurs outils d’analyse. Par exemple, il existe des filtres qui ne sélectionnent que certains événements selon toutes sortes de propriétés : nombre de leptons, p_T minimal du jet le plus énergétique et des milliers d’autres.

Bien que la collaboration derrière l’expérience ATLAS ait conçu des outils très performants et très exacts de la simulation des interactions avec le détecteur, qui représente l’étape la plus compliquée de la simulation, il existe aussi d’autres programmes publics qui troquent de la précision pour une plus petite demande de ressources et une rapidité de simulation. En effet, les logiciels de l’expérience ATLAS prennent une dizaine de minutes [18] pour ne produire qu’un seul événement, tandis qu’il est possible de générer des centaines de milliers, voire des millions, d’événements en une heure pour ces logiciels tiers. La simulation

d'événements officiels ATLAS représente d'ailleurs un beau défi, et elle se fait dans des centaines de centres de calcul répartis un peu partout dans le monde proposant au total plusieurs millions d'heures de calcul par jour [18].

Dans le cadre d'un usage de l'apprentissage machine pour étudier les collisions au LHC, ces simulations peuvent être utiles pour s'assurer de n'entraîner un modèle que sur des événements connus et compris. Elles sont aussi utilisées pour créer des données qui ne sont pas incluses dans le modèle standard dans le but de simuler de la nouvelle physique et de tester si les modèles sont assez robustes pour la découvrir (voir chapitre 3). Nous pouvons aussi entraîner un modèle sur des données qui ne présentent pas une certaine particule, par exemple le quark top, puis exposer le modèle à des événements comportant le quark top pour savoir s'il est capable de le découvrir, en simulant un environnement de découverte de nouvelle physique. Pour simuler des événements en dehors du modèle standard, il est possible de changer la masse de certaines particules connues ou de modifier les couplages qu'elles possèdent avec d'autres particules. Il est aussi possible de simplement créer un nouveau modèle théorique avec de nouvelles particules, spécifier leurs propriétés et simuler les événements qui en découlent.

Toutefois, puisque notre projet s'inscrit dans une optique de recherche et de développement rapide où les résultats sont en constante évolution, nous avons priorisé d'utiliser des logiciels permettant une génération rapide des événements au coût d'une perte de précision sur ces échantillons. De plus, dans notre projet, il a été nécessaire de générer divers types d'échantillons de plusieurs configurations différentes couvrant un large spectre en deux dimensions de masse et de p_T (voir sections 4.3 et 4.3.3) pour les utiliser en tant que *données hors distribution*. De plus, nous avons dû générer à la fois du bruit de fond, des données hors distribution et du signal en plus de modifier certains paramètres du modèle standard tout en créant des événements de processus physiques n'existant pas (voir section 4.3.3), justifiant également cette décision. C'est pourquoi, pour la partie traitant des interactions avec le détecteur, je me concentrerai sur la description des simulateurs publics plutôt que les simulateurs directement issus d'ATLAS.

2.5.1. Les sous-étapes de la simulation

La simulation se fait souvent en trois sous-étapes qui sont effectuées par trois logiciels distincts. Dans notre projet, ces trois étapes sont la simulation de la collision parton-parton et le calcul des éléments de matrice en découlant (effectuée par **MadGraph** [7]), la simulation des cascades de partons (effectuée par **Pythia8** [19]) et la simulation de la réponse du

détecteur (effectuée par **Delphes** [64]).

La première étape consiste au calcul du résultat de la collision initiale des partons, c'est-à-dire l'état du système immédiatement après la collision proton-proton via le calcul des éléments de matrice. Ici sont donc calculés les diagrammes de Feynman de la collision, qui proposent souvent des quarks et des gluons dans l'état final. Entre autres, cette sous-étape se termine tout juste avant le calcul des cascades de partons (*parton showers*) qui forment les jets. Nous nous intéressons entre autres aux particules créées ainsi que de leur énergie et leur quantité de mouvement. Cette étape n'est pas triviale, car le proton est constitué de deux quarks up et d'un quark down, en plus des différents gluons qui donnent l'énergie de liaison de ces particules. Toutefois, lors de la collision proton-proton, ce n'est qu'un parton par proton qui entre en collision. On peut donc assister à une collision quark-quark, quark-gluon ou gluon-gluon. Il est toutefois important de savoir que la force forte peut momentanément faire apparaître des *quarks de la mer*, des quarks qui existent environ 10^{-24} secondes avant de disparaître. Ceux-ci ne sont pas les constituants « officiels » du proton, mais sont bien réels puisque lors d'une collision au LHC, il est probable que la collision implique un ou deux quarks de la mer. Notons qu'il est donc possible de collisionner des quarks qui ne sont ni des up, ni des down grâce à ce phénomène et qu'il devient aussi possible de collisionner des antiquarks.

Une autre difficulté de ce calcul réside dans le fait que l'énergie et l'impulsion de chaque parton ne sont pas connues d'avance puisque chaque parton transporte une fraction probabiliste de la quantité de mouvement totale du proton. Ainsi, la collision peut être plus ou moins énergétique selon les partons impliqués. Heureusement, il existe les fonctions de distribution des partons (PDF), représentant la densité de probabilité de trouver une particule donnée à une énergie (quantité de mouvement) donnée dans un proton d'énergie Q . Le plus souvent, elles sont données par x , la fraction de l'énergie totale du proton transportée par le parton. Les PDF ne peuvent pas être calculées théoriquement et doivent donc être trouvées expérimentalement. Ces fonctions sont utilisées dans les simulateurs pour bien reproduire la probabilité de collision des différents constituants des protons à énergie donnée. Notons que les PDF varient en fonction de l'énergie du proton (par exemple, si l'énergie du proton double, on ne peut pas simplement garder la même PDF en doublant l'énergie des particules). Finalement, les quarks de la mer sont inclus dans les PDF. Tous ces facteurs rendent les PDF longues et ardues à déterminer et augmentent leurs incertitudes. Deux exemples de PDF sont montrés à la figure 2.10.

Une fois la nature et l'énergie des partons collisionnés calculées, la première étape de simulation se termine de même que le rôle de **MadGraph** dans la simulation, qui envoie

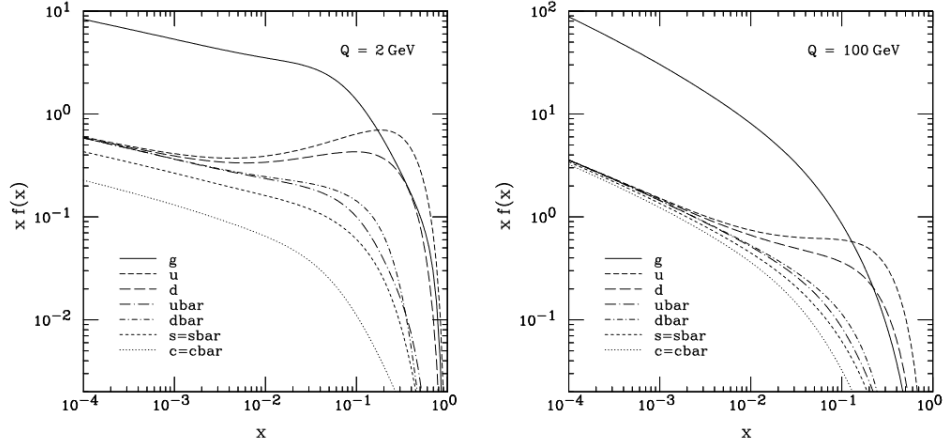


Fig. 2.10. Fonctions de distribution des partons pour une énergie du proton de 2 GeV (gauche) et de 100 GeV (droite). Celles-ci sont mesurées expérimentalement. Nous notons la différence importante des PDF selon l'énergie du proton, la domination du gluon et la présence d'antiparticules ainsi que de partons qui ne font pas partie de la composition du proton $p = uud$ [109].

ensuite l'information sur l'état actuel de la collision à `Pythia8`.

Ce logiciel simule les cascades de partons, qui se produisent par le principe de confinement des quarks, où les quarks s'éloignant les uns des autres produisent des nouvelles paires de quark-antiquark jusqu'à ce qu'il n'y ait plus suffisamment d'énergie pour produire de nouvelles paires et que l'hadronisation survienne. Celle-ci se manifeste quand les quarks s'agglomèrent en hadrons une fois l'énergie suffisamment basse. `Pythia8` calcule ces deux phénomènes, qui avait été illustrés à la figure 1.8 et qui sont à l'origine de la formation des jets. Cette étape est difficile, car elle implique la QCD et ses fameux calculs non perturbatifs qui doivent plutôt se faire par ajustements des modèles aux données.

Finalement, la simulation d'événements donne en résultat (*output*) leur lecture telle qu'obtenue via le détecteur ATLAS. Le logiciel `Delphes` s'assure de faire le pont entre les particules produites et leur trace dans le détecteur, c'est donc dire qu'il simule la réponse du détecteur face aux événements. Il traduira ainsi les particules qui le traversent en signaux détectés. De plus, il est possible pour les particules de « foncer » dans les noyaux atomiques du détecteur, un phénomène qui peut déclencher une gerbe électromagnétique ou hadronique. `Delphes` approximera aussi ces gerbes. C'est ici que se situe la différence majeure entre `Delphes` et d'autres simulateurs comme `GEANT4`, qui propose de simuler de façon beaucoup plus complète les gerbes produites. Par exemple, `GEANT4` simulera précisément les cascades électromagnétiques particule par particule, ce qui demande

énormément de ressources.

Delphes peut aussi reconstruire certaines particularités des événements qui sont plus complexes que des simples particules, comme par exemple les jets et l'énergie manquante. Lorsqu'il reconstruit des jets, il peut identifier quelles particules font partie dudit jet et calculer ses propriétés ainsi que ses variables de haut niveau (masse, p_T , ainsi qu'une panoplie d'autres variables telles que d_{12} , τ_{32}). Une liste de ces variables est donnée au tableau 1 de la source [72] où des références pour les expliquer sont données à même le tableau. Ces variables ont été créées par les scientifiques puisqu'elles sont déterminantes pour différencier ou catégoriser certains événements.

2.6. Conclusion

Avec tout l'appareillage au LHC maintenant expliqué, nous comprenons désormais comment le détecteur ATLAS prend ses mesures, quels événements il juge intéressant de conserver ainsi que la reconstruction de ses événements. Toutefois, nous ne savons pas si tout ce que produit le LHC est issu du modèle standard. Il nous faut donc plus d'outils pour mieux analyser les données produites par ATLAS dans l'espoir de trouver des signes de nouvelle physique encore jamais observée et potentiellement jamais expliquée. Le prochain chapitre se concentre sur l'apprentissage machine, un type d'intelligence artificielle que nous avons utilisé dans le cadre de ce projet de recherche. Celui-ci voit la quantité gigantesque d'événements d'ATLAS non pas comme un obstacle, mais plutôt comme un allié, puisqu'il excelle à évoluer dans un environnement avec énormément de données pour s'entraîner. De plus, il facilite la recherche qui n'est pas guidée par des idées théoriques, une philosophie au coeur de notre projet.

Chapitre 3

L'apprentissage automatique

Cette section peut présenter des éléments très nouveaux pour les lecteurs familiers avec la physique, mais moins avec l'apprentissage automatique. Une lecture de la référence [75], une introduction complète de l'apprentissage machine, peut aider les lecteurs les moins à l'aise avec le sujet.

L'intelligence artificielle (IA) est de plus en plus répandue et permet d'accomplir des tâches très variées et disjointes. Par exemple, elle peut reconnaître des visages et leurs émotions [132], recommander aux utilisateurs d'une plateforme web des articles ou des services qui risquent de leur plaire [134, 95], traduire des textes [106], traiter des courriels pour y détecter des pourriels ou des tentatives de fraude (*spam*) [107], battre les grands maîtres d'échecs ou du jeu de go [90, 81], déclasser les médecins quant à la détection de certains cancers [97], servir de robot conversationnel comme le tout nouveau **ChatGPT**, et j'en passe. Avec les développements récents de l'industrie, l'IA permet même de créer de l'art, de la musique, des images ou des visages.

La nomenclature de tout ce qui touche à l'intelligence artificielle peut être un brin complexe et non unanime. Ici, seul l'apprentissage automatique (AA mais aussi plus connu sous le nom de *machine learning*), un type particulier d'IA sera examiné. Celui-ci apprend en s'entraînant à l'aide d'exemples, puis peut ensuite être utilisé pour diverses applications sur de nouvelles données. Il constitue aujourd'hui presque l'entièreté du domaine de l'intelligence artificielle en vertu de ses performances inégalées. L'AA excelle dans plusieurs problèmes trop complexes pour l'humain ou présentant des données trop nombreuses pour être analysées efficacement. Comme les données réelles ou simulées du LHC sont présentes en très grand nombre, l'AA est un parfait candidat pour ce projet de mémoire.

3.1. Les réseaux de neurones

Depuis l'avènement des cartes graphiques puissantes (*graphics processing unit* ou *GPU*) dans les années 2010, les réseaux de neurones (*neural networks* ou *NN*) dominent outrageusement le domaine de l'apprentissage automatique. En effet, leur entraînement nécessite une très grande puissance de traitement,¹ mais leur performance et leur capacité à traiter de problèmes excessivement complexes en font l'outil de prédilection de tous.

Il existe deux grandes catégories de tâches effectuelles via les NN : la classification et la régression. La classification permet d'attribuer à chaque donnée une classe parmi plusieurs. Par exemple, reconnaître la race d'un chien grâce à une image de celui-ci est un problème de classification. Ainsi, la *prédiction* du modèle consiste en une classe ou une probabilité d'appartenir à chaque classe. Ensuite, la régression tente de *prédire* un nombre réel et non pas une classe. Par exemple, prédire le prix d'une action à la bourse ou du litre d'essence dans trois mois sont des exemples de problèmes de régression.

3.1.1. Le flux d'information du réseau de neurones

Les NN prennent en entrée \mathbf{x} , une donnée, performant toutes sortes d'opérations et retournent \mathbf{y} , qui sont tous les deux tenseurs de nombres réels.² Notons que la dimensionnalité de \mathbf{x} et de \mathbf{y} sont libres et peuvent différer. Ces entrées sont assez générales pour pouvoir appliquer presque n'importe quelle situation à des NN. Par exemple, pour classer des images encodées via les couleurs rouge, vert, bleu (RGB) d'animaux de $p_1 \times p_2$ pixels en m classes, l'entrée \mathbf{x} peut être composée des valeurs de chaque couleur dans l'encodage RGB. Dans ce cas, \mathbf{x} est un tenseur de dimension $3 \times p_1 \times p_2$ et \mathbf{y} peut être un vecteur qui retourne la probabilité d'une image d'appartenir à chaque classe : $\mathbf{y} \in \mathbb{R}^m$ ou \mathbf{y} est un tenseur de dimension m .

Au niveau structurel, les réseaux de neurones sont constitués d'une multitude de couches de neurones reliées entre elles de façon séquentielle où l'information est transmise d'une couche à l'autre. La première couche contient \mathbf{x} , l'information fournie au réseau, tandis que la dernière contient \mathbf{y} , la sortie du réseau. Les autres couches portent le nom de *couches cachées* et servent à transformer l'information successivement. Celles-ci peuvent toutes varier de dimensionnalité de l'une à l'autre, sans suivre de règle. Autrement dit, les tenseurs

¹ Comme leur nom l'indique, les cartes graphiques peuvent être utilisées pour traiter des images à être affichées à un écran. Par contre, elles sont de nos jours surtout utilisées pour effectuer des calculs machine, c'est-à-dire tout calcul mathématique qu'un ordinateur devrait pouvoir faire. C'est cette deuxième caractéristique qui est utilisée dans le cadre de l'AA puisqu'elles sont beaucoup plus rapides que leurs compétitrices, les unités centrales de calcul (*CPU*).

² Quelques rares NN peuvent fonctionner avec des nombres complexes, mais ils sont excessivement rares et ne seront pas traités dans ce mémoire. Presque tous les réseaux de neurones les excluent entièrement et ne les utilisent dans aucune partie de leur architecture.

qui représentent chaque couche cachée sont de taille arbitraire et ces tailles sont totalement indépendantes d'une couche à l'autre.

Pour transmettre l'information d'une couche à l'autre, les neurones communiquent entre eux à l'aide de transformations linéaires et de fonctions d'activation non-linéaires, dénotées σ . En réalité, chaque neurone est obtenu par combinaison linéaire des neurones de la couche précédente qui est ensuite fournie à une fonction d'activation non-linéaire prédéterminée. Comme la composition de fonctions linéaires demeure linéaire, les fonctions d'activation sont nécessaires pour permettre au réseau d'atteindre une non-linéarité afin d'être plus profond et complexe pour une même quantité de couches et de neurones par couche. Le réseau peut donc réduire sa taille sans diminuer sa capacité de prise de décision complexe. En effet, un réseau linéaire est très restreint puisqu'il ne peut que créer des frontières linéaires dans l'espace de phase des paramètres fournis en entrée pour délimiter les différentes sorties (*outputs*) qu'il peut offrir. L'introduction des fonctions σ permet au réseau de pouvoir moduler ces frontières de décision de façon beaucoup plus malléable et ainsi offrir des prises de décision plus complexes et raffinées. Notons que chaque couche peut avoir une fonction σ différente des autres. La figure 3.1 présente une vue schématique du flux d'information dans un NN, tandis que la figure 3.2 fournit quelques fonctions d'activation typiquement utilisées dans le domaine de l'apprentissage automatique.

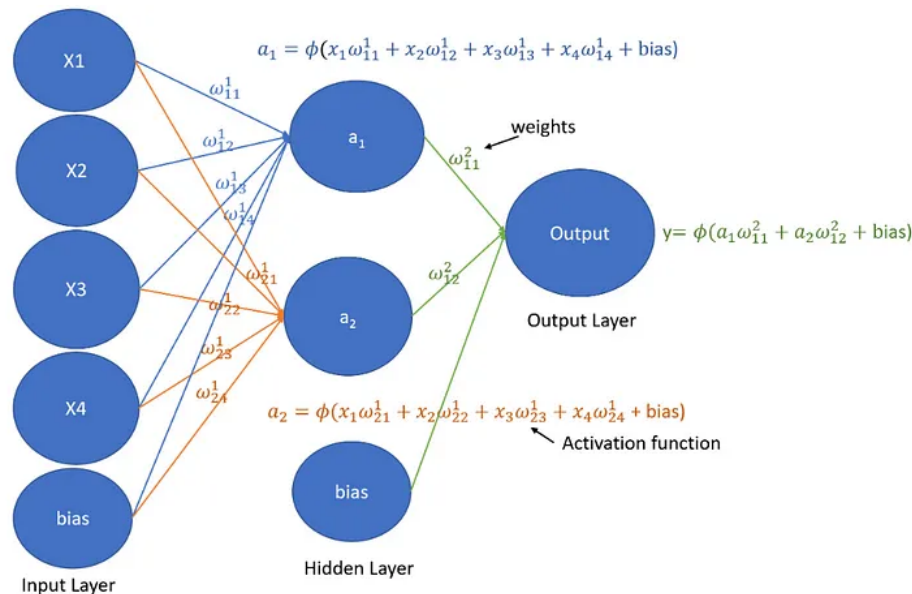
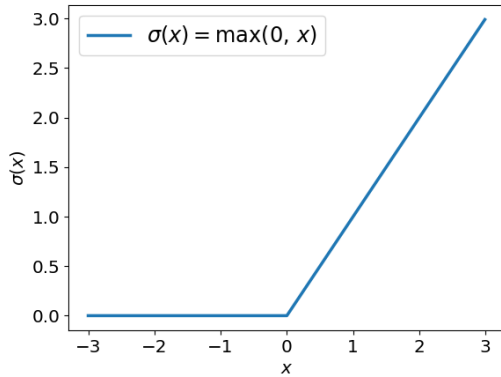
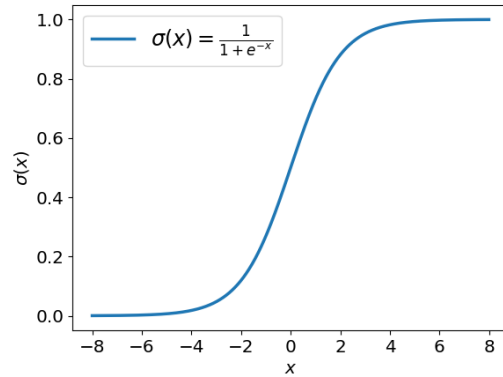


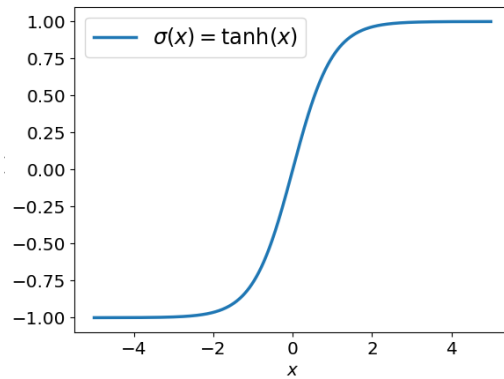
Fig. 3.1. Schéma du flux d'information dans un réseau de neurones. L'information provient de l'entrée \mathbf{x} et circule au travers des couches cachées (une seule dans cet exemple) et de la couche finale à l'aide de poids, de biais et de fonctions d'activation. Ici, les a_i représentent les neurones de la couche cachée (les $x_{i,1}$ de l'équation 3.1.2) [82].



(a) Fonction d'activation unitaire rectifiée linéaire ReLU (*rectified linear unit*).



(b) Fonction d'activation sigmoïde.



(c) Fonction d'activation tanh.

Fig. 3.2. Fonctions d'activation fréquemment utilisées dans les NN. N'importe quelle fonction pourrait être utilisée, mais celles-ci ont tendance à présenter des gradients doux et améliorant la stabilité de l'entraînement, qui sera présenté à la section 3.2.

Mathématiquement, la valeur de la k -ième couche \mathbf{x}_k est donnée par ³:

$$\mathbf{x}_k = \sigma_k \left(\mathbf{w}_k^\top \mathbf{x}_{k-1} + \mathbf{b}_k \right) \quad (3.1.1)$$

où σ_k représente une fonction non-linéaire quelconque, \mathbf{w}_k représente le tenseur de poids qui relie \mathbf{x}_{k-1} à \mathbf{x}_k avec la dimensionnalité variable nécessaire pour les relier, et \mathbf{b}_k représente le tenseur de biais de la couche k . Notons que les \mathbf{b}_k sont des constantes, que les tenseurs et les vecteurs sont en gras tandis que les scalaires ne le sont pas et finalement que la transposée sur la matrice de poids est une convention. Pour tenter de mieux comprendre intuitivement

³ Certains modèles, comme par exemple les réseaux de neurones convolutifs (CNN), experts en traitement d'images, n'utilisent pas exactement cette forme, mais une variante très rapprochée. Par exemple, les CNN utilisent des convolutions autour de chaque neurone, c'est donc dire qu'il existe autant de tenseurs de poids \mathbf{w}_k qu'il y a de neurones dans la k -ième couche. Toutefois, nous pouvons nous tenir aux formes données par 3.1.1 et 3.1.2 sans trop perdre de généralité.

l'équation 3.1.1, regardons ce qui se passe pour un neurone prédéterminé x_j de la k -ième couche :

$$x_{j,k} = \sigma_k \left(\sum_{i=0}^m w_{ij,k} x_{i,k-1} + b_{j,k} \right) \quad (3.1.2)$$

où σ_k représente encore une fonction d'activation non-linéaire quelconque, les w_{ij} représentent le poids reliant le neurone j de la couche k et le neurone i de la couche $k-1$, les $x_{i,k-1}$ représentent le neurone i de la couche $k-1$ et b_j représente le poids scalaire et constant associé au neurone j de la couche k .

Il est donc révélé que chaque neurone peut être relié plus ou moins fortement à chaque neurone de la couche précédente en fonction de la valeur du w_{ij} associé. Nous voyons aussi que ce que contiennent les couches cachées est souvent incompréhensible à l'humain. En effet, les NN peuvent être vus comme des énormes boîtes noires où plusieurs milliers ou millions d'opérations mathématiques très simples sont effectuées sans toutefois pouvoir être logiquement comprises par l'humain. Ceci est vrai en particulier lorsque nous savons que nous devons souvent appliquer l'équation 3.1.2 sur des dizaines ou des centaines de couches comportant chacune plusieurs milliers de neurones. Ces couches peuvent malgré tout parfois servir à comprendre la performance d'un réseau et à l'améliorer dans sa phase de développement, mais cette tâche est ardue et constitue une science imprécise. Ainsi, la plupart du temps, seule la dernière couche, qui représente la prédiction du réseau \mathbf{y} , est analysée.

La quantité de couches et de neurones par couche de même que l'architecture globale du réseau varient grandement d'un modèle à l'autre. Il existe plusieurs types pré-déterminés de modèles. Par exemple, les réseaux de neurones convolutifs (CNN, *convolutional neural networks*) [78] excellent dans tout travail qui se penche sur des images ou toute structure de données où l'information locale est très importante tandis que la grande famille des réseaux de neurones récurrents (RNN, *recurrent neural networks*) [114] est souvent la meilleure pour les tâches de reconnaissance vocale ou de traduction.

Ces structures sont capables d'encapsuler énormément d'information. Le théorème d'approximation universelle, l'un des plus merveilleux de l'apprentissage automatique, stipule grossièrement qu'un NN avec suffisamment de neurones est capable d'approximer arbitrairement près n'importe quelle fonction⁴ de \mathbb{R}^n , $n \in \mathbb{N}$. Or, comme chaque problème peut être reproduit, mathématisé ou approximé par une fonction $f(\mathbf{x}) \rightarrow \mathbf{y}$, les réseaux de neurones sont donc théoriquement capables de résoudre presque n'importe quoi. Nous

⁴ Qui est ensuite généralisable à n'importe quelle fonction de dimensionnalité supérieure puisque chaque dimension est indépendante.

reconnaissons le parallèle avec la structure même de l'apprentissage automatique, où \mathbf{x} et \mathbf{y} représentent l'entrée et la sortie du réseau, et où ces deux tenseurs sont de taille arbitraire (\mathbf{x} est de taille $n_1 \times n_2 \times \dots \times n_k$, $k \in \mathbb{N}$ et \mathbf{y} est de taille $m_1 \times m_2 \times \dots \times m_j$, $j \in \mathbb{N}$).

Par exemple, on peut alors découper l'espace des entrées \mathbf{x} en toutes sortes de fonctions limites où chaque région appartient à une classe pour un problème de classification, ou simplement calquer la prédiction \mathbf{y} sur la *vraie* fonction qu'elle devrait prendre en fonction de l'entrée \mathbf{x} . C'est pourquoi les NN sont si puissants et peuvent être appliqués à toutes sortes de tâches.

3.2. L'entraînement des réseaux de neurones

La section 3.1.1 explique comment l'information circule dans un NN, de la première couche à la dernière couche. C'est ce qu'on appelle la « passe en avant » (*forward pass*). L'entraînement d'un réseau se fait en appliquant successivement cette passe en avant et la « passe en arrière » (*backward pass*) qui sera décrite dans cette section.

3.2.1. La fonction de perte et la descente de gradient

Le but d'un réseau performant réside à donner en sortie le résultat attendu. Par exemple, un modèle conçu pour prévoir le cours d'une action à la bourse est performant s'il prédit avec justesse le cours de l'action. Cela revient à attiturer les « bonnes » valeurs aux neurones de la dernière couche, qui représentent la prédiction du modèle. Ceci est fait à l'aide d'une « fonction de perte » L (*loss function*). Celle-ci peut prendre plusieurs formes mais pour être performante, elle doit s'assurer d'être différentiable et d'être de plus en plus petite à mesure que \mathbf{y} prend la forme que l'on veut. Par exemple, un réseau qui veut s'entraîner à prédire le prix de l'immobilier pourrait s'entraîner sur certaines constructions dont le prix de vente z est connu au préalable. Alors, le réseau pourrait retourner le prix prédit par celui-ci $\mathbf{y} = y \in \mathbb{R}$ et la fonction de perte pourrait être par exemple $(y - z)^2$ ou $|y - z|$. Ceci constitue une bonne fonction de perte puisqu'elle diminue au fur et à mesure que le réseau prédit avec justesse le prix de vente des exemples fournis et puisqu'elle est dérivable.

De cette manière, lors de l'entraînement, après chaque passe en avant, le réseau peut juger de sa propre performance avec cette fonction de perte – plus elle est petite, plus le réseau a eu du succès – et s'ajuster en conséquence. En effet, la dernière couche n'est dépendante que de l'information entrante \mathbf{x} (ou en d'autres mots, l'exemple fourni au NN), une donnée fixe, et des poids et des biais de l'équation 3.1.1, que l'on appelle *paramètres internes*. Ils

représentent donc la seule chose qui peut être modifiée pour augmenter la performance du réseau, en modifiant sa prédiction \mathbf{y} pour en arriver à une perte moins grande à l'aide d'une *descente de gradient* appliquée sur L .⁵ Ceci explique aussi pourquoi L doit être différentiable. En plus, de par la transmission simple de l'information d'une couche à l'autre, le gradient est facile à calculer en utilisant la règle en chaîne. C'est d'ailleurs pourquoi cette étape est appelée la passe par derrière : contrairement à la passe avant qui transmet l'information de couche en couche de la première à la dernière du réseau, la passe en arrière calcule le gradient de L de par la dernière couche séquentiellement jusqu'à la première grâce à la règle de la dérivée en chaîne. Par exemple, dans un NN de k couches, pour calculer le gradient par rapport aux biais \mathbf{b}_k où k est la dernière couche, on retrouve,⁶ se rappelant que $\mathbf{y} = \mathbf{x}_k$:

$$\begin{aligned}
\frac{\partial L}{\partial \mathbf{b}_k} &= \left(\frac{\partial L}{\partial \mathbf{y}} \right) \left(\frac{\partial \mathbf{y}}{\partial \mathbf{b}_k} \right) \\
&= \left(\frac{\partial L}{\partial \mathbf{x}_k} \right) \left(\frac{\partial \mathbf{x}_k}{\partial \mathbf{b}_k} \right) \\
&= \left(\frac{\partial L}{\partial \mathbf{x}_k} \right) \frac{\partial}{\partial \mathbf{b}_k} \left[\sigma_k \left(\mathbf{w}_k^\top \mathbf{x}_{k-1} + \mathbf{b}_k \right) \right] \\
&= \left(\frac{\partial L}{\partial \mathbf{x}_k} \right) \sigma'_k \left(\mathbf{w}_k^\top \mathbf{x}_{k-1} + \mathbf{b}_k \right)
\end{aligned} \tag{3.2.1}$$

Puis, pour retrouver le gradient de la couche $k - 1$:

$$\begin{aligned}
\frac{\partial L}{\partial \mathbf{b}_{k-1}} &= \left(\frac{\partial L}{\partial \mathbf{x}_k} \right) \left(\frac{\partial \mathbf{x}_k}{\partial \mathbf{x}_{k-1}} \right) \left(\frac{\partial \mathbf{x}_{k-1}}{\partial \mathbf{b}_{k-1}} \right) \\
&= \left(\frac{\partial L}{\partial \mathbf{x}_k} \right) \left[\frac{\partial}{\partial \mathbf{x}_{k-1}} \left[\sigma_k \left(\mathbf{w}_k^\top \mathbf{x}_{k-1} + \mathbf{b}_k \right) \right] \right] \left[\frac{\partial}{\partial \mathbf{b}_{k-1}} \left[\sigma_{k-1} \left(\mathbf{w}_{k-1}^\top \mathbf{x}_{k-2} + \mathbf{b}_{k-1} \right) \right] \right] \\
&= \left(\frac{\partial L}{\partial \mathbf{x}_k} \right) \sigma'_k \left(\mathbf{w}_k^\top \mathbf{x}_{k-1} + \mathbf{b}_k \right) \left[\frac{\partial}{\partial \mathbf{x}_{k-1}} \left(\mathbf{w}_k^\top \mathbf{x}_{k-1} + \mathbf{b}_k \right) \right] \times \\
&\quad \sigma'_{k-1} \left(\mathbf{w}_{k-1}^\top \mathbf{x}_{k-2} + \mathbf{b}_{k-1} \right) \left[\frac{\partial}{\partial \mathbf{b}_{k-1}} \left(\mathbf{w}_{k-1}^\top \mathbf{x}_{k-2} + \mathbf{b}_{k-1} \right) \right] \\
&= \left(\frac{\partial L}{\partial \mathbf{x}_k} \right) \sigma'_k \left(\mathbf{w}_k^\top \mathbf{x}_{k-1} + \mathbf{b}_k \right) \mathbf{w}_k \sigma'_{k-1} \left(\mathbf{w}_{k-1}^\top \mathbf{x}_{k-2} + \mathbf{b}_{k-1} \right)
\end{aligned} \tag{3.2.2}$$

Et procédant ainsi pour redescendre jusqu'à \mathbf{b}_1 . Ici, les σ' représentent la dérivée des fonctions σ telle que nous le ferions par la règle de dérivée de composition de fonctions.

⁵ Il est aussi possible de maximiser la fonction de perte. Toutefois, par convention, c'est presque toujours la minimisation qui est recherchée. Ceci ne cause jamais problème car maximiser une fonction L revient à minimiser $-L$, alors il est toujours possible de trouver une L optimale cherchant à être minimisée.

⁶ Pour le calcul matriciel qui suit, une excellente référence est donnée en [104] : le livre de recettes des matrices (*matrix cookbook*).

Fonction	Équation	Dérivée
ReLU	$\sigma(x) = \begin{cases} 0 & \text{si } x < 0 \\ x & \text{si } x > 0 \end{cases}$	$\sigma'(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1 & \text{si } x > 0 \end{cases}$
Sigmoïde	$\sigma(x) = \frac{1}{1+e^{-x}}$	$\sigma'(x) = \sigma(x)(1 - \sigma(x))$
Tanh	$\sigma(x) = \tanh(x)$	$\sigma'(x) = 1 - \sigma^2(x)$

Tableau 3.1. Fonctions d'activation σ ainsi que leurs dérivées fréquemment utilisées dans les NN. Il s'agit des mêmes fonctions que celles présentées à la figure 3.2, mais nous avons désormais le bagage nécessaire pour discuter de leurs dérivées. Elles sont cruciales à la passe par en arrière, et doivent être substituées dans les équations 3.2.1, 3.2.2 et 3.2.3 élément par élément. Toutes les dérivées sont soit extrêmement simples comme pour la fonction ReLU ou proposent d'utiliser $\sigma(x)$ ou $\sigma^2(x)$ qui sont déjà connues, ce qui allège leur calcul qui doit être fait pour chaque neurone. De plus, les dérivées sont reconnues pour contribuer à la stabilité numérique de l'entraînement puisqu'elles n'ont jamais une très grande magnitude. Finalement, notons que la fonction ReLU n'est pas dérivable en $x = 0$, mais il est facile d'entrer manuellement une dérivée de 0 dans ce scénario presque impossible pour les fins de la descente de gradient.

Notons que tout ce qui est à l'intérieur des σ et des σ' sont des scalaires puisque nous faisons opérer σ individuellement sur chaque neurone (*element-wise operation*) comme il est montré à l'équation 3.1.2. Le tableau 3.1 présente les dérivées des fonctions d'activation les plus populaires. De plus, la forme exacte de $\partial L / \partial \mathbf{x}_k$ diffère d'un problème à l'autre mais est facilement calculable. Par exemple, dans la situation de régression des prix de l'immobilier où l'erreur quadratique moyenne était utilisée ($L = (y - z)^2 = (x_k - z)^2$, x_k est un scalaire), alors $\partial L / \partial x_k = 2(x_k - z)$.

Pour les matrices de poids \mathbf{w}_i , nous procédons de façon similaire :

$$\begin{aligned}
\frac{\partial L}{\partial \mathbf{w}_k} &= \left(\frac{\partial L}{\partial \mathbf{x}_k} \right) \left(\frac{\partial \mathbf{x}_k}{\partial \mathbf{w}_k} \right) \\
&= \left(\frac{\partial L}{\partial \mathbf{x}_k} \right) \frac{\partial}{\partial \mathbf{w}_k} \left[\sigma_k \left(\mathbf{w}_k^\top \mathbf{x}_{k-1} + \mathbf{b}_k \right) \right] \\
&= \left(\frac{\partial L}{\partial \mathbf{x}_k} \right) \sigma'_k \left(\mathbf{w}_k^\top \mathbf{x}_{k-1} + \mathbf{b}_k \right) \left[\frac{\partial}{\partial \mathbf{w}_k} \left(\mathbf{w}_k^\top \mathbf{x}_{k-1} + \mathbf{b}_k \right) \right] \\
&= \left(\frac{\partial L}{\partial \mathbf{x}_k} \right) \sigma'_k \left(\mathbf{w}_k^\top \mathbf{x}_{k-1} + \mathbf{b}_k \right) \mathbf{x}_{k-1}
\end{aligned} \tag{3.2.3}$$

Et répétant une fois de plus ce procédé jusqu'à \mathbf{w}_1 .

En conclusion, le processus d'entraînement est donc une répétition de ces mêmes étapes : prendre un exemple d'entraînement et lui faire la passe en avant du réseau, puis ajuster les poids et les biais des équations 3.1.1 et 3.1.2 grâce à un algorithme de descente de gradient au choix.

3.2.2. Le surentraînement et les données d'entraînement, de validation et de test

Il est plus ou moins clair que de ne donner qu'une seule fois chaque exemple possible à un réseau lors de l'entraînement n'est pas suffisant pour en soutirer une bonne performance. En effet, les paramètres internes sont initialisés au hasard et engendrent un espace dont la dimensionnalité est aussi grande que le nombre de paramètres. Ainsi, ces paramètres initiaux aléatoires n'ont aucune chance raisonnable d'être proches d'un minimum de cette fonction sur les données de validation, qui seront introduites sous peu. Alors, il faudra exposer le modèle aux mêmes exemples plusieurs fois pour espérer atteindre une performance raisonnable. Puisque le gradient servant d'ajustement aux paramètres internes sera grandement différent d'exemple en exemple, il est coutume de séparer l'entraînement en plusieurs époques (*epochs*) qui consistent chacune à montrer l'entièreté des données d'entraînement sans répétition dans un ordre aléatoire et variable d'une époque à l'autre. À première vue, il pourrait sembler optimal d'utiliser le plus grand nombre d'époques possible, ou quand la fonction de perte cesse de diminuer sur les données d'entraînement. En réalité, cette pratique est néfaste et porte le nom de *surentraînement* ou *surapprentissage*.

En effet, les données d'entraînement sont supposées représenter une certaine *catégorie*.⁷ Par contre, ces données ne forment qu'un sous-ensemble de celle-ci et n'en représentent donc qu'un *proxy* qu'on souhaite le plus représentatif possible de la « vraie » *catégorie*. Ainsi, le minimum de la fonction de perte sur ce sous-ensemble n'est pas le même que son minimum sur la « vraie » *catégorie* dans son ensemble. Une autre façon de visualiser ce concept est de se rendre compte que ce même minimum (ou l'entraînement du réseau dans son ensemble) est dépendant des données d'entraînement et que si l'entraînement était fait sur deux ensembles différents représentant la même *catégorie*, il convergerait vers deux états différents. Il faut en comprendre que le modèle entraîné est légèrement fait « sur mesure » pour les données d'entraînement et que si celles-ci sont montrées trop souvent au modèle, il tentera de converger trop près d'un minimum qui n'est pas celui de la *catégorie* de départ, mais du sous-ensemble auquel il a été exposé. On parle aussi d'un modèle qui *mémorise* ses données d'entraînement, et qui est capable d'ajuster ses paramètres internes pour que

⁷ Ici, le mot *catégorie* est défini comme étant très large. Il peut s'agir d'une classe dans un problème de classification, mais aussi d'un sous-ensemble des données qui présentent une structure similaire. Ce mot sera utilisé pour les prochains paragraphes dans cette optique en *italique*.

chaque exemple ou presque possède une petite valeur de L , mais sans pour autant en comprendre la structure de sa *catégorie* associée. Ceci mène à une perte de la *généralisation* du modèle, c'est-à-dire de sa performance sur des données jamais vues auparavant. Un parallèle est possible avec l'humain : dans une tâche, si une personne n'apprend qu'à résoudre des problèmes d'une certaine tâche par coeur, les mathématiques par exemple, et qu'elle n'apprend jamais la structure de ces problèmes ou de leur résolution, elle ne pourra pas transférer ses acquis à de nouveaux problèmes semblables mais jamais vus auparavant. Une visualisation du surapprentissage est présentée à la figure 3.3.

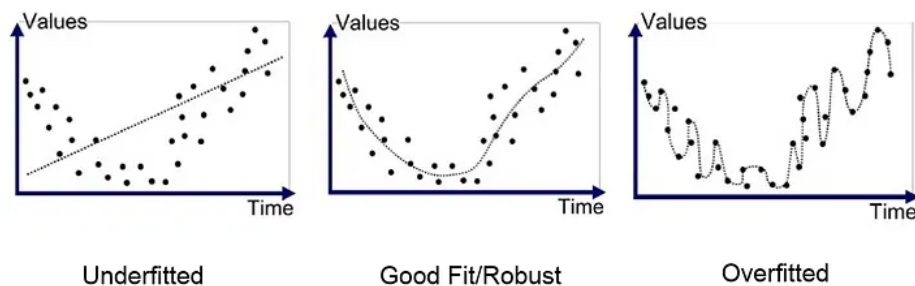


Fig. 3.3. Schéma global du sous-apprentissage, du surapprentissage et du « bon » apprentissage. Les points représentent les données d'entraînement. Si le modèle n'est pas assez entraîné ou n'a pas assez de capacité (n'est pas assez complexe) pour la tâche, il ne parviendra pas à tirer suffisamment d'information sur les données comme sur la figure de gauche. En revanche, s'il est trop complexe pour la tâche ou entraîné sur un trop grand nombre d'époques, il tentera de trop se mouler aux données d'entraînement ou de les mémoriser, ce qui est néfaste (figure de droite). En effet, il est beaucoup plus probable qu'un modèle bien entraîné comme celui du centre propose des meilleures performances sur des nouvelles données que le modèle surentraîné qui perd sa capacité de généralisation. La figure est issue de [79].

La réponse à ce problème vient des *données de validation*. Celles-ci subissent généralement la passe en avant du modèle à la fin de chaque époque, c'est-à-dire que chaque exemple se voit donner une perte via L , mais *elles ne sont pas utilisées pour actualiser les paramètres internes via la passe en arrière*. Ainsi, elles se voient confier le rôle de pseudo-nouvelles données jamais vues par le modèle et servent d'exemples de données que pourrait voir le réseau lors de son application. Encore une fois, nous espérons que ces données soient représentatives de la vraie *catégorie* à laquelle appartiennent les données. Ainsi, à la fin de chaque époque, il est possible d'estimer la généralisation du réseau avec celles-ci. La progression typique de la performance d'un modèle en fonction du nombre d'époques sur les deux ensembles de données est présentée à la figure 3.4.

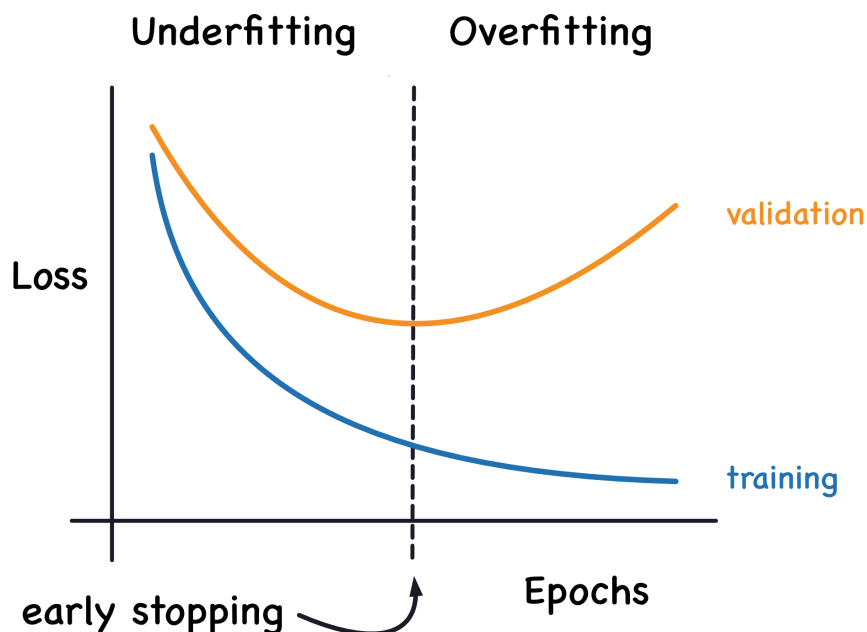


Fig. 3.4. Progression typique de la fonction de perte évaluée sur les données d’entraînement et de validation en fonction du nombre d’époques. Bien que la performance sur les données d’entraînement (ligne bleue) ne cesse de s’améliorer, ceci n’aide pas le réseau dans sa quête de généralisation puisque les données de validation commencent à augmenter leur perte. Il est possible de faire un parallèle avec la figure 3.3 : quand le réseau commence à atteindre son état démontré à la droite de cette figure, il augmente sa perte sur les exemples qui ne sont pas montrés, qui devraient plutôt suivre le modèle de la figure du centre [80].

La solution la plus globalement utilisée pour empêcher le surentraînement est donc d’arrêter l’entraînement lorsque L atteint son minimum sur les données de validation, c’est-à-dire que L commence à augmenter pour quelques époques consécutives pour ces données. Ensuite, le réseau retourne à son état qui a le mieux performé sur ces données, qui avait préalablement été sauvegardé. Ceci permet d’entraîner suffisamment longtemps pour que le modèle puisse être performant et comprenne la structure de ce qui lui est appris, mais pas trop au point où il commence à mémoriser ses données d’entraînement et en perde sa généralisation.

Une fois un modèle entraîné, il est légitime de vouloir tester sa performance dans le cadre d’une *vraie* application avant son utilisation dans le monde réel, c’est-à-dire d’effectivement utiliser le modèle avec des paramètres internes fixes sur des exemples. Pour ce faire, tout lecteur avisé comprendrait qu’il est fortement déconseillé d’y utiliser les données d’entraînement. Par contre, la raison pour laquelle il est non recommandé d’utiliser les données

de validation est moins évidente. Même s'il est vrai qu'elles n'ont pas servi à changer les paramètres internes du réseau, elles ont tout de même eu un impact sur l'entraînement du réseau puisqu'elles ont décidé à elles seules de l'état « optimal » du réseau. Ainsi, si nous voyons encore les données de validation comme un *proxy* de la *catégorie* qu'elles devraient représenter, le minimum de L sur celles-ci ne représente pas le minimum de L sur la « vraie » *catégorie*. Il convient donc d'utiliser trois ensembles de données lors de l'entraînement :

- Les données d'entraînement, qui servent à effectivement entraîner en ajustant les paramètres internes grâce à la combinaison de la passe en avant et de la passe en arrière ;
- Les données de validation, qui servent à surveiller les tendances de surapprentissage du réseau et décident du moment où arrêter l'entraînement, en servant d'exemples de la *catégorie* entraînée que pourrait voir le réseau et sur lequel le réseau est appliqué, mais pas entraîné ;
- Les données de test, qui servent à tester la performance réelle du modèle une fois l'entraînement terminé et qui ne doivent pas être vues du tout lors de celui-ci.

Souvent, par abus de langage, ces trois catégories sont appelées *données d'entraînement* puisqu'elles sont utilisées avant la mise en application du modèle dans le monde réel, mais aussi parce qu'elles représentent toutes des exemples de données (*proxy*) valides de la classe à évaluer. Notons que ces données sont toutes de même nature et ne sont divisées ainsi que pour les raisons expliquées plus haut, mais que cette division peut être faite n'importe comment, car chaque exemple est utilisable dans chaque catégorie. Par exemple, si nous voulions entraîner un réseau à reconnaître des images de chats et de chiens en utilisant comme exemples des photos de ces animaux, chacune de ces images pourrait servir soit de donnée d'entraînement, de validation ou de test.

Pour donner une idée de l'ordre de grandeur de leur importance relative au sein d'un entraînement, il est coutume d'utiliser environ 70% des données pour l'entraînement, 15% pour la validation et 15% pour le test, mais ces nombres peuvent énormément varier en raison de plusieurs facteurs, notamment la taille du jeu de données initial.

3.2.3. L'entraînement supervisé et l'entraînement non supervisé

Tout ce qui a été décrit jusqu'ici traite de l'entraînement supervisé, c'est-à-dire que chaque exemple comporte son étiquette (*label*) durant l'entraînement qui est utilisée dans la fonction de perte. C'est la façon la plus naturelle de fonctionner dans plusieurs problèmes d'apprentissage automatique. Par contre, elle suppose aussi de connaître d'avance l'étiquette de tous les exemples préalablement à l'entraînement, ce qui n'est pas toujours possible. C'est cette impossibilité qui est la cause la plus fréquente du recours à l'entraînement non

supervisé, qui, vous l'aurez deviné, ne comporte pas d'étiquette lors de l'entraînement.

Cette méthode d'entraînement force le réseau à comprendre la structure des exemples donnés. En effet, puisqu'il n'a pas accès aux étiquettes, il n'a d'autre choix que d'extraire les caractéristiques déterminantes présentes dans les données d'entraînement. En s'entraînant sur suffisamment d'exemples, un modèle non supervisé est capable de regrouper les différents exemples en différentes classes ou catégories non étiquetées. Deux cas principaux d'usage de cette méthode sont disponibles :

- Dans le cas de données analysables mais difficiles, coûteuses ou longues à identifier, l'apprentissage non supervisé peut tout de même répartir les exemples en plusieurs catégories, sans toutefois connaître la nature de celles-ci ;
- D'autre part, et c'est cet usage qui sera utile pour nous dans ce mémoire, l'entraînement non supervisé peut reconnaître des subtilités indétectables dans des données qui sont non étiquetables (*unlabelable*) pour l'humain. Cette incapacité est causée par des données qui sont trop similaires, trop compliquées ou carrément incompréhensibles pour l'humain. Un réseau peut toutefois comprendre des relations entre les données invisibles et inaccessibles pour nous. Ainsi, l'apprentissage non supervisé peut créer des sous-classes ou des sous-catégories au sein des exemples d'entraînement tout en reconnaissant les caractéristiques de celles-ci. L'une des applications de ce type de réseau est d'ensuite l'exposer à des nouveaux exemples et de tenter de voir dans quelle sous-catégorie ils vont tomber.

Dans le cadre de ce mémoire, l'apprentissage non supervisé servira à tenter de comprendre la structure et les caractéristiques des jets à grand rayon au sein de données simulées du LHC. Puis, une fois exposé à des réelles données, le réseau tentera de les caractériser comme semblables ou non à des données issues du modèle standard actuel. Les jets jugés comme les plus différents par ce réseau feront ensuite l'objet d'une analyse détaillée au chapitre 4.

3.3. L'auto-encodeur

Un auto-encodeur est un type de réseau non supervisé qui tente de reproduire l'entrée \mathbf{x} en sortie. Il est composé de deux réseaux de neurones : un encodeur et un décodeur. L'encodeur envoie l'information d'entrée \mathbf{x} vers une couche latente, le plus souvent de dimensionnalité significativement inférieure aux données,⁸ puis le décodeur prend l'information de cette couche latente et tente de reconstruire \mathbf{x} . Typiquement, les structures de l'encodeur

⁸ Ceci permet de s'assurer que le réseau n'apprenne pas une fonction identité. En effet, puisqu'il doit significativement réduire puis augmenter la dimensionnalité de l'information qu'il transmet de couche en couche, il est alors forcé d'apprendre quelque chose de plus fondamental au niveau des données et d'en comprendre leur structure.

et du décodeur sont miroitées, c'est-à-dire qu'elles proposent le même nombre de couches et le même nombre de neurones en sens inverse comme vu à la figure 3.5.

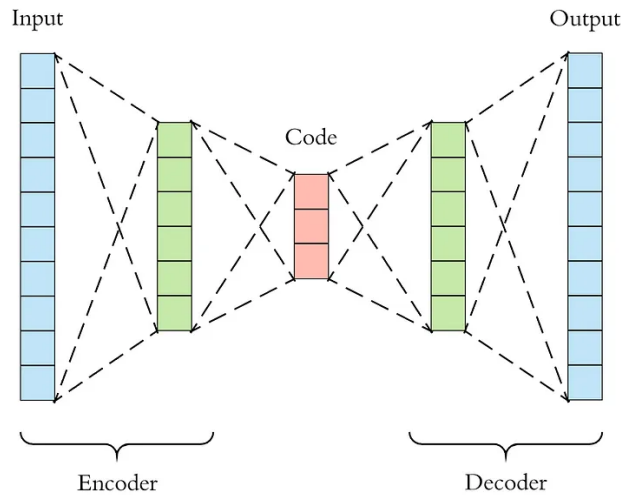


Fig. 3.5. Schéma de l'architecture d'un auto-encodeur. L'encodeur envoie l'entrée \mathbf{x} sur la couche latente de dimensionnalité inférieure (en rouge sur l'image), puis le décodeur prend l'information de cette couche latente et la reconstruit au meilleur de sa capacité. Les couches de l'encodeur et du décodeur sont miroitées [66].

Pour s'assurer de la volonté du modèle à reproduire l'entrée \mathbf{x} en sortie, la fonction de perte utilisée compare habituellement la sortie de l'auto-encodeur à son entrée avec des fonctions comme l'erreur quadratique moyenne (MSE) tel que vu à l'équation 3.3.1 :

$$L_{\text{auto-encodeur}} = \text{MSE}(\mathbf{y}, \mathbf{x}) = \frac{1}{N} \sum_{i=1}^N (y_i - x_i)^2 \quad (3.3.1)$$

où les y_i et les x_i représentent les N entrées de \mathbf{x} et de \mathbf{y} . Notons que dans l'application d'un tel problème, la normalisation via le $1/N$ pourrait être omise puisqu'on ne cherche qu'à minimiser cette fonction de perte, donc toute constante est superflue. Notons que plusieurs autres fonctions continues et différentiables qui tentent de rapprocher l'entrée et la sortie sont adaptées. Il suffit ensuite d'appliquer les mêmes algorithmes de descente du gradient habituels pour entraîner un modèle à recréer l'entrée en sortie et créer un auto-encodeur. Les auto-encodeurs sont d'excellents candidats à la recherche d'anomalies : en effet, toute donnée structurellement différente des données d'entraînement devrait être moins bien reconstruite et proposer une fonction de perte plus grande. En utilisant la magnitude de cette fonction de perte, il est possible d'isoler les cas les plus anormaux observés par l'auto-encodeur.

3.3.1. L'auto-encodeur variationnel (VAE)

Un auto-encodeur variationnel (VAE) possède une structure identique à celle d'un auto-encodeur, aux différences près que la couche latente est hautement *régularisée* et que

cette couche sera ensuite échantillonnée (*sampled*) avant de passer à la couche suivante. En effet, nous imposons une certaine structure à l'espace latent dans le but de pouvoir générer des nouvelles données. C'est ce qu'on appelle la régularisation de la couche latente. Tandis que l'espace latent associé à un auto-encodeur traditionnel peut être excessivement épars dans l'espace qu'il engendre, celui du VAE est concentré dans une région dense. Alors, il est possible de choisir au hasard un point de ces régions de l'espace latent et de le faire passer au-travers le décodeur, créant ainsi une sortie \mathbf{y} qui devrait avoir un sens puisque la couche latente utilisée était topologiquement voisine d'exemples réels. Ainsi se créent des nouvelles données qui devraient avoir une sous-structure semblable à celles des données fournies en entraînement. Ceci est plus difficile dans le cas d'un auto-encodeur traditionnel puisqu'il n'y a que peu de régions suffisamment densément peuplées pour que le modèle puisse générer quelque chose d'intelligible à partir d'un point qui n'appartient pas aux données. Cette distinction est illustrée à la figure 3.6.

Examinons maintenant l'échantillonnage appliqué sur la couche latente. La technique la plus utilisée est de considérer que cette couche est composée de $2n$ neurones dont les valeurs représentent $2n$ paramètres associés à n gaussiennes,⁹ soient n paramètres μ_i et n paramètres σ_i . Ceux-ci représentent l'espérance et l'écart-type des gaussiennes, respectivement. La définition de la fonction de densité d'une gaussienne (aussi appelée loi normale) est donnée à l'équation 3.3.2 :

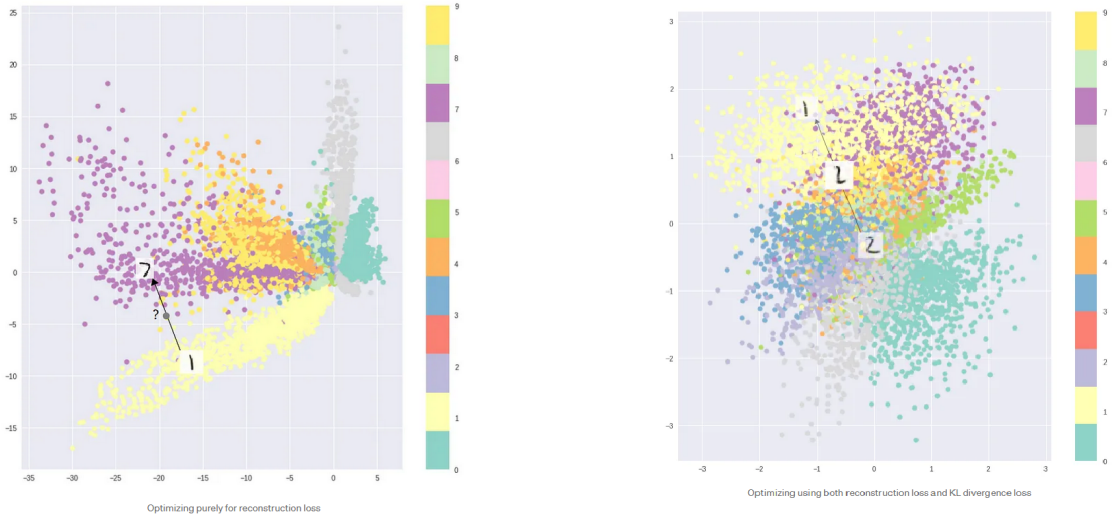
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} . \quad (3.3.2)$$

Nous tirons ensuite de ces fonctions de densité n nombres au hasard pour créer le nouveau vecteur de la couche latente, dénoté \mathbf{z} . C'est ce nouveau vecteur \mathbf{z} de dimension n qui est utilisé pour poursuivre la passe avant du réseau. La couche latente possède donc en réalité deux vecteurs : celui des $2n$ paramètres des gaussiennes, puis celui des n nombres échantillonnés individuellement pour chaque gaussienne. Comme ces paramètres proviennent d'une couche latente, ils dépendent de l'entrée \mathbf{x} et ils sont entraînaibles. L'architecture du réseau est donnée à la figure 3.7.

Notons que l'instauration de cet élément de hasard est non différentiable, ce qui semble empêcher la descente de gradient pour mettre à jour les paramètres internes du modèle. Toutefois, ceci est contourné en réécrivant l'étape d'échantillonnage. En effet, la construction du vecteur \mathbf{z} peut être réécrite de la sorte :

$$z_i = \mu_i + \sigma_i \epsilon_i , \quad (3.3.3a)$$

⁹ Il est toutefois possible d'utiliser plusieurs représentations, par exemple tenter de forcer les valeurs de la couche latente à résider sur une hypersphère [133].



(a) Représentation en deux dimensions de l'espace latent d'un auto-encodeur traditionnel.

(b) Représentation en deux dimensions de l'espace latent d'un auto-encodeur variationnel.

Fig. 3.6. Comparaison des espaces latents représentés en deux dimensions d'un auto-encodeur traditionnel et d'un auto-encodeur variationnel. Dans cet exemple, les modèles s'entraînent à reconstruire des images des chiffres 0 à 9 manuscrits. Comme on peut le voir, l'auto-encodeur (gauche) possède un espace latent épars comportant beaucoup de trous. Il est donc difficile de générer des nouvelles images de ces chiffres en inventant un vecteur de la couche latente, puis en le faisant passer par le décodeur. Le VAE (droite) possède quant à lui un espace latent beaucoup plus densément peuplé, et il est donc possible de générer facilement des nouvelles images de chiffres en créant un nouveau vecteur \mathbf{z} issu de cette région et en le faisant passer par le décodeur. Notons la différence entre les échelles des deux images. Notons aussi que ce ne sont qu'une représentation réduite à deux dimensions de l'espace latent qui possède une plus grande dimensionnalité, ce qui rend la comparaison visuelle possible [112].

$$\epsilon_i \sim \mathcal{N}(0, 1) . \tag{3.3.3b}$$

Or, l'élément de hasard est maintenant contenu dans les ϵ_i , qui sont désormais considérés comme étant des constantes dans le calcul de la descente de gradient. Il est donc possible de calculer un gradient du vecteur \mathbf{z} aux couches latentes puis jusqu'à l'encodeur, mais impossible de le faire pour le vecteur $\boldsymbol{\epsilon}$. Ceci ne pose pas problème, car les ϵ_i sont toujours tirés au hasard à chaque exemple fourni au VAE et ne sont pas entraînaibles. Cette reparamétrisation est elle aussi démontrée à la figure 3.7.

Il existe maintenant diverses façons de contraindre l'espace latent, mais dans le cas où nous interprétons la couche latente comme étant les paramètres des lois normales, la plus logique consiste à imposer au VAE un espace latent constitué de gaussiennes les plus standard possibles. Cela revient à forcer tous les μ_i à être le plus proche de 0 et tous les σ_i à être

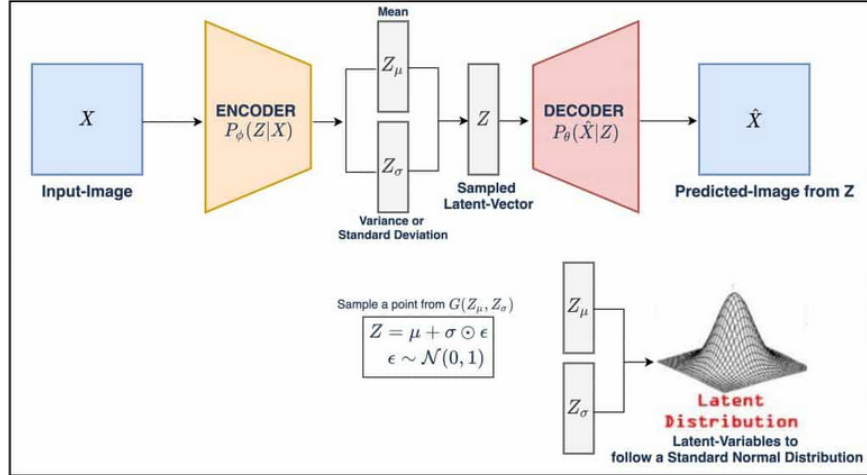


Fig. 3.7. Architecture standard d'un VAE. La structure d'un VAE est très semblable à celle d'un auto-encodeur traditionnel, sauf que la couche latente est à la fois régularisée et échantillonnée pour passer au vecteur \mathbf{z} qui sera ensuite soumis au décodeur. Nous apercevons aussi la réécriture de l'échantillonnage qui permettra de différencier la fonction de perte jusqu'au décodeur pour la descente de gradient de l'algorithme [113].

le plus proche de 1.¹⁰ Il est toutefois moins clair de quelle façon nous pouvons imposer cette contrainte au modèle. Ceci se fera lors de l'entraînement, où le réseau sera pénalisé dans les occurrences où le VAE propose des espaces latents avec des lois normales très peu standard. Pour ce faire, nous allons ajuster la fonction de perte du VAE avec l'addition de la *divergence de Kullback-Leibler* (KLD) [125]. Cette divergence est une mesure de la « distance statistique » entre deux distributions et bien qu'elle ne soit pas formellement une distance, nous pouvons l'interpréter étant une sorte de distance topologique généralisée. Même si elle peut être utilisée sur deux ensembles de valeurs discrètes, nous utiliserons sa définition $D_{KL}(P || Q)$ qui s'applique sur deux distributions continues P et Q :

$$D_{KL}(P || Q) = \int_{-\infty}^{\infty} p(x) \log \left(\frac{p(x)}{q(x)} \right) dx , \quad (3.3.4)$$

où $p(x)$ et $q(x)$ sont les fonctions de densité de probabilité de P et de Q . Cette quantité est nulle si $p(x) = q(x) \forall x$ et positive sinon. Notons que de façon générale, cette divergence n'est pas symétrique dans le sens où

$$D_{KL}(P || Q) \neq D_{KL}(Q || P) . \quad (3.3.5)$$

Nous utiliserons cette divergence pour évaluer l'écart entre les gaussiennes observées de la couche latente et les gaussiennes désirées, qui sont des gaussiennes $\mathcal{N}(0, 1)$. Plus spécifiquement, nous utilisons le résultat intéressant de la divergence de Kullback-Leibler

¹⁰ Notons qu'il devient alors très trivial de générer des nouvelles données : il suffit d'échantillonner n nombres issus de lois normales $\mathcal{N}(0, 1)$ et de les donner en entrée du décodeur.

entre deux gaussiennes. En effet, en substituant P et Q de l'équation 3.3.4 par des lois normales définies par $P = \mathcal{N}(\mu_1, \sigma_1)$ et $Q = \mathcal{N}(\mu_2, \sigma_2)$ et $p(x)$ et $q(x)$ par les fonctions de densité d'une loi normale de l'équation 3.3.2, nous retrouvons que la KLD ne dépend que des espérances et des écarts-types des gaussiennes [74] :

$$D_{KL}(\mathcal{N}(\mu_1, \sigma_1) \parallel \mathcal{N}(\mu_2, \sigma_2)) = \log\left(\frac{\sigma_2}{\sigma_1}\right) + \frac{\sigma_1^2 + (\mu_1 - \mu_2)^2}{2\sigma_2^2} - \frac{1}{2} . \quad (3.3.6)$$

Cela se simplifie davantage si $Q = \mathcal{N}(0, 1)$, soit si $(\mu_2, \sigma_2) = (0, 1)$:

$$D_{KL}(\mathcal{N}(\mu_1, \sigma_1) \parallel \mathcal{N}(0, 1)) = -\frac{1}{2} [\log(\sigma_1^2) - \sigma_1^2 - \mu_1^2 + 1] . \quad (3.3.7)$$

Pour régulariser l'espace latent d'un VAE, il suffit d'ajouter à la fonction de perte d'un AE la moyenne des divergences de l'équation 3.3.7 calculée sur les n gaussiennes de la couche latente multipliée par un hyperparamètre β . Ainsi, puisque nous voulons à la fois bien reconstruire les données de sortie \mathbf{y} vers le vecteur d'origine \mathbf{x} de dimension m , une fonction de perte typique d'un VAE ressemble à :

$$L_{\beta\text{-VAE}} = \text{MSE}(\mathbf{y}, \mathbf{x}) + \frac{\beta}{n} \sum_{i=1}^n D_{KL}(\mathcal{N}(\mu_i, \sigma_i) \parallel \mathcal{N}(0, 1)) \quad (3.3.8)$$

$$= \frac{1}{m} \sum_{i=1}^m (y_i - x_i)^2 - \frac{\beta}{2n} \sum_{i=1}^n [\log(\sigma_i^2) - \sigma_i^2 - \mu_i^2 + 1] , \quad (3.3.9)$$

où les x_i et y_i représentent les composantes des vecteurs \mathbf{x} et \mathbf{y} . Le deuxième terme tend à amener les gaussiennes latentes vers des gaussiennes $\mathcal{N}(0, 1)$, ce qui explique la concentration des espaces latents dans une région dense tel que vu à la figure 3.6(b). Finalement, l'ajout de β comme hyperparamètre permet de choisir l'importance de la régularisation de cet espace par rapport à la reconstruction des données. En effet, en augmentant β , il devient plus robuste de générer des nouvelles données à partir d'un espace latent mieux régularisé et plus dense, mais il est possible que la reconstruction des objets soit moins exacte, et vice-versa.

En conclusion, un VAE est un type d'auto-encodeur qui possède une couche latente régularisée qui paramétrise des lois normales les plus centrées réduites possible. Ceci permet au VAE de générer des nouvelles données en créant artificiellement des nouveaux vecteurs \mathbf{z} de la couche latente en échantillonnant n nombres de lois normales centrées réduites qui seront envoyés dans le décodeur. Cette régularisation se fait à partir de l'ajout d'un terme qui calcule la divergence de Kullback-Leibler entre les lois normales obtenues et les lois normales standard. Moins l'espace latent est régularisé, c'est-à-dire plus ses gaussiennes sont éloignées de lois normales centrées réduites, alors plus ce terme sera important et plus le modèle tentera de régulariser l'espace latent.

Tout comme les auto-encodeurs, les auto-encodeurs variationnels sont des modèles de choix pour la détection d'anomalie puisque leurs données anormales devraient être moins bien reconstruites et proposer une fonction de perte plus grande. En plus, en raison de la capacité générative du VAE, il devrait mieux comprendre la sous-structure des données que l'auto-encodeur, qui ne fait que trouver une façon de déconstruire puis reconstruire les données. Pour cette raison, le VAE pourrait être un meilleur détecteur d'anomalie que l'auto-encodeur traditionnel.

Chapitre 4

La détection d'anomalies au Grand Collisionneur de Hadrons

Maintenant que nous comprenons les motivations physiques qui dictent notre recherche, les appareillages expérimentaux qui s'offrent à nous au CERN et les méthodes d'apprentissage automatique à notre portée, nous sommes prêts à nous lancer dans le projet de recherche de notre groupe : la détection d'anomalies au Grand Collisionneur de Hadrons.

Pour rappel, nous nous concentrons sur la recherche de nouvelle physique sous forme de jets boostés au sein de l'expérience ATLAS. Comme nous sommes toujours en phase de recherche et développement, nous utilisons des simulations de données plutôt que les données réelles enregistrées par ATLAS. Nous n'étudions que le jet à plus grand p_T produit dans chaque événement. Nous entraînerons des VAE sur des jets QCD énergétiques à grand rayon ($R = 1$) issus du modèle standard (ainsi que sur des données hors distribution,¹ ou *outliers*, que nous verrons plus tard), puis nous exposerons ces modèles à divers signaux à la fois contenus dans le MS et hors du MS. L'entraînement permet aux VAE d'apprendre la structure des jets issus du MS tandis que l'exposition à de nouvelles données permet d'analyser leur capacité à découvrir de la nouvelle physique. En effet, la reconstruction et l'espace latent des jets de signal devraient être de moindre qualité par rapport à ceux issus du bruit de fond puisque le modèle n'a jamais vu de jets semblables au signal lors de l'entraînement. Ainsi, nous pouvons identifier des événements « anormaux » via la qualité de leur reconstruction et de leur espace latent, qui peut être vue comme un « score d'anomalie ». Dans ce projet, seule la qualité de l'espace latent sert de score d'anomalie. En effet, bien que nous entraînions le modèle à simultanément posséder un espace latent de qualité et à bien reconstruire les jets, cette reconstruction ne sera pas utilisée pour déterminer du niveau d'anomalie des jets lors de l'application du modèle. Nous étudions

¹ D'ailleurs, l'utilisation du terme « données hors distribution » est réservée uniquement à ces *outliers* qui sont au coeur du OE-VAE. Aucun autre usage n'y sera rattaché.

ensuite ces événements anormaux en utilisant un certain seuil d’anomalie sous lequel tout événement trop « normal » est rejeté de l’étude. Puis, nous espérons que les événements survivant à ce seuil sont majoritairement issus du signal et que seuls quelques jets du bruit de fond y survivront. Nous analyserons ensuite ces données anormales qui devraient avoir une plus grande concentration de jets issus de la nouvelle physique que l’ensemble des jets produits et analysés au détecteur ATLAS.

La technique au coeur de notre projet est l’exposition aux données hors distribution, qui tente de décorrélérer le score d’anomalie à la masse des jets. Elle sera décrite à la section 4.2.2 à la suite du survol de l’architecture du VAE sans exposition aux données hors distribution de la section 4.2.1. Cette méthode utilise un autre type de données lors de l’entraînement qui ne se veut ni du signal, ni du bruit de fond. Son but est de montrer au réseau de neurones que de la physique qui n’est pas associée au bruit de fond peut être observée à n’importe quelle masse et non seulement à haute masse.

Finalement, bien que l’objectif ultime soit de trouver de la nouvelle physique dans les données du LHC, nous attendons d’avoir des modèles suffisamment polis avant d’effectuer cette *vraie* recherche. Nous entraînons plutôt des jets sur des données simulées d’événements banals présents en très grand nombre au LHC : les événements à deux jets à haute énergie. Ces jets ne sont constitués que de deux gluons ou quarks légers ayant une grande impulsion qui produisent des jets reconstruits avec un rayon $R = 1$. Ensuite, nous exposons le VAE à des différents signaux de jets boostés, dont la particule mère est parfois associée au modèle standard (quark top) ou parfois issue de modèles théoriques hors du MS (Higgs du modèle 2HDM et boson Z' , voir section 4.3.3). Ces différents signaux tentent de simuler un environnement où le modèle est exposé à des données jamais vues auparavant, une imitation de la « nouvelle physique » permettant de mesurer la performance du modèle. Nous utilisons plusieurs outils d’analyse de performance sur les VAE entraînés pour parfaire notre connaissance de ceux-ci et optimiser leur performance.

4.1. Méthodes d’analyse

Une fois les jets les plus anormaux sélectionnés, il faut maintenant analyser cet ensemble de données pour y trouver des résultats intéressants. Or, une nouvelle particule massive pourrait créer une résonance et augmenter significativement le nombre de jets retrouvés dans un intervalle autour de sa masse. Nous nous intéressons donc à la « chasse aux bosses » (*bump hunting*). L’idée générale de cette technique est de retrouver des « bosses » (ou « pics ») dans l’histogramme du nombre de jets observés selon la masse. La présence de ces bosses impliquerait alors la formation de jets boostés via la désintégration d’une

particule massive encore inconnue, où le centre du pic correspondrait approximativement à la masse de cette nouvelle particule.

Nous comprenons aussi pourquoi il est important d’augmenter la concentration des jets de la nouvelle physique. En effet, dans les vraies données d’ATLAS, si ces jets sont présents dans les données, il serait toutefois très ardu de les retrouver s’ils sont noyés dans un océan de jets du MS. On se doute alors que leur pic est trop petit pour être observé. En effet, les fluctuations statistiques d’une fenêtre de masse à l’autre peuvent surpasser l’importance de ces bosses, et un tout petit excès de données dans une fenêtre de masse ne peut être analysé en profondeur. En augmentant drastiquement la concentration de nouveaux jets, cette bosse prend une plus grande importance relative et commence à être visible ou à tout le moins, statistiquement intéressante. Ce phénomène est illustré à la figure 4.1.

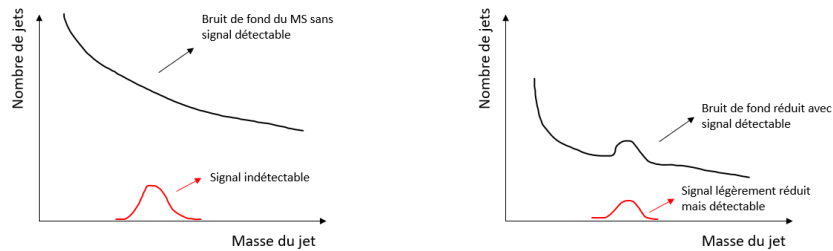


Fig. 4.1. Illustration du principe de recherche de bosses. À gauche, même si un petit pic existe autour de la masse d’une nouvelle particule offrant une résonance, il est trop petit pour être visible dans les données. À droite, le pic est légèrement plus grand, mais il revêt une importance plus grande puisque le bruit de fond a été diminué par un facteur beaucoup plus important. Un signal devient donc détectable autour de la masse du pic. La figure imite une échelle logarithmique, mais n’est pas à l’échelle.

Il existe un algorithme particulier pour faire cette chasse aux bosses : l’algorithme de *bump hunting* [48] que nous avons implémenté avec `pyBumpHunter` [110]. Il sert à trouver des excès d’événements dans des distributions (histogrammes) 1D ou 2D par rapport à ce qui serait attendu. Pour ce faire, il compare la distribution à analyser avec une distribution de référence, qui sont les deux distributions à fournir à l’algorithme. Il cherche ensuite où sont les excès ou les déficits de la distribution à analyser par rapport à ce qui serait attendu si elle respectait les mêmes proportions que la distribution de référence.

L’algorithme de *bump hunting* fait partie de la grande famille des algorithmes de tests d’hypothèse [94] dont le but est de vérifier si une hypothèse est confirmée ou rejetée. Cette famille comporte entre autres le test du χ^2 et de Kolmogorov-Smirnov. Plus spécifiquement,

ces algorithmes comparent des données D avec une hypothèse nulle H_0 .² Ces algorithmes séparent les deux ensembles en plusieurs fenêtres (dans notre cas, des fenêtres de masse, *mass bins*) identiques pour D et H_0 . Puis, un scalaire dénoté t est calculé qui quantifie la différence entre les deux distributions. Par exemple, dans le test du χ^2 , la valeur de t est donnée comme suit : [48]

$$t = \sum_i \left(\frac{d_i - b_i}{\sqrt{b_i}} \right)^2 \quad (4.1.1)$$

où d_i et b_i représentent les valeurs des données et de l'hypothèse de la fenêtre i , respectivement. Nous voyons que t tente de mesurer une certaine distance entre les deux distributions, et que plus D et H_0 diffèrent, plus t augmente. Lorsque nous comparons D à H_0 , il est coutume de dénoter cette valeur t_0 . Les différents algorithmes de test utilisent différentes définitions de t et c'est ce qui les différencie.

Ensuite, de multiples ensembles de pseudo-données sont générées à partir de l'hypothèse nulle et la valeur de t est calculée sur ces différents ensembles, permettant de former une distribution de la valeur de t attendue en échantillonnant H_0 . Puis, la valeur de t_0 obtenue avec l'ensemble de données D est comparée avec cette distribution. Si D suit l'hypothèse nulle, c'est-à-dire que D est un échantillonnage de H_0 , alors sa valeur de t_0 devrait être échantillonnée selon la distribution obtenue avec les pseudo-données. Il devient alors possible de trouver une valeur p (*p-value*) qui caractérise la probabilité que t soit plus grand ou égal à la valeur obtenue t_0 :

$$p\text{-value} \equiv P(t \geq t_0 | H_0). \quad (4.1.2)$$

Cette valeur p est interprétable comme étant la probabilité d'avoir un faux-positif, c'est-à-dire d'avoir une fausse découverte. En effet, plus cette valeur est petite, plus t_0 devient surprenant à obtenir en vertu de l'équation 4.1.2. Par exemple, si la valeur p vaut 0.001, alors D aurait une chance sur 1000 d'obtenir une valeur t_0 si élevée s'il était échantillonné à partir de H_0 . Dans ce cas, nous pourrions alors affirmer que D ne respectait pas l'hypothèse nulle avec une probabilité de 0.1% de se tromper.

La méthode de calcul du scalaire t dans le cadre de l'algorithme de *bump hunting* est quelque peu long et technique pour ce mémoire, mais sa définition est donnée en référence [48, page 11].

Dans notre cas, les distributions seront des histogrammes 1D du nombre de jets par fenêtre de masse, où la largeur des fenêtres est un paramètre de l'algorithme déterminé

² En physique des particules, H_0 devrait habituellement représenter un bruit de fond exempt de toutes données tandis que D devrait représenter des données qui contiennent à la fois le bruit de fond sur lequel il se calque et aussi, potentiellement, du signal.

au choix. La distribution D à analyser est celle de tous les jets (issus du signal ou du bruit de fond) ayant survécu à un seuil choisi d’anomalie et nous la comparerons avec les jets seulement issus du bruit de fond ayant survécu à ce même seuil (H_0). Même si cette technique n’est pas utilisable dans l’éventualité où nous appliquerions notre projet sur les données d’ATLAS (il serait impossible d’isoler le bruit de fond tel que nous le faisons), elle est moins biaisée que d’utiliser l’échantillon complet de signal et de bruit de fond avant la coupe d’événements. En effet, en retirant des événements jugés normaux – ou non anormaux –, la forme de la courbe demeure légèrement corrélée à la masse, ce qui occasionne des excès à hautes masses et des déficits à basse masse non désirés. Ainsi, dans une vraie analyse, nous pourrions plutôt utiliser des données de bruit de fond simulées par des logiciels de simulation de détecteur plus précis que les simulateurs publics tels que **Delphes**. Une autre option serait d’utiliser des échantillons de contrôle dans les vraies données où nous sommes confiants qu’il n’existe pas ou très peu de signal, par exemple à haute pseudo-rapidité, pour établir une estimation de la forme du bruit de fond. Peu importe l’avenue choisie, elle présenterait des erreurs systématiques qui seraient à évaluer. Puis, nous utiliserions la forme de ce bruit de fond isolé après l’application du seuil d’anomalie comme référence de distribution mère.

Concrètement, l’algorithme fournit une signification statistique de l’excès observé pour chaque fenêtre de masse. Ces significations sont données en unités d’écarts-types σ qui séparent la distribution réelle de la distribution attendue. En raison de sa nature, cet algorithme est robuste face aux fluctuations statistiques naturelles des distributions de masse. Ces significations constitueront la métrique de performance la plus importante de notre modèle lorsque nous le testerons avec des signaux connus.

4.2. Description de l’architecture du VAE

Nous avons tout d’abord commencé notre projet de recherche avec le VAE *simple*, c’est-à-dire sans exposition aux données hors distribution. Nous rendant compte de la corrélation importante du score d’anomalie à la masse des jets, nous avons ajouté le mécanisme d’exposition aux données hors distribution. Cette section définit tout d’abord le VAE simple, puis décrit les modifications à y apporter pour arriver au VAE avec exposition aux données hors distribution.

Tous les VAE ont été codés en Python à l’aide de la librairie TensorFlow. Ils lisent des fichiers de type hdf5 (*.h5*) où est enregistrée l’information des constituants des jets d’entraînement ou de test. Un modèle entraîné consiste en son architecture (nombre de couches, nombre de neurones, etc.) et l’ensemble de ses paramètres internes (poids et biais).

Toute l’analyse des résultats se fait aussi par Python et Tensorflow, que ce soit pour faire passer les données de test dans le modèle, en soutirer l’information de la sortie ou pour faire les graphiques des résultats.

4.2.1. Le VAE sans exposition aux données hors distribution

Pour mener à terme notre projet, nous avons utilisé un VAE qui respecte la structure donnée dans la section 3.3.1. Nous donnons en entrée les quadrivecteurs des particules constituant le jet (amas topologiques ou *topoclusters* explorés à la section 2.4.1).

Les quadrivecteurs sont représentés en coordonnées cartésiennes, soit sous la forme (E, p_x, p_y, p_z) , et nous n’utilisons que l’information des 100 constituants avec le plus haut p_T . Nous avons fait plusieurs tests en tentant de modifier le nombre de quadrivecteurs fournis, mais cela n’impacte pas ou peu les résultats. Au-delà d’environ 20 constituants, les particules sont très peu énergétiques et leurs quadrivecteurs sont remplis de valeurs très près de 0. Nous croyons que les modèles ont alors peu d’informations additionnelles à aller chercher sur ces constituants, et c’est ce qui expliquerait cette stabilité dans les résultats. Toutefois, la décision de conserver plus de constituants des jets n’influence pas négativement le modèle.

La fonction de perte qui a été utilisée respecte la forme standard d’un β -VAE qui avait été donnée à l’équation 3.3.9 :

$$\mathcal{L}_{VAE} = \text{MSE}(\mathbf{x}, \hat{\mathbf{x}}) + \beta D_{KL}(\mathbf{z}, \mathcal{N}(0, I)) \quad (4.2.1)$$

Pour rappel, le premier terme calcule la qualité de la reconstruction du VAE : plus la sortie $\hat{\mathbf{x}}$ ressemble à l’entrée \mathbf{x} , plus ce terme est petit. Le deuxième terme tente quant à lui de régulariser l’espace latent \mathbf{z} en amenant chacune des gaussiennes de celui-ci le plus près de $\mathcal{N}(0, 1)$, donnant ainsi un pouvoir de génération au modèle. Tel que mentionné à la section 3.3.1, cette capacité générationnelle aide le réseau à mieux comprendre la sous-structure des données, ce qui peut l’aider à mieux différencier le signal du bruit de fond.

L’architecture propose donc une couche d’entrée de dimension 400 (quatre entrées pour chacun des 100 constituants), puis des couches cachées de dimension 40 et 20 qui forment l’encodeur. La couche latente est de dimension 10, puis le décodeur est constitué de couches de dimension 20, 40 et de la couche de sortie de dimension 80, miroitant la structure de l’encodeur. Par manque de jets d’entraînement, nous utilisons une partie des jets QCD à la fois pour la validation du modèle lors de l’entraînement et pour l’évaluation du modèle une

fois l'entraînement terminé.

4.2.2. Le VAE avec exposition aux données hors distribution (OE-VAE)

L'un des buts du VAE exposé aux données hors distribution (OE-VAE pour *Outlier-Exposure VAE*) est de décorréler les résultats à la masse des jets. Pour ce faire, nous exposons durant l'entraînement le VAE à des « données hors distribution » (ou « données aberrantes »), soient des jets qui ne sont ni issus du signal, ni issus du bruit de fond à analyser. Nous demandons ensuite au OE-VAE d'offrir à ces jets une couche latente à la fois de mauvaise qualité et différente des jets QCD à l'aide de la fonction de perte (voir équation 4.2.2). L'OE-VAE s'étant entraîné sur des événements qu'il ne doit pas identifier comme du bruit de fond à basse masse de jet, il apprendra à ne pas associer une basse masse à une faible probabilité d'anomalie, décorrélant ainsi la masse au score d'anomalie. Cela a aussi l'effet d'aider le modèle à ne pas identifier les hautes masses comme anormales puisqu'il y a maintenant un meilleur équilibre entre les jets de haute et basse masse. Pour ce faire, il est crucial que les données aberrantes montrées lors de l'entraînement proposent des jets sur le même intervalle de masse que le bruit de fond. En effet, nous avons réalisé que les modèles sont sensibles aux ensembles hors distribution, et l'absence de ces données dans une fenêtre de masse diminue la valeur de la fonction de perte de tous les jets dans cette région.

Nous avons aussi réalisé que le p_T pouvait être corrélé aux résultats et au score d'anomalie au même titre que la masse. Ceci est néfaste pour le modèle, car il devient alors, de façon vulgarisée, un « catalogueur de haut p_T » (*high- p_T tagger*). Il est toutefois déjà facile d'appliquer un seuil sur le p_T , ce qui ne permet pas de découvrir de la nouvelle physique et nous désirons évidemment avoir un modèle plus robuste. Nous avons donc décidé de traiter le p_T et la masse de la même façon. Par exemple, il est préférable de proposer un même intervalle de p_T pour les deux types de jets montrés lors de l'entraînement. Ceci résulte en la volonté de proposer des événements hors distribution qui remplissent l'espace de phase engendré par la masse et le p_T des jets QCD. Plus généralement, nous tentons de remplir tout l'espace disponible pour des jets boostés produits au LHC puisque de futures générations de bruit de fond, notamment à haute masse et à haut p_T , peuvent agrandir l'espace à remplir.

Finalement, pour maximiser la décorrélation simultanée de la masse et du p_T , chaque jet QCD sera associé à un jet hors distribution qui possède simultanément une masse et un p_T très semblables aux siens lors de l'entraînement. Concrètement, nous forçons les deux jets à être sélectionnés dans la même fenêtre de l'espace 2D de masse et de p_T préalablement

quadrillé. Les deux jets compléteront parallèlement et indépendamment la passe avant du VAE, c'est-à-dire en utilisant les mêmes paramètres internes du modèle. Ils seront ensuite comparés l'un à l'autre dans la fonction de perte (voir équation 4.2.2), qui, au lieu de fournir une valeur de perte pour chaque jet, n'en produira plutôt qu'une seule pour les deux, qui sont vus comme une paire. Ce faisant, nous forçons davantage le OE-VAE à utiliser la sous-structure des jets pour les différencier, puisque la masse et le p_T seront toujours semblables pour les deux types de jets, et ainsi inutilisables comme données discriminantes.

Pour implémenter cette technique, nous rajoutons un terme à la fonction de perte du β -VAE standard vue à l'équation 4.2.1 qui prend en compte la nouvelle liaison des jets QCD avec ceux hors distribution :

$$\mathcal{L}_{OE-VAE} = \mathcal{L}_{VAE} + \lambda \text{RELU} \left[D_{KL}(\mathbf{z}, \mathcal{N}(0, I)) - D_{KL}(\mathbf{z}', \mathcal{N}(0, I)) + \text{marge} \right] \quad (4.2.2)$$

où \mathbf{z} est le vecteur de l'espace latent associé au jet QCD, \mathbf{z}' celui des jets hors distribution tandis que λ est un hyperparamètre scalaire jouant sur l'importance de ce nouveau terme dans la fonction de perte. Ce dernier terme tente d'augmenter la différence entre :

- (a) la distance séparant l'espace latent des données de bruit de fond et la loi centrée réduite multivariée ;
- (b) la distance séparant l'espace latent des données aberrantes et de cette même loi centrée réduite multivariée

à au moins *marge*, qui est un autre hyperparamètre du modèle. En effet, la fonction RELU (dont le rôle n'a rien à voir avec les fonctions d'activation entre les couches de neurones dans ce cas-ci) peut être réécrite comme $\text{RELU}(x) = \max(0, x)$. Nous comprenons alors que le modèle considère optimale la situation où :

$$D_{KL}(\mathbf{z}, \mathcal{N}(0, I)) + \text{marge} \leq D_{KL}(\mathbf{z}', \mathcal{N}(0, I)) \quad (4.2.3)$$

puisqu'il retournera une valeur de perte de 0, le minimum accordé par la fonction RELU. C'est effectivement le comportement attendu du modèle. Sinon, il retournera $D_{KL}(\mathbf{z}, \mathcal{N}(0, I)) - D_{KL}(\mathbf{z}', \mathcal{N}(0, I)) + \text{marge}$, qui grandit au fur et à mesure que les données hors distribution ont une qualité de l'espace latent trop similaire aux jets de bruit de fond. Le modèle doit donc envoyer les espaces latents des données aberrantes à une distance additionnelle d'au moins *marge* de la loi centrale réduite multivariée par rapport à la distance qui sépare déjà les données du bruit de fond à la valeur « parfaite » de $\mathcal{N}(0, I)$ pour obtenir une note parfaite.

Tangiblement, plus l'espace latent du bruit de fond est de bonne qualité, plus le premier terme est petit, ce que nous recherchons. De son côté, la magnitude du deuxième terme se veut maximale en vertu de sa soustraction à la fonction de perte et grandit au fur et

à mesure que l'espace latent des données hors distribution est de mauvaise qualité. Nous encourageons alors le comportement attendu du modèle à offrir un bon espace latent pour les jets du bruit de fond, et d'en avoir un mauvais sinon. Ce faisant, il doit traiter différemment les jets observés selon leur sous-structure. En combinant cette idée avec le fait que les jets aberrants ont une distribution de masse très semblable, voire identique à celle des jets de bruit de fond, il apprend encore davantage à utiliser la sous-structure des jets plutôt que leur masse pour identifier leur nature. D'ailleurs, seule la valeur de perte de l'espace latent (deuxième terme de l'équation 4.2.1) sera utilisée pour déterminer le score d'anomalie d'un jet puisque l'exposition aux données aberrantes n'affecte que l'espace latent. De plus, c'est la métrique qui performe le mieux. Notons que le deuxième terme de l'équation 4.2.2 ne peut être utilisé de la sorte, puisque nous n'utilisons les données aberrantes que lors de l'entraînement.

L'exposition aux données aberrantes permet aussi d'aider le modèle dans son pouvoir de discrimination simplement en montrant au modèle des jets qui ne sont pas issus du bruit de fond. Nous passons donc d'un entraînement non supervisé à un entraînement semi-supervisé. Ce faisant, il explore déjà des exemples anormaux lors de son entraînement. Par exemple, en n'entraînant que sur les jets QCD, le modèle a la chance d'en comprendre la sous-structure. Cependant, en lui montrant d'autres exemples qui n'appartiennent pas à la même catégorie, il peut apercevoir quelques exemples de caractéristiques de la sous-structure qui indiquent que le jet n'est pas issu du bruit de fond. Ainsi, pour statuer le score d'anomalie d'un jet donné, non seulement une non-ressemblance avec la QCD peut mener à un haut score, mais la ressemblance avec des jets fournis comme données aberrantes aussi.

Notons que l'effet des données hors distribution sur la performance du modèle a aussi été un aspect du projet sur lequel j'ai personnellement beaucoup travaillé. En effet, j'ai mené toutes sortes d'études sur le sujet. Par exemple, j'ai déjà tenté d'entraîner un modèle avec exposition aux données hors distribution en retirant tout événement hors distribution possédant une masse entre deux valeurs, par exemple entre 200 et 300 GeV, en variant l'intervalle retiré et notant s'il était dans le pic du signal recherché. En observant que tout jet contenu dans cet intervalle de masse avait par la suite des valeurs de fonction de perte beaucoup plus petites (avaient beaucoup plus de chances de se faire étiqueter comme « normal »), nous avons pu en déduire qu'il était important que les données hors distribution couvrent le même intervalle de masse que le bruit de fond. À l'aide de ces études, nous avons aussi porté une conclusion encore plus fondamentale : tout jet ayant été bien représenté par des données hors distribution, c'est-à-dire que des jets hors distribution avec des caractéristiques semblables à lui avaient été montrés lors de l'entraînement, avait plus de chances d'être identifié comme du signal. C'est entre autres pour cela que nous avons généré des données

hors distribution à deux et quatre sous-jets couvrant un très large spectre de masse et de p_T , pour assurer que cette représentation soit plus égale entre les différents jets et pour minimiser l'écart entre ceux de basse masse souvent mieux représentés que ceux de haute masse. D'ailleurs, plus les jets hors distribution sont variés, plus ils ont de chances de bien représenter un signal inconnu.

4.3. Description des échantillons

Pour entraîner et évaluer le modèle, nous avons généré plusieurs ensembles de données à l'aide des logiciels `Madgraph`, `Pythia` et `Delphes`. Ceux-ci proviennent de plusieurs processus physiques différents et sont présentés dans cette section. Bien que chaque ensemble de données puisse être utilisé comme signal, bruit de fond ou données hors distribution, ils sont tout de même séparés selon ces trois catégories selon leur rôle dans le projet.

4.3.1. Le bruit de fond

Pour entraîner notre modèle, nous utilisons des jets représentant le bruit de fond associé aux jets issus de la QCD. Ceux-ci sont créés simplement via la commande `Madgraph` suivante :

```
generate p p > j j
```

qui spécifie uniquement que la collision initiale proton-proton doit résulter en deux jets. Nous avons un total d'environ 8.5 millions de jets QCD à notre disposition. Ces jets ne peuvent provenir que de gluons et de quarks légers, c'est-à-dire tous les quarks et leurs antiquarks respectifs à l'exception du top et du bottom. Ces jets sont les plus fréquents au LHC et servent donc d'entraînement sur le bruit de fond. Lors de l'entraînement, nous demandons un p_T minimal de 450 GeV pour le jet ayant le plus grand p_T , puisque c'est une des conditions qui est demandée pour enregistrer un événement [70] dans le système d'acquisition de données que nous avons étudié à la section 2.3. C'est aussi une condition que nous imposons lors de la génération des événements. Avoir un modèle qui applique déjà les mêmes seuils qu'ATLAS servira au futur du projet. Ces jets sont reconstruits avec un rayon $R = 1$.

4.3.2. Les signaux

Viennent ensuite divers ensembles de *signal*. C'est ce que nous rechercherons après l'entraînement. Ces ensembles tentent d'imiter la nouvelle physique en se présentant comme des processus qui n'ont pas été observés par le VAE lors de l'entraînement. En effet, la nouvelle physique peut être vue comme un signal pour un modèle appliqué sur les vraies données. Ainsi, dans les deux cas, le VAE tente de détecter avec justesse un signal parmi un

bruit de fond. Or, tant et aussi longtemps que le VAE est incapable de détecter les signaux que nous avons nous-mêmes intégrés, nous savons qu’il demeure trop peu performant pour partir à la quête de nouvelle physique.

Parmi ces signaux générés, nous avons le quark top boosté qui est généré à partir de la commande suivante :

```
generate p p > t t~
```

qui demande à ce que la collision initiale résulte en un quark top et un antiquark top, où la nécessité d’avoir une particule et une antiparticule est requise par conservation de la charge électrique et de la couleur, entre autres. Nous demandons aussi qu’au moins un jet (pouvant provenir du top ou du anti-top) ait une quantité de mouvement d’au moins 450 GeV, ce qui garantit que le top (anti-top) en question sera à l’origine d’un jet boosté. Plus de 99.7% du temps [130, page 32], ces quarks top se désintégreront en un boson W et un quark b . Nous avons ensuite forcé le W à se désintégrer en une paire de quark-antiquark (autrement, il se désintègre un tiers du temps en leptons [130, page 32] qui ne sont presque jamais à l’origine de jets). Ceci mène à un jet boosté comportant trois sous-jets : l’un issu du quark b initial et deux issus de la paire quark-antiquark produite. Nous avons généré environ 9 millions de ces jets.

Ce signal a été choisi car il est résonant avec les jets créés centrés autour de la masse du top, qui est d’environ 172 GeV [130, page 32]. Il s’agit donc d’un bon échantillon pour tester la recherche de nouvelles particules avec l’algorithme de chasse aux bosses. Il est aussi un résultat du modèle standard et donc de « vraie » physique, ce qui simule réellement l’exposition du VAE à de la nouvelle physique qui existe. Aucun résultat avec ce signal ne sera présenté dans le reste de ce mémoire, mais il est mentionné car il a joué un rôle important dans notre projet en étant le premier que nous avons utilisé et ayant souvent servi de référence pour nos résultats.

Nous avons aussi simulé un signal basé sur un modèle hors modèle standard qui introduit un boson Z' [76], implémenté à l’aide du modèle *Top-Philic-Zprime* [85]. Cette nouvelle particule couple avec les quarks les plus lourds, donc prédominamment le quark top. Ainsi, il est très fréquent de voir des tops dans les produits de désintégration du Z' . Plus précisément, notre groupe a généré le processus suivant :

```
generate p p > Vz > t t~
```

où Vz représente le Z' et t et $t~$ des quarks (antiquarks) lourds de charge $+2/3$. Nous pouvons les voir comme jouant le rôle du top et du antitop à masse variable, qui sont des particules hypothétiques. Il s’agit ici d’un autre signal à 3 sous-jets, puisque nous nous assurons d’avoir un Z' suffisamment massif et avons encore une fois forcé une désintégration

hadronique du W . Il s’agit aussi d’un processus résonant, où la masse des jets devrait être agglomérée autour de la masse du top. Dans le modèle *Top-Philic-Zprime*, il nous a été possible de choisir indépendamment la masse du Z' et du top. Nous avons notamment généré un échantillon où la masse du Z' était de 5000 GeV tandis que celle du top était de 500 GeV (900,000 événements) et un autre où les masses étaient de 2000 et 200 GeV (820,000 événements). L’avantage de cet ensemble de données par rapport à celui du top du modèle standard est qu’il nous permet de créer des échantillons où le signal possédera un pic à diverses masses et où nous pouvons aussi contrôler la distribution en p_T des jets produits. Dans notre cas, non seulement la masse des jets n’est pas équivalente à la masse réelle du top de 172 GeV, mais puisque les tops proviennent d’une particule si massive, les jets possèdent en moyenne un p_T très élevé. Finalement, rappelons-nous que nous n’utiliserons que le jet à plus haut p_T produit dans chaque événement. Par exemple, si l’état final est composé d’un top et d’un antitop, chacun de ceux-ci produira un jet à trois sous-jets. Nous choisirons donc celui à plus haut p_T à étudier avec le VAE, tandis que l’autre sera écarté de l’étude.

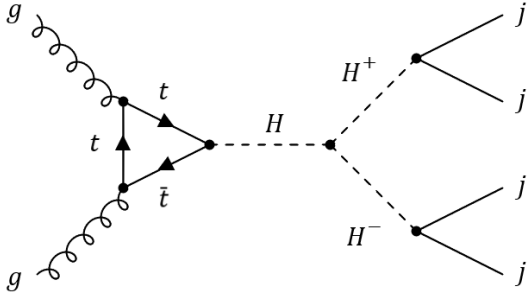
Les diagrammes de masse et de p_T de ces ensembles peuvent être observés à la section 4.4. Par exemple, la figure 4.6 montre les histogrammes de masse et de p_T des ensembles de données du bruit de fond QCD et du signal du Z' au pic à 200 GeV en vertu de la masse du top. Ils ne sont donc pas inclus dans cette section par souci de redondance.

4.3.3. Les données hors distribution

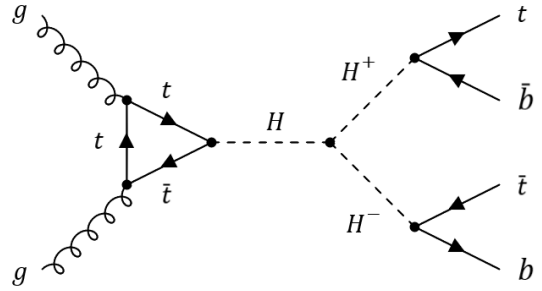
Pour la génération de jets hors distribution, nous avons voulu créer deux ensembles de données représentant des jets de grand rayon à 2 sous-jets (*2-prong*) ainsi qu’à 4 sous-jets (*4-prong*). Nous jugeons que le nombre de sous-jets dans un jet est une caractéristique suffisamment importante pour justifier la simulation de ces deux ensembles de données. De cette manière, nous avons voulu tester le modèle sur au moins deux échantillons ayant un nombre différent de sous-jets. Notons que les signaux du top et du Z' possèdent trois sous-jets, et ainsi, nous pouvons tester la capacité du modèle de généraliser l’idée d’anomalie au-delà d’une ressemblance avec les jets hors distribution fournis en entraînement. Malgré tout, nous voudrions idéalement posséder à la fois du signal et des données aberrantes pour tous les nombres de sous-jets possibles, notamment des données aberrantes à trois sous-jets.

Pour ce faire, nous avons généré des événements avec le modèle 2HDM (Modèle au doublet de deux Higgs), qui n’est pas inclus dans le MS [21]. Ce modèle suppose l’existence d’un deuxième doublet de Higgs, résultant entre autres en deux bosons de Higgs neutres h et H en plus de deux bosons chargés H^\pm . Si le H est significativement plus massif que les H^\pm , alors les Higgs chargés seront produits avec une impulsion importante, produisant

des jets boostés tel que vu à la section 1.3.6. Évidemment, c'est ce genre de processus boosté que nous avons généré. Crucialement, l'implémentation de ce modèle que nous avons utilisée pour générer les événements [68] permet entre autres de choisir manuellement la masse du H et des H^\pm , ce qui nous a autorisé de générer des jets boostés avec des masses et des impulsions diverses, remplissant l'espace de phase engendré par la masse et le p_T des jets QCD.³ La figure 4.2 montre les processus que nous avons générés pour les jets hors distribution à deux et quatre sous-jets.



(a) Diagramme de Feynman du processus $H^\pm \rightarrow jj$ généré comportant deux sous-jets.



(b) Diagramme de Feynman du processus $H^\pm \rightarrow tb$ généré comportant quatre sous-jets. Les tops dans l'état final montré se désintègreront hadroniquement pour un total de quatre sous-jets par Higgs chargé en vertu des trois provenant du top et de celui provenant du b .

Fig. 4.2. Diagrammes de Feynman représentant les processus des jets hors distribution générés à deux et quatre sous-jets. Deux gluons collisionnent et produisent un H via une boucle de fermions neutres (représentée ici par le top puisqu'elle est la plus courante en vertu du couplage du Higgs à la masse) ou de bosons W . Cette boucle est nécessaire puisque le gluon ne possède pas de masse. Le Higgs se désintègre ensuite en H^+ et en H^- , qui se désintègrent chacun en deux quarks légers qui produisent deux jets (figure 2(a)) ou qui se désintègrent en $t\bar{b}$ ou $\bar{t}b$ (figure 2(b)).

Comme mentionné précédemment, nous voulons proposer des événements qui couvrent plus ou moins de façon uniforme une région de l'espace en masse et en p_T . Or, il est possible d'estimer la masse et le p_T des jets à grand rayon ainsi générés. En effet, sa masse doit être équivalente à celle du H^\pm puisque le jet est boosté (voir section 1.3.6). Pour le p_T , le H est si massif qu'il est raisonnable d'approximer qu'il sera produit au repos, c'est-à-dire avec une impulsion nulle.⁴ Ainsi, en utilisant des règles de cinématique relativiste de base, via

³ C'est pourquoi les autres échantillons, comme le quark top et le Z' , n'ont pas été utilisés comme données hors distribution dans le projet : ils ne remplissent pas une grande région de l'espace de phase engendré par la masse et le p_T puisqu'ils n'ont pas été générés avec plusieurs configurations de masse.

⁴ Plus le H est massif, plus cette approximation devient juste. En effet, une grande quantité d'énergie doit être utilisée pour lui donner sa masse, ce qui limite considérablement l'énergie restante pour lui donner de la quantité de mouvement.

l'équation 4.3.1 :

$$E^2 = p^2 + m^2 \quad (4.3.1)$$

où E représente l'énergie de la particule, p sa quantité de mouvement et m sa masse et où l'énergie de repos d'une particule est égale à sa masse via $E = m$. Nous pouvons donc estimer que le p_T des jets équivaut à :

$$p_T \approx \sqrt{\left(\frac{m_H}{2}\right)^2 - m_{H^\pm}^2} \quad (4.3.2)$$

où nous avons approximé que $p = p_T$. Nous devons aussi nous assurer que les produits de désintégration de la particule boostée soient contenus dans un cône de rayon 1 dans le plan $\eta \times \phi$. En effet, rappelons-nous que c'est dans ce plan qu'est défini le rayon d'un jet et que l'algorithme de reconstruction utilise la valeur de 1. Or, pour des jets boostés, nous pouvons estimer la distance entre les produits de désintégration, dénotée ici R puisqu'elle représente le rayon nécessaire à sa reconstruction, comme valant [88, page 7] :

$$R \approx \frac{2m_{jet}}{p_T} \approx \frac{2m_{H^\pm}}{p_T} \quad (4.3.3)$$

où m_{jet} sa masse et p_T sa quantité de mouvement transverse et où la dernière approximation provient du fait que la masse du jet est environ égale à la masse du Higgs chargé. En remplaçant p_T par son expression de l'équation 4.3.2, nous trouvons après quelques manipulations mathématiques que la masse du H doit être au moins $\sqrt{20}$ fois plus grande que celle des H^\pm . Cette condition est très utile puisqu'elle ne fait intervenir que les deux masses que nous décidons de fournir au programme de simulation. Évidemment, plusieurs suppositions ont été utilisées pour en arriver à ces équations, mais ces dernières servent de bonne approximation de la position des jets dans l'espace engendré par leur masse et leur p_T . De plus, nous avons réalisé que les jets étaient effectivement agglomérés près de ces points une fois générés. Un exemple de cette approximation est donné à la figure 4.3.

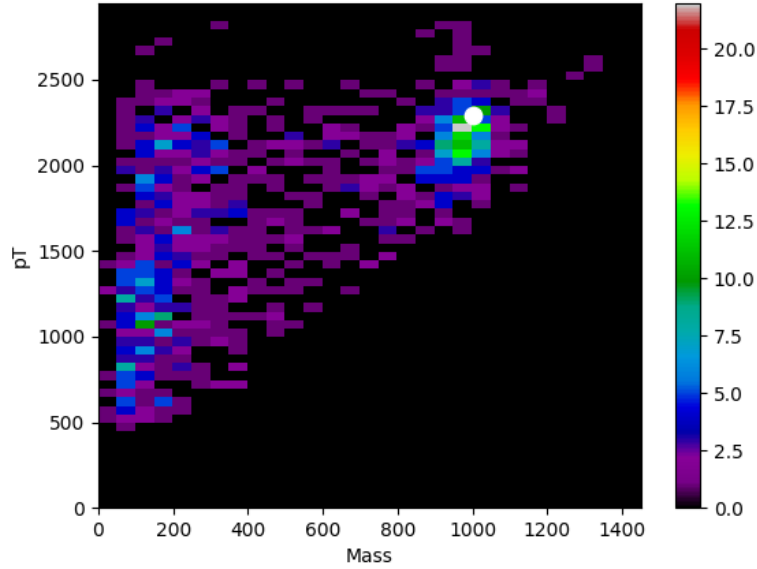


Fig. 4.3. Comparaison de la distribution des événements $H^\pm \rightarrow tb$ dans l’espace de phase 2D de masse et p_T et de la prédiction théorique pour la configuration de masse $\{m_H = 5000 \text{ GeV}, m_{H^\pm} = 1000 \text{ GeV}\}$. Le point blanc indique la masse et le p_T attendus via les calculs théoriques approximatifs. Nous verrons dans les prochains paragraphes que les événements ayant une masse sous 800 GeV ont été rejetés et pourquoi il y en a autant.

Finalement, les jets doivent encore une fois avoir un p_T plus grand que 450 GeV pour passer le seuil du système d’acquisition de données. En dernier lieu, puisque l’énergie de collision des protons est de 13.6 TeV, il devient impossible de générer des jets trop énergétiques, donc nous avons décidé de générer les jets jusqu’à $m_H = 5000 \text{ GeV}$.

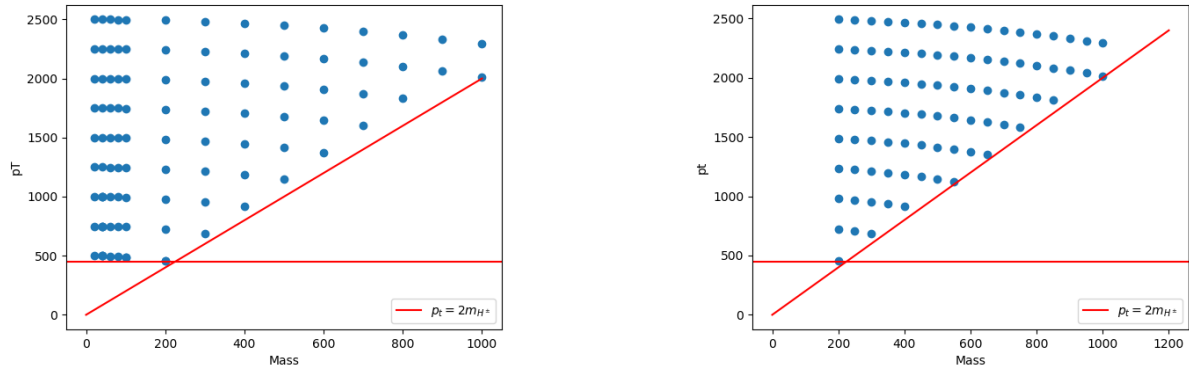
En résumé, nous devons respecter trois conditions lors de la génération des jets boostés. Premièrement, il faut qu’ils aient un p_T supérieur à 450 GeV. Ensuite, nous devons pouvoir les reconstruire dans un rayon plus petit que 1 dans le plan $\eta \times \phi$. Finalement, nous devons nous assurer que l’événement soit accessible aux énergies du LHC. De plus, puisque l’état final est composé d’un top et d’un bottom pour le processus $H^\pm \rightarrow tb$, il faut s’assurer que la masse du H^\pm soit au moins aussi grande que celle des deux quarks combinés d’environ 180 GeV.

De cette manière, nous avons généré toute combinaison respectant ces conditions en choisissant des masses du H allant de 1000 à 5000 GeV par intervalles de 500 GeV, ainsi que des masses de H^\pm étalées de 20 à 100 GeV, par intervalles de 20 GeV ainsi que de 200 à 1000 GeV, par intervalles de 100 GeV pour le processus $H^\pm \rightarrow jj$. Au total, 91 configurations

ont été retenues.

Pour le processus $H^\pm \rightarrow tb$, nous avons généré toute combinaison valide en choisissant des masses de H de 1000 à 5000 GeV par intervalles de 500 GeV, puis des masses de H^\pm de 200 à 1000 GeV par intervalles de 50 GeV. Au total, ce sont 87 configurations de masse qui ont été choisies. La figure 4.4 montre les p_T estimés via l'équation 4.3.2 et la masse attendue (qui est égale à celle du H^\pm) des configurations sélectionnées pour les processus $H^\pm \rightarrow jj$ et $H^\pm \rightarrow tb$.

Ainsi, nous avons généré 1000 événements par configuration de masse lors de la première phase de génération pour en analyser leur teneur afin de réajuster le tir au besoin . Finalement, comme nous le verrons dans le reste de cette section, ces tests ont permis d'élaborer des contraintes sur la masse des jets à conserver et du nombre de jets à générer pour chaque configuration de masse en vertu de leur densité autour de leur masse et leur p_T attendu.



(a) Masse et p_T attendus des différentes configurations de masses sélectionnées du processus $H^\pm \rightarrow jj$, qui possède deux sous-jets au coeur du jet de grand rayon généré.

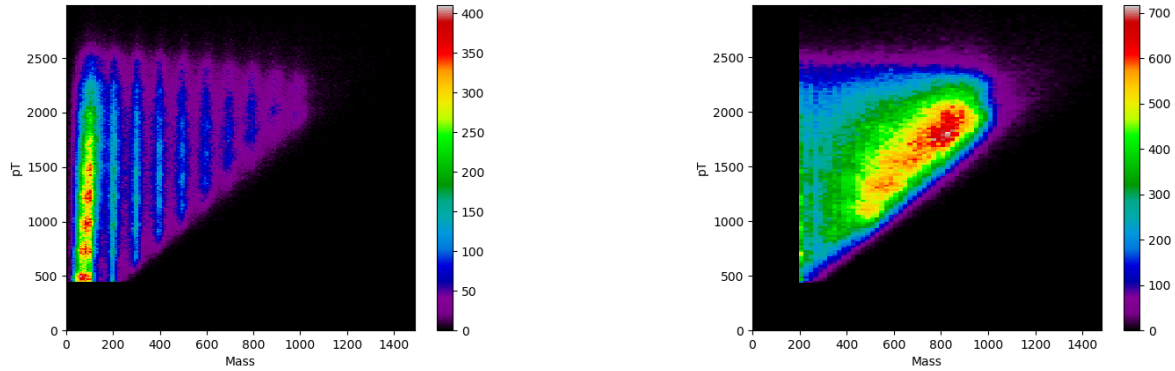
(b) Masse et p_T attendus des différentes configurations de masse sélectionnées du processus $H^\pm \rightarrow tb$, qui possède quatre sous-jets au coeur du jet de grand rayon généré.

Fig. 4.4. Masse et p_T attendus des différentes configurations de masse sélectionnées des jets hors distribution générés. Les lignes rouges obliques représentent le cas limite théorique $R = 1$ que les jets ne peuvent pas franchir ; cette limite est donnée par l'équation 4.3.3. Les lignes horizontales rouges représentent la limite de 450 GeV appliquée sur le p_T des jets.

Or, bien entendu, nous nous attendons à ce que les jets soient agglomérés autour de ces points pour chaque configuration donnée. Toutefois, il arrivait parfois que le jet ait une distance trop grande entre les produits de désintégration de la particule boostée dans le plan $\eta \times \phi$ pour qu'il soit entièrement reconstruit par l'algorithme de reconstruction. Ceci menait le jet à avoir une masse beaucoup plus petite que ce à quoi nous nous attendions

puisque certaines particules n'étaient pas comptabilisées dans le calcul de la masse du jet, étant trop éloignées. Ceci était particulièrement vrai pour les ensembles de données où la distance théorique entre les produits de désintégration de la particule boostée (voir équation 4.3.3) correspondait à un rayon de reconstruction qui se rapprochait de 1. C'est par ailleurs ce que l'on observe à la figure 4.3, où cette distance théorique valait 0.87. Le rayonnement QCD, où les quarks et les gluons peuvent émettre d'autres quarks et gluons, explique en partie ce phénomène. En effet, lors de cette fragmentation, il est possible qu'un nouveau parton (à l'origine d'un jet) apparaisse à grand angle. Ainsi, si le Higgs chargé à l'origine d'un jet produit des partons à grand angle à l'origine de jets, alors il n'est possible de reconstruire qu'un seul de ces jets dans un cône de rayon $R = 1$, ce qui explique la basse masse des jets. De plus, la probabilité d'émettre un parton à grand angle augmente lorsque le jet initial est de haute énergie. Ainsi, pour chaque configuration de masse, nous avons appliqué un certain seuil variable au-dessous duquel nous rejetons les jets produits. Par exemple, pour tous les ensembles de données ayant une m_{H^\pm} de 1000 GeV, nous avons rejeté tout jet dont le jet principal avait une masse sous 800 GeV. Finalement, pour s'assurer que l'espace engendré par la masse et le p_T des jets soit le plus équitablement produit, nous avons renormalisé le nombre de jets à générer pour chaque configuration de masse, non seulement avec l'inverse du ratio du nombre de jets conservés après la coupure, mais aussi en fonction de l'inverse du nombre de jets inclus autour du pic de résonance observé dans cet espace. Cette étape est importante : la configuration de masse la moins peuplée (et donc la plus dense autour de son pic attendu) n'a que 5684 événements simulés, tandis que les plus peuplées en comportent autour de 80 000. Ainsi, nous avons produit environ 1 million d'événements par processus (2 millions au total) conservés après l'application du seuil de masse, mais répartis selon un nombre très variable d'événements par configuration de masse lors de la seconde phase de génération destinée à l'entraînement des modèles.

La figure 4.5 montre ces distributions réellement produites en masse et en p_T des jets à grand rayon des processus générés conservés. Avec ces échantillons en poche, nous sommes maintenant prêts à utiliser le OE-VAE.



(a) Histogramme en deux dimensions de la masse et du p_T des jets conservés lors de la génération du processus $H^\pm \rightarrow jj$.

(b) Histogramme en deux dimensions de la masse et du p_T des jets conservés lors de la génération du processus $H^\pm \rightarrow tb$.

Fig. 4.5. Histogramme en deux dimensions de la masse et du p_T des jets conservés lors de la génération des événements hors distributions. Ces événements serviront d'exemples de jets hors distribution qui seront exposés au OE-VAE lors de son entraînement. La partie en bas à droite des histogrammes n'est pas peuplée puisqu'elle représente des jets dont les produits de désintégration de la particule boostée possèdent une distance qui correspond à un rayon supérieur à 1.

Cette génération non triviale a représenté une de mes plus importantes contributions au projet. Plus précisément, je connaissais la nature physique des événements à générer (voir figure 4.2) et je savais qu'il fallait remplir l'espace de phase en masse et en p_T . J'ai ensuite décidé de la méthode à employer, des configurations de masse à tester et à rejeter, de faire une première génération avec ces configurations, de vérifier leurs distributions en masse et en p_T , de réaliser que de nombreux événements possédaient une masse non conforme aux attentes, de choisir d'appliquer un seuil au-dessous duquel les événements étaient rejetés, de sélectionner la valeur variable de ces seuils et finalement de déterminer le nombre d'événements à sélectionner en élaborant le test du comptage d'événements autour du pic de masse attendu. D'un point de vue technique, j'ai profité d'un gabarit (*template*) d'un script de génération pouvant générer plusieurs processus différents que j'ai du adapter aux besoins de mes générations, devant notamment sélectionner des configurations de masse désirées. Notons que ce script faisait référence à plusieurs autres scripts, comme les cartes des paramètres ou de processus (`parameter_card`, `run_card`, `process_card` et ainsi de suite). J'ai légèrement modifié certaines de ces cartes en fonction de gabarits qui m'avaient été donnés dans le modèle 2HDM ou par des collègues. J'ai aussi profité de deux scripts généraux que j'ai pu modifier pour sélectionner quels événements conserver et quels retirer ainsi que pour convertir les fichiers de type *root* (qui sont produits lors de la génération) en fichiers *hdf5* (qui sont requis pour l'entraînement des modèles d'apprentissage automatique). Finalement, j'ai

fusionné les fichiers *hdf5* en deux gros fichiers pour l’entraînement, un par type de processus. Je dirais que mon travail a plutôt servi pour l’analyse et l’élaboration d’un plan solide pour s’assurer de bien remplir l’espace de phase. Évidemment, j’ai aussi fait un travail technique important, mais plutôt sur les outils d’analyse plutôt que sur les outils de génération.

4.4. Résultats du OE-VAE

Cette section présente les résultats obtenus par différents OE-VAE. Pour ce faire, nous présenterons toujours les mêmes six graphiques associés à la performance de deux OE-VAE provenant d’entraînements semblables, un utilisant les données aberrantes $H^\pm \rightarrow jj$ et l’autre utilisant $H^\pm \rightarrow tb$. Les modèles seront évalués sur les signaux du Z' avec pic de masse à 200 et 500 GeV selon la masse du top choisie.

Le tableau 4.1 décrit les hyperparamètres communs aux deux modèles. En fait, la seule différence entre eux réside en l’ensemble de données hors distribution utilisé. L’optimiseur de descente de gradient est une technique utilisée pour s’assurer d’une convergence du modèle lors de l’entraînement. Par exemple, l’optimiseur Adam [86] attribue un taux d’apprentissage dynamique à chaque paramètre entraînable du modèle en fonction de la sensibilité du modèle par rapport à ce paramètre. Les techniques de mise à l’échelle (*scalers*) [63] permettent quant à elles d’appliquer un traitement sur les données pour qu’elles soient groupées sur des intervalles plus restreints dans le but d’essayer d’obtenir des gradients plus égaux d’un exemple à l’autre. Sinon, des données proposant des valeurs extrêmes par rapport à l’ensemble des données peuvent mener à des fonctions de pertes gigantesques dont les gradients sont beaucoup plus importants que pour d’autres.

Une autre de mes contributions notables au projet aura été d’optimiser les valeurs de β et de λ . Puisque notre projet demeure en phase de recherche et de développement, l’optimisation des hyperparamètres n’a jamais été une priorité. En effet, les valeurs optimales peuvent changer d’une situation à l’autre et le peaufinage (*fine-tuning*) de ces valeurs est une perte de temps alors que des changements majeurs à notre projet ne sont pas exclus. J’ai tout de même optimisé ces valeurs en testant la performance de modèles possédant des valeurs différentes de ces hyperparamètres. Ainsi, j’ai entraîné deux modèles pour chaque configuration possible en sélectionnant β parmi les valeurs [0, 0.1, 0.2, 1, 2, 5, 10, 20, 100] et λ parmi [0, 0.1, 10, 50, 100, 200, 500, 1000, 2000]. Finalement, j’ai sélectionné la configuration optimale de $\beta = 2$, $\lambda = 50$ présentée au tableau 4.1. Celle-ci était perçue comme étant celle qui performait le mieux en analysant les résultats sur l’ensemble des données hors distribution et des signaux mis à notre disposition. Incidemment, les trois termes de la fonction de perte du OE-VAE (équations 4.2.2 et 4.2.1) sont environ

tous égaux lors de l’entraînement des modèles avec cette configuration, signifiant que le OE-VAE accorde plus ou moins une importance égale à la qualité de la reconstruction et de l’espace latent du bruit de fond ainsi qu’à la correction due aux données hors distribution.

Les modèles s’entraînent sur 7.5 millions de jets QCD et sont validés sur un ensemble disjoint d’un autre million de jets QCD. Faute de statistiques, cet ensemble de validation sera réutilisé lors de l’évaluation, qui se fera de concert avec 1 million de jets de signal, peu importe le signal choisi. Nous avons toutefois utilisé un poids de 10^{-3} pour les jets de signal afin d’imiter une application réelle où le signal est rare dans les données.

Hyperparamètre	Valeur
Taille des <i>batches</i> d’entraînement	10 000
Nombre d’époques	100
Optimiseur de descente de gradient (<i>optimizer</i>)	Adam
Rythme d’apprentissage utilisé lors de la descente de gradient	0.001
β	2
λ	50
Nombre de constituants (taille du vecteur d’entrée)	100 (400)
Technique de mise à l’échelle (<i>scaler</i>)	QuantileTransformer

Tableau 4.1. Liste des hyperparamètres communs aux deux modèles entraînés et évalués.

4.4.1. Avec les données hors distribution $H^\pm \rightarrow jj$

Commençons tout d’abord avec les résultats avec les données hors distribution $H^\pm \rightarrow jj$. Comme elles sont générées sur un très large spectre de masse, il est pertinent d’évaluer le modèle sur des signaux à basse et haute masses pour pouvoir en comparer les effets. Ainsi, un seul modèle a été entraîné et a ensuite été évalué de façon indépendante sur les signaux du Z' avec pic à 200 et 500 GeV selon la masse du top choisie. Les figures 4.6 et 4.7 présentent les résultats pour le signal de 200 GeV.

Discutons rapidement des graphiques qui seront présentés pour bien comprendre ce que les résultats représentent. Le premier graphique (figure 4.6(a)) représente la visualisation de l’algorithme de *bump hunting*, appliqué sur les données conservées au seuil optimal. Celui-ci est défini par le seuil qui offre la signification statistique maximale sur une seule fenêtre : pour chaque seuil, nous enregistrons la fenêtre de masse individuelle de la plus haute signification statistique, puis nous sélectionnons la valeur du seuil qui offre la valeur

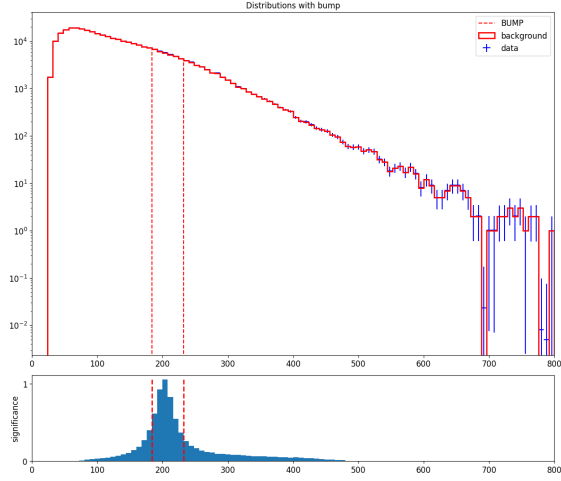
maximale de cette fenêtre. Nous y voyons la signification statistique de l’excès des données comprenant à la fois le signal et le bruit de fond par rapport à la courbe attendue du bruit de fond uniquement. Cette signification fluctue de fenêtre de masse en fenêtre de masse et les lignes rouges pointillées représentent l’intervalle où l’algorithme placerait le centre du pic observé. Le deuxième graphique (figure 4.6(b)) représente la signification statistique maximale atteinte sur une seule fenêtre de masse en fonction de l’efficacité du bruit de fond. Cette efficacité, représentée sur l’axe des abscisses, est déterminée en fonction du pourcentage de jets issus du bruit de fond qui survivent après l’application de différents seuils pour la valeur de la fonction de perte. La signification atteinte à 100% d’efficacité du signal, donc en utilisant l’ensemble des données, n’est pas nulle. Cela s’explique par la distribution de référence que nous utilisons, qui représente simplement la forme du bruit de fond une fois un seuil appliqué dans l’algorithme de *bump hunting*.⁵ Ainsi, en la comparant avec la distribution incluant le signal mais sans application de seuil d’anomalie, il y a déjà des excès même en conservant toutes les données. Nous devons donc nous intéresser à la variation de cette signification par rapport à son point initial et non pas à sa valeur absolue. Par exemple, une signification statistique importante de 5σ peut être associée à un mauvais modèle si la signification à 100% d’efficacité du bruit de fond est supérieure à 5σ .

Puis, les troisième et quatrième graphiques (figures 4.6(c) et 4.6(d)) représentent la distribution en masse ou en p_T , respectivement, des jets QCD (bruit de fond) et issus du signal, à la fois avant et après l’application du seuil d’anomalie optimal. Ces figures permettront de visualiser la corrélation de la masse et du p_T au score d’anomalie. Idéalement, les distributions avant et après l’application du seuil devraient avoir la même forme (les distributions où plusieurs événements sont supprimés devraient être une translation vers le bas des distributions de référence).

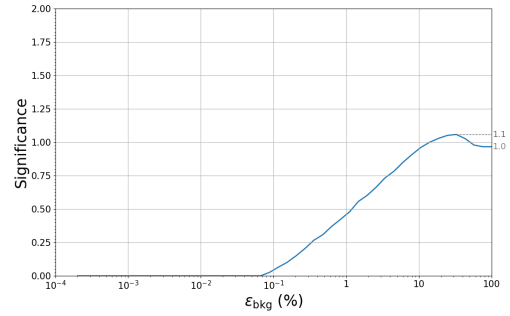
Finalement, nous présenterons aussi la distribution des valeurs des fonctions de perte des jets issus du signal en comparaison avec ceux issus du bruit de fond. Nous le ferons pour la qualité de la reconstruction des jets en sortie et pour la qualité de l’espace latent en cinquième et sixième graphiques, respectivement (figures 4.7(a) et 4.7(b)). Comme c’est en fonction de la valeur de la perte liée à l’espace latent (figure 4.7(a)) que le modèle sélectionne quels jets sont normaux et lesquels sont anormaux, ce graphique permet de visualiser la prise de décision du OE-VAE quant au score d’anomalie des jets. Idéalement, les jets issus

⁵ Chaque valeur d’efficacité du bruit de fond représente un seuil d’anomalie différent. Tel que décrit à la section 4.1, l’algorithme de *bump hunting* sera appliqué à chacune de ces valeurs de seuil en comparant la distribution totale des jets survivants au seuil (représentant la somme du bruit de fond et du signal) à la distribution des jets seulement issus du bruit de fond survivant à ce même seuil. Ainsi, à 100% d’efficacité du bruit de fond, nous comparons la distribution entière du signal et du bruit de fond par rapport à la distribution entière du bruit de fond, où des excès seront assurément observés.

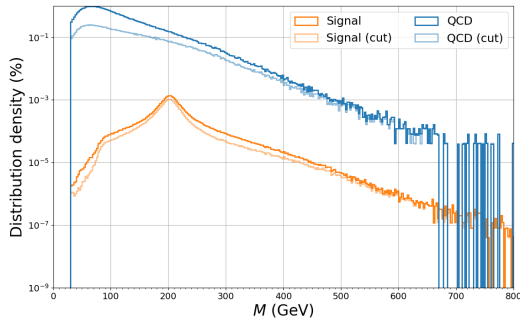
du signal devraient avoir une valeur de fonction de perte moyenne plus élevée que ceux du bruit de fond. En particulier, l'extrémité droite du graphique devrait être très concentrée en jets de signal, puisque cette extrémité survivra toujours aux seuils d'anomalie choisis.



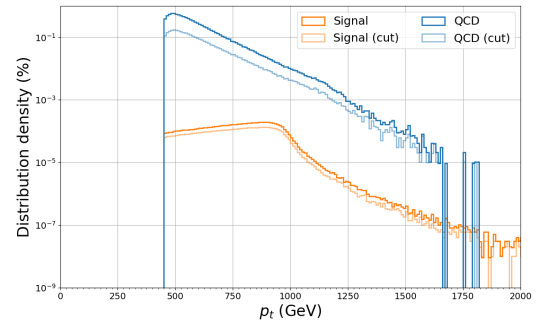
(a) Visualisation de l'algorithme de *bump hunting* appliqué au seuil d'anomalie optimal. Les cases en rouge représentent les données attendues par l'algorithme, tandis que les données en bleu représentent les jets conservés après l'application du seuil. La partie du bas montre la signification statistique des excès observés pour chaque fenêtre de masse.



(b) Signification statistique de l'excès retrouvé dans les données après l'application de divers seuils d'anomalie. L'axe des abscisses représente l'efficacité du bruit de fond, c'est-à-dire quelle proportion des événements QCD sont conservés après l'application des divers seuils.

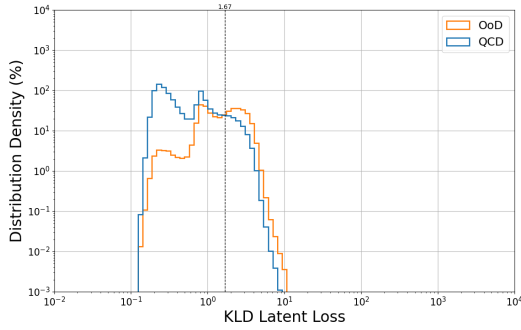


(c) Distributions de masse du bruit de fond QCD ainsi que du signal, à la fois avant et après l'application du seuil d'anomalie optimal.

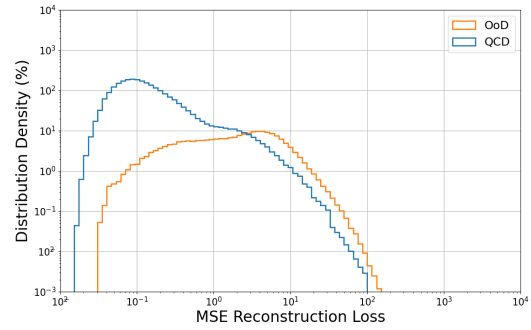


(d) Distributions en p_T du bruit de fond QCD ainsi que du signal, à la fois avant et après l'application du seuil d'anomalie optimal.

Fig. 4.6. Résultats obtenus pour l'entraînement effectué sur les jets hors distributions $H^\pm \rightarrow jj$ et une évaluation sur le signal du Z' avec pic à 200 GeV selon la masse du top choisie.



(a) Distributions de la perte liée à l'espace latent (deuxième terme de l'équation 4.2.1), à la fois pour le bruit de fond et le signal lors de l'évaluation du modèle. La ligne grise pointillée verticale correspond à la valeur de la fonction de perte correspondant au seuil optimal.

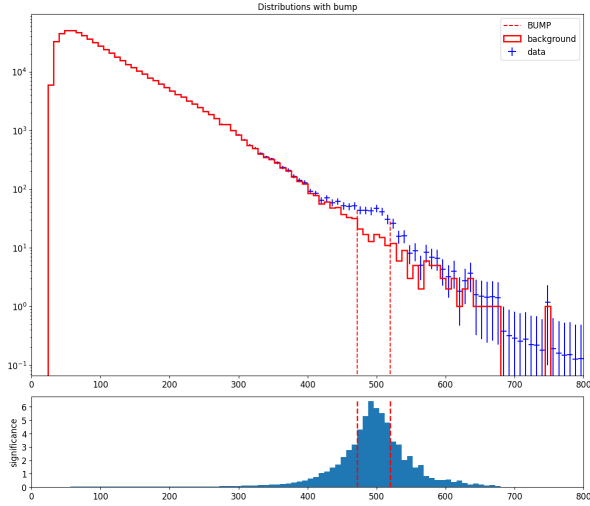


(b) Distributions de la perte liée à la reconstruction des jets (premier terme de l'équation 4.2.1), à la fois pour le bruit de fond et le signal lors de l'évaluation du modèle.

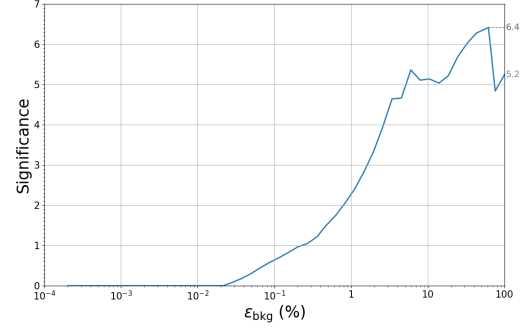
Fig. 4.7. Résultats obtenus pour l'entraînement effectué sur les jets hors distributions $H^\pm \rightarrow jj$ et une évaluation sur le signal du Z' avec pic à 200 GeV selon la masse du top choisie (suite).

Globalement, les résultats sont plutôt décevants. Le seuil optimal est atteint à une efficacité du bruit de fond d'environ 30% (figure 4.6(b)), ce qui ne permet pas une très grande suppression de ce dernier. De plus, l'augmentation par rapport aux données brutes n'est que de 0.1σ , ce qui signifie que le modèle ne possède pas un pouvoir de discrimination suffisant pour espérer trouver de la nouvelle physique. De plus, le score d'anomalie est corrélé à la masse, puisque presque tous les événements supprimés sont sous les 500 GeV (figure 4.6(c)). Par contre, le p_T est très bien décorrélé (figure 4.6(d)). Nous pouvons visualiser le pouvoir de discrimination du VAE aux figures 4.7(a) et 4.7(b), où le modèle peine à attribuer des valeurs de fonction de perte supérieures pour le signal. Notamment, la toute fin des distributions n'est pas suffisamment peuplée en jets de signal, rendant donc les seuils sévères inefficaces.

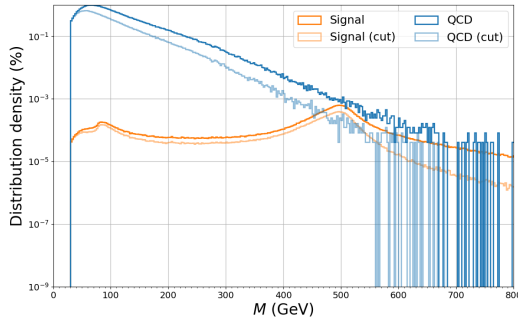
Toutefois, notons que le problème est particulièrement difficile à basse masse, puisque le pic d'événements de signal est situé où il y a énormément de jets QCD. Pour espérer que le signal devienne statistiquement intéressant, il faut alors supprimer énormément de bruit de fond puisqu'il est omniprésent à l'endroit du pic du signal, le tout en gardant une bonne décorrélation à la masse. Nous en demandons beaucoup au modèle. Les résultats pour le signal du Z' où la masse du top est de 500 GeV sont quant à eux présentés aux figures 4.8 et 4.9 et possèdent les mêmes six graphiques en ordre.



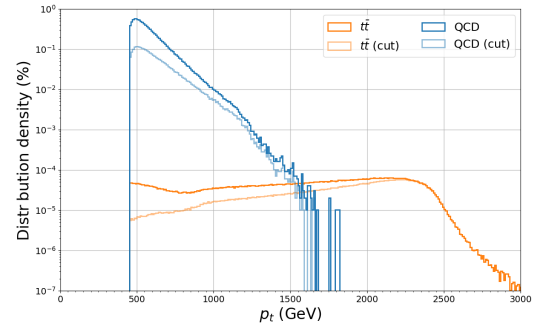
(a) Visualisation de l'algorithme de *bump hunting* appliqué au seuil d'anomalie optimal. Les cases en rouge représentent les données attendues par l'algorithme, tandis que les données en bleu représentent les jets conservés après l'application du seuil. La partie du bas montre la signification statistique des excès observés pour chaque fenêtre de masse.



(b) Signification statistique de l'excès retrouvé dans les données après l'application de divers seuils d'anomalie. L'axe des abscisses représente l'efficacité du bruit de fond, c'est-à-dire quelle proportion des événements QCD sont conservés après l'application des divers seuils.

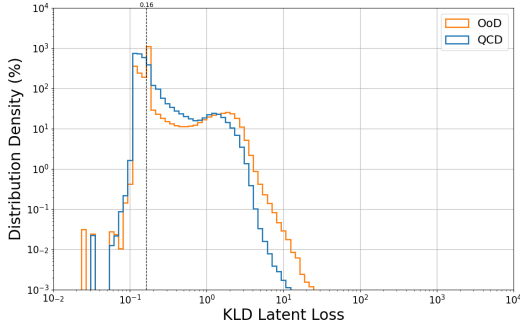


(c) Distributions de masse du bruit de fond QCD ainsi que du signal, à la fois avant et après l'application du seuil d'anomalie optimal.

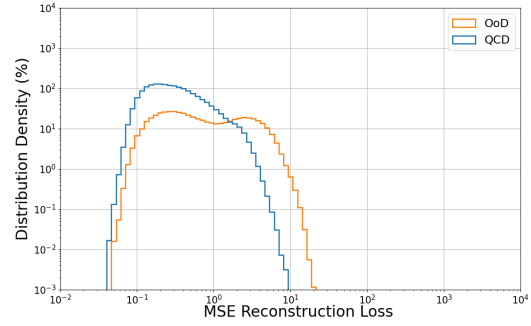


(d) Distributions en p_T du bruit de fond QCD ainsi que du signal, à la fois avant et après l'application du seuil d'anomalie optimal.

Fig. 4.8. Résultats obtenus pour l'entraînement effectué sur les jets hors distributions $H^\pm \rightarrow jj$ et une évaluation sur le signal du Z' où la masse du top et du signal est de 500 GeV.



(a) Distributions de la perte liée à l'espace latent (deuxième terme de l'équation 4.2.1), à la fois pour le bruit de fond et le signal lors de l'évaluation du modèle. La ligne grise pointillée verticale correspond à la valeur de la fonction de perte correspondant au seuil optimal.



(b) Distributions de la perte liée à la reconstruction des jets (premier terme de l'équation 4.2.1), à la fois pour le bruit de fond et le signal lors de l'évaluation du modèle.

Fig. 4.9. Résultats obtenus pour l'entraînement effectué sur les jets hors distributions $H^\pm \rightarrow jj$ et une évaluation sur le signal du Z' où la masse du top et du signal est de 500 GeV (suite).

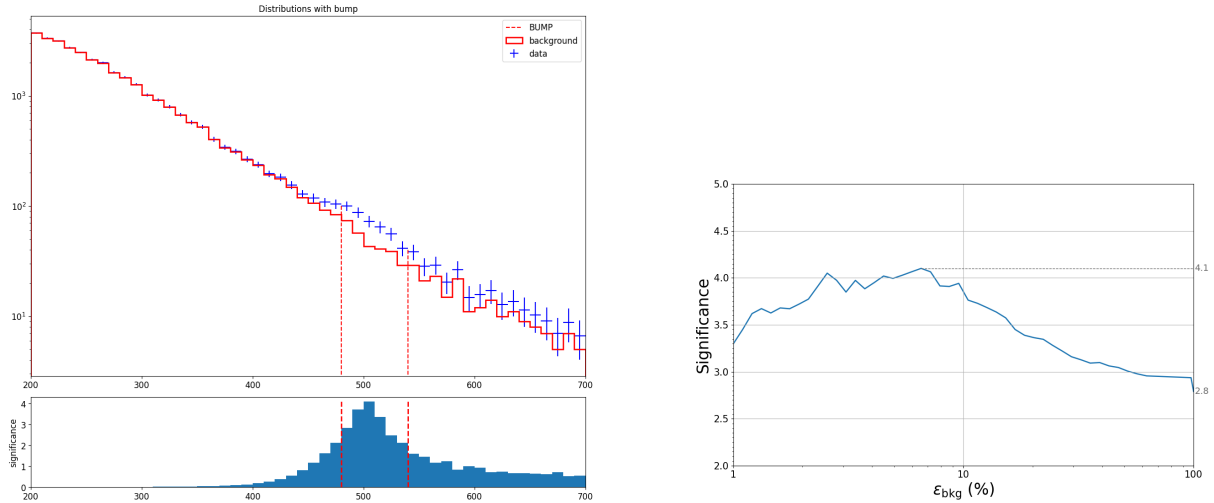
Dans ce cas, le modèle performe beaucoup mieux. Nous voyons tout de suite aux figures 4.8(a) et 4.8(b) que la signification statistique est bien meilleure, où le seuil optimal nous fait gagner 1.2σ . Toutefois, le seuil optimal demeure à une grande efficacité de bruit de fond aux alentours de 60%. Cela diminue nos probabilités de retrouver de la nouvelle physique dans une application réelle, puisqu'une telle coupe risque de ne pas diminuer suffisamment le bruit de fond pour voir apparaître un signal plus rare. Il serait donc préférable que ce seuil corresponde à une suppression plus sévère. En revanche, si nous tentions d'appliquer un tel seuil, il serait inefficace puisque la signification statistique diminue rapidement sous les 5.2σ initiaux des données brutes.

La figure 4.8(c) montre que la corrélation à la masse demeure, mais est moins importante, et semble presque parfaite sous les 500 GeV, surtout pour le signal. Étrangement, les hautes masses sont considérées comme étant plus normales que les basses masses. Le p_T s'en tire bien lui aussi à la figure 4.8(d), où la corrélation est la plus forte à bas et à haut p_T , mais qui n'est pas mauvaise entre 1000 et 1500 GeV. Il n'est cependant pas étonnant de voir que les jets du signal qui ont des p_T supérieurs à environ 1700 GeV, là où il n'y a pas de jets QCD lors de l'entraînement, soient considérés comme étant presque tous anormaux. En outre, le modèle possède une belle discrimination dans les valeurs des fonctions de perte entre les jets du signal et de QCD aux figures 4.9(a) et 4.9(b), beaucoup plus impressionnante que celles du modèle précédent présentées aux figures 4.7(a) et 4.7(b). Cela est surtout vrai en observant les parties les plus à l'extrémité droite des graphiques, où le présent modèle comporte une

beaucoup plus grande concentration de jets de signal. Le problème était toutefois beaucoup plus facile cette fois puisqu’une diminution moindre du bruit de fond permettait d’obtenir de bons résultats et parce que le pic des jets de signal était contenu dans une région de p_T dans laquelle le modèle ne s’est jamais entraîné à reconnaître des jets QCD.

4.4.2. Avec les données hors distribution $H^\pm \rightarrow tb$

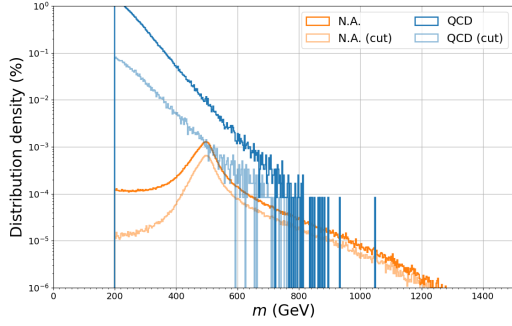
Puisque les données hors distribution $H^\pm \rightarrow tb$ n’ont été générés qu’à partir d’une masse des Higgs chargés H^\pm de 200 GeV en vertu de la grande masse-énergie de leur état final, nous n’avons utilisé que des jets QCD et hors distribution de 200 GeV et plus lors de l’entraînement. Pour ne pas avoir à chercher un signal dans une fenêtre de masse où le VAE n’a pas été entraîné, nous n’évaluons donc le modèle que sur des résonances à 500 GeV. Nous présenterons ici les résultats obtenus pour le signal du Z' où la masse du top, et ainsi le pic du signal, est de 500 GeV. Encore une fois, les mêmes six graphiques dans le même ordre sont proposés.



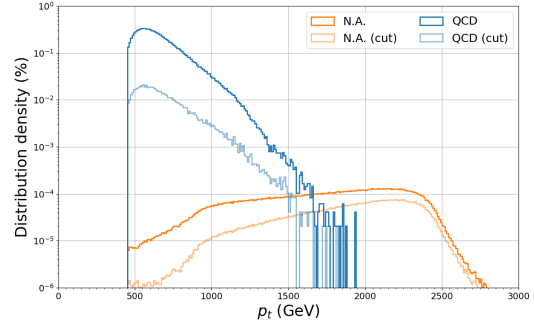
(a) Visualisation de l’algorithme de *bump hunting* appliqué au seuil d’anomalie optimal. Les cases en rouge représentent les données attendues par l’algorithme, tandis que les données en bleu représentent les jets conservés après l’application du seuil. La partie du bas montre la signification statistique des excès observés pour chaque fenêtre de masse.

(b) Signification statistique de l’excès retrouvé dans les données après l’application de divers seuils d’anomalie. L’axe des abscisses représente l’efficacité du bruit de fond, c’est-à-dire quelle proportion des événements QCD sont conservés après l’application des divers seuils.

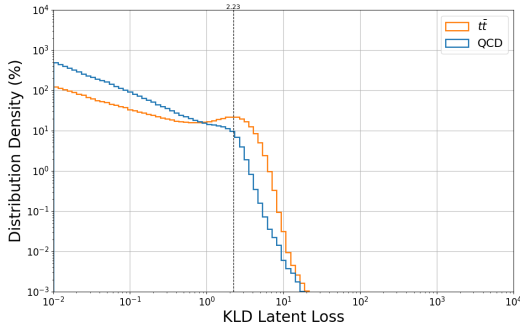
Fig. 4.10. Résultats obtenus pour l’entraînement effectué sur les jets hors distributions $H^\pm \rightarrow tb$ et le signal du Z' à 500 GeV en vertu de la masse du top sélectionnée.



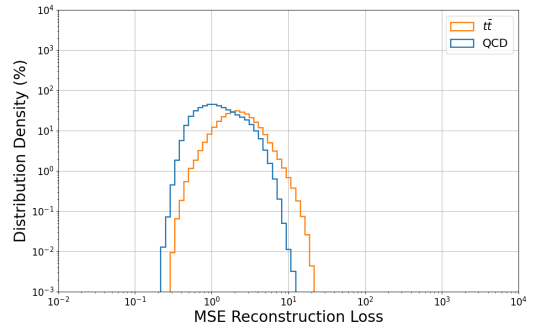
(a) Distributions de masse du bruit de fond QCD ainsi que du signal, à la fois avant et après l'application du seuil d'anomalie optimal.



(b) Distributions en p_T du bruit de fond QCD ainsi que du signal, à la fois avant et après l'application du seuil d'anomalie optimal.



(c) Distributions de la perte liée à l'espace latent (deuxième terme de l'équation 4.2.1), à la fois pour le bruit de fond et le signal lors de l'évaluation du modèle. La ligne grise pointillée verticale correspond à la valeur de la fonction de perte correspondant au seuil optimal.



(d) Distributions de la perte liée à la reconstruction des jets (premier terme de l'équation 4.2.1), à la fois pour le bruit de fond et le signal lors de l'évaluation du modèle.

Fig. 4.11. Résultats obtenus pour l'entraînement effectué sur les jets hors distributions $H^\pm \rightarrow tb$ et le signal du Z' à 500 GeV en vertu de la masse du top sélectionnée (suite).

Ici, le pouvoir de discrimination semble évident aux figures 4.10(a) et 4.10(b) alors que le modèle augmente la signification statistique du signal de 1.3σ et ce, à une efficacité du bruit de fond record de 7%. Malheureusement, cela se fait au détriment de la décorrélation de la masse et du p_T , qui perd de son lustre par rapport aux résultats de la section 4.4.1. En effet, bien que la corrélation à la masse soit acceptable (figure 4.11(a)), nous voyons que le VAE a tout de même tendance à conserver plus de jets à haute masse et à rejeter plus de jets à basse masse, tant pour le signal que pour le bruit de fond. Il conserve toutefois beaucoup plus de jets de signal que de jets de bruit de fond, toutes proportions gardées, surtout autour du pic de masse du signal, ce qui est encourageant. Nous observons un

comportement semblable pour le p_T à la figure 4.11(b), où le VAE rejette davantage les événements à bas p_T et juge anormaux les jets à haut p_T . Toutefois, nous devons encore noter que l'échantillon du signal présente un p_T qui couvre un plus large intervalle que l'échantillon QCD. Le pic du p_T du signal est notamment compris dans cette région. Il est donc normal que de tels jets, dont le modèle n'a jamais vu d'équivalent QCD, les traite d'anormaux. Finalement, la figure 4.11(c) montre que la discrimination du modèle est ambiguë : bien qu'il y ait un excès clair des jets de signaux à haute fonctions de perte dans l'espace latent, il reste tout de même plusieurs jets issus de la QCD. Toutefois, le graphique est quelque peu trompeur puisque l'axe des y est logarithmique. La figure 4.11(d) possède une séparation plus claire à haute valeurs de perte, mais rappelons que la perte liée à la reconstruction n'est pas utilisée pour la détermination du seuil optimal.

En conclusion, bien que la problématique que nous tentons de résoudre soit plus facile avec des signaux à haute masse où le bruit de fond doit diminuer de façon moindre pour voir la bosse du signal, il est encourageant d'observer une augmentation significative de la signification statistique à une si basse efficacité du signal. De plus, nous voyons que ce seuil conserve, toutes proportions gardées, beaucoup plus de signal dans son pic de masse que de jets QCD, une autre preuve directe de son pouvoir de discrimination.

4.5. Discussion

Nous constatons des résultats mitigés des performances de nos modèles. Entre autres, la décorrélation de la masse et du p_T coûte trop cher à établir actuellement. Les résultats utilisant le VAE simple, qui n'ont pas été présentés dans ce mémoire, présentaient un bien meilleur pouvoir de discrimination au coût d'une corrélation à la masse beaucoup plus importante. Le OE-VAE permet de décorréler ces deux variables au score d'anomalie, mais il ne le fait pas parfaitement, en plus de diminuer considérablement le pouvoir de discrimination des modèles. Bien que nous nous attendons à une telle diminution, celle que nous observons actuellement possède une magnitude tout simplement trop importante. Il nous faut donc penser à des solutions qui permettent une décorrélation de la masse et du p_T de façon plus efficace. De plus, une partie du pouvoir de discrimination sur l'échantillon du Z' où la masse du top est de 500 GeV doit être attribuée aux jets à haut p_T qui n'ont pas vu de semblable QCD lors de l'entraînement.

Nous sommes cependant conscients de la difficulté de notre projet. En effet, distinguer la nature de jets en fonction des quadrivecteurs de leurs constituants est un défi colossal, rendu encore plus difficile en vertu de la demande claire de décorrélation de masse et de p_T . De plus, notre projet se fait via un entraînement semi-supervisé, qui possède des

étiquettes de bruit de fond mais pas de signal, ce qui complique la tâche aux OE-VAE. Il est ambitieux de vouloir faire un catalogueur anti-QCD (*anti-QCD tagger*) qui ne s’entraîne que sur des jets QCD et qui voit quelques exemples de jets qui ne sont pas du QCD, mais sans pour autant être du signal. De plus, nous avons arbitrairement choisi un signal qui représente un millième des jets QCD. Toutefois, si le signal que nous rechercherons dans les vraies données d’ATLAS est encore plus rare, ce qui est très possible, la difficulté du problème augmente encore de beaucoup. Non seulement faut-il un modèle encore plus robuste dans son pouvoir de discrimination (nous devons rejeter encore beaucoup plus de jets QCD, toutes proportions gardées, que de jets de signal), mais il nous faut aussi beaucoup plus de statistiques pour tester le modèle. Par exemple, si le signal représente un millionième de la taille des jets QCD, alors nous serons forcés d’utiliser une suppression très importante, sous 1% d’efficacité du bruit de fond, pour espérer voir du signal. Actuellement, un tel seuil ne permettrait pas d’avoir des résultats intéressants faute de statistiques.

Ainsi, peut-être que l’envergure du projet pourrait être revue à la baisse, en ne se concentrant que sur des signaux à haute masse, où le pouvoir de discrimination n’a pas besoin d’être si important pour espérer voir du signal. Par contre, une telle approche demanderait une génération importante de jets à haute masse que nous ne possédons pas pour le moment.

4.6. Voies d’amélioration possibles et futur du projet

Malgré des résultats mitigés, notre groupe conserve en tête plusieurs idées pour améliorer la performance de nos modèles, dont certaines sont déjà soit en développement, soit déjà développées et en phase de test. Par exemple, nous avons déjà instauré la possibilité d’utiliser certaines variables de haut niveau des jets (HLV) discutées à la section 2.5.1 en remplacement des quadrivecteurs des particules du jet en entrée. Les résultats que nous obtenons sont visiblement meilleurs que ceux obtenus avec les quadrivecteurs, mais malheureusement insuffisamment polis pour être inclus dans ce mémoire.

Nous avons aussi pour plan de générer plus de données aberrantes de processus différents, en particulier à hautes masses. Par exemple, nos seuls exemples de signaux sont les processus $Z' \rightarrow t\bar{t}$, $H \rightarrow H^+H^- \rightarrow jj$ et $H \rightarrow H^+H^- \rightarrow tb$, dont les deux derniers sont aussi utilisés comme données aberrantes et n’ont jamais été utilisés comme étant du signal lors de ce mémoire. C’est donc dire que pour nos données hors distribution $H \rightarrow H^+H^- \rightarrow tb$, nous n’avons que deux signaux à tester à haute masse tandis que les signaux à basse masse ne sont pas utilisables en vertu de la masse minimale de 200 GeV utilisée lors de la génération de ces événements. Nous voulons aussi avoir accès à un plus grand espace de phase des jets

issus du bruit de fond. En effet, nos données proposent des jets jusqu'à une masse d'environ 700 GeV et un p_T de plus ou moins 1600 à 1700 GeV, ce qui ne nous permet pas d'aller sonder des signaux à haute masse et/ou haut p_T puisque le modèle n'aura pas été entraîné à ces endroits, ou de le faire avec des biais importants, tel que nous l'avons vu dans ce texte.

Nous avons aussi quelques idées pour potentiellement améliorer notre performance en peaufinant nos outils d'analyse sans changer les modèles que nous utilisons. À titre d'exemple, nous avons implémenté une technique de décorrélation de la masse et du p_T qui permet une décorrélation parfaite des jets QCD lors de l'entraînement et presque parfaite dans l'évaluation (pour les jets QCD). Celle-ci provient de nos collaborateurs de l'Université de Genève [87] qui l'ont tout d'abord développée avant que nous l'implémentions nous-mêmes dans notre projet. Qui plus est, cette méthode ne nécessite pas l'utilisation de données aberrantes. Pour ce faire, nous n'entraînons qu'avec les jets du bruit de fond, puis nous décidons quelle fraction des jets doit être conservée dans chaque fenêtre de masse et de p_T . C'est donc dire que nous séparons les données en deux catégories, et ce, fenêtre par fenêtre : les jets faisant partie de l'ensemble anormal et les jets ne faisant pas partie de l'ensemble anormal. Ensuite, lors de l'étape de la validation de la dernière époque, nous trouvons le « dernier » jet faisant partie de la catégorie anormale, c'est-à-dire celui qui possède la plus petite valeur de la fonction de perte. Nous trouvons donc le jet le plus normal parmi les jets anormaux, pour ainsi dire. Nous utilisons ensuite cette valeur pour déterminer quels jets sont anormaux lors de l'évaluation du modèle. En théorie, les jets QCD anormaux seront presque totalement indépendants simultanément de la masse et du p_T , et c'est aussi ce que nous observons expérimentalement. Ainsi, seuls les jets du signal qui sont plus anormaux que leurs compatriotes QCD dans une même fenêtre de masse et de p_T seront jugés anormaux. Cette technique permet aussi de contourner le problème où les valeurs absolues de la fonction de perte augmentent avec la masse et le p_T . En effet, nous comparons les valeurs de ces fonctions de façon relative entre les jets dans un même voisinage topologique plutôt que de se servir de leurs valeurs absolues pour déterminer le degré d'anomalie d'un jet.

Conclusion

Ce mémoire a traité de notre projet de recherche qui se veut novateur en utilisant l'apprentissage automatique pour rechercher de la nouvelle physique dans les données d'ATLAS. L'élément le plus important de l'utilisation de cet outil s'avère à être la recherche des signes de cette nouvelle physique de façon indépendante des modèles théoriques. Cette façon novatrice de procéder pourrait accélérer ou même débloquer la recherche de nouvelle physique.

Nous avons tout d'abord décrit le modèle standard actuel au premier chapitre dans lequel ont été étalées les forces et les faiblesses de cette théorie. Il y aussi a été révélé que bien que le modèle standard soit très efficace pour ce qu'il décrit, il existe plusieurs phénomènes observés qu'il ne peut expliquer. À la fin de ce chapitre, nous avons aussi fait un bref survol de plusieurs études réalisées au LHC qui ont recherché de la nouvelle physique sous forme de modèles théoriques existants. Ces recherches n'ont pas abouti, motivant notre désir de procéder autrement avec l'élaboration d'une recherche via l'intelligence artificielle.

Puis, le deuxième chapitre a décrit l'appareillage expérimental de l'expérience ATLAS mis à notre disposition. Nous avons discuté de l'accélérateur en tant que tel, le LHC, mais aussi décrit l'architecture et le fonctionnement du détecteur ATLAS, en plus d'avoir abordé la détection et la reconstruction des jets au sein d'ATLAS.

Le troisième chapitre a enchaîné en décrivant le fonctionnement de l'apprentissage automatique. Nous avons pu comprendre comment les réseaux de neurones s'entraînent et évaluent des données lors de leur application. L'architecture des auto-encodeurs et des auto-encodeurs variationnels a été détaillée. Une analyse a indiqué pourquoi il s'agit d'architectures très performantes dans le cadre de détection d'anomalies.

Le mémoire s'est conclu avec le quatrième chapitre avec les détails de notre démarche scientifique dans le cadre de ce projet. Il a été expliqué comment nous avons entraîné des réseaux de neurones à reconnaître des signes de nouvelle physique sous forme d'anomalies

dans les données. Nous avons aussi discuté du problème de la corrélation de la masse avec le score d'anomalie des jets et de la solution utilisée pour remédier à ce problème, soit l'exposition aux données hors distribution. Pour ce faire, il a été décrit précisément comment nous avons procédé pour simuler ces événements, ce qui a été un long processus et qui a représenté un des efforts les plus notables à ma contribution personnelle au projet. Finalement, les résultats de nos modèles ont été présentés, qui se veulent mitigés. En effet, les réseaux de neurones sont en mesure d'accéder à une bonne discrimination entre le signal et le bruit de fond, mais cela vient au coût de la corrélation à la masse, ce qui est inacceptable dans le cadre d'une recherche par résonance. Puis, l'exposition aux données hors distribution permettait de réduire cette dépendance, mais en paire avec une diminution du pouvoir de discrimination. Nous devons donc poursuivre nos efforts pour être en mesure d'accéder à cette décorrélation sans nuire d'une telle façon à la capacité de discrimination du réseau. Heureusement, nous avons ensuite déterminé des innovations possibles et réalistes à la poursuite de cet objectif. Par exemple, la simulation d'événements à de plus larges régions de masse et de p_T serait bénéfique pour améliorer les performances des modèles sur ces intervalles. Nous avons en outre déjà développé une technique de décorrélation de la masse qui s'applique lors de l'analyse des résultats plutôt que lors de l'entraînement. Ainsi, il reste de nombreuses idées pour poursuivre ce projet et améliorer la qualité des modèles entraînés.

Finalement, nous avons exploré quelques unes de mes contributions personnelles au projet, comme l'optimisation des hyperparamètres β et λ et l'étude de l'effet des données hors distribution sur la performance des modèles.

Références bibliographiques

- [1] M. et al. Aaboud. Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at $\sqrt{s} = 13$ TeV with the atlas detector. *Phys. Rev. D*, 96:072002, Oct 2017. URL: <https://link.aps.org/doi/10.1103/PhysRevD.96.072002>, doi:10.1103/PhysRevD.96.072002.
- [2] G. et al. Aad. Topological cell clustering in the atlas calorimeters and its performance in lhc run 1. *The European Physical Journal C*, 77(7):490, Jul 2017. doi:10.1140/epjc/s10052-017-5004-5.
- [3] M. Agostini, G. Benato, S. Dell’Oro, S. Pirro, and F. Vissani. Discovery probabilities of majorana neutrinos based on cosmological data. *Phys. Rev. D*, 103:033008, Feb 2021. URL: <https://link.aps.org/doi/10.1103/PhysRevD.103.033008>, doi:10.1103/PhysRevD.103.033008.
- [4] Q. et al. Ahmad. Measurement of the rate of $\nu_e + d \rightarrow p + p + e^-$ interactions produced by 8b solar neutrinos at the sudbury neutrino observatory. *Phys. Rev. Lett.*, 87:071301, Jul 2001. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.87.071301>, doi:10.1103/PhysRevLett.87.071301.
- [5] Evgeny Akhmedov. *Majorana neutrinos and other Majorana particles: Theory and experiment*. 12 2014. arXiv:1412.3320.
- [6] Guido Altarelli and James Wells. *Gauge Theories and the Standard Model*, pages 1–25. Springer International Publishing, Cham, 2017. doi:10.1007/978-3-319-51920-3_1.
- [7] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *Journal of High Energy Physics*, 2014(7):79, Jul 2014. doi:10.1007/JHEP07(2014)079.
- [8] D. et al. Amaral. Constraints on low-mass, relic dark matter candidates from a surface-operated supercdms single-charge sensitive detector. *Phys. Rev. D*, 102:091101, Nov 2020. URL: <https://link.aps.org/doi/10.1103/PhysRevD.102.091101>, doi:10.1103/PhysRevD.102.091101.
- [9] C. et al. Amole. Dark matter search results from the pico-60 cf₃I bubble chamber. *Phys. Rev. D*, 93:052014, Mar 2016. URL: <https://link.aps.org/doi/10.1103/PhysRevD.93.052014>, doi:10.1103/PhysRevD.93.052014.
- [10] Leila Haegel and. The latest t2k neutrino oscillation results. In *Proceedings of The European Physical Society Conference on High Energy Physics — PoS(EPS-HEP2017)*. Sissa Medialab, oct 2017. URL: <https://doi.org/10.22323/2F1.314.0112>, doi:10.22323/1.314.0112.
- [11] E. et al. Aprile. Search for inelastic scattering of wimp dark matter in xenon1t. *Phys. Rev. D*, 103:063028, Mar 2021. URL: <https://link.aps.org/doi/10.1103/PhysRevD.103.063028>, doi:10.1103/PhysRevD.103.063028.
- [12] ATLAS. Summary Plots for Exotics Heavy Particle Searches and Exotics/SUSY Long-lived Particle Searches. Technical report, CERN, Geneva, 2019. All figures including auxiliary figures are available

- at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2019-023>.
URL: <http://cds.cern.ch/record/2682064>.
- [13] ATLAS. Standard Model Summary Plots Spring 2020. Technical report, CERN, Geneva, 2020. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2020-010>. URL: <https://cds.cern.ch/record/2718937>.
 - [14] ATLAS. SUSY July 2020 Summary Plot Update. Technical report, CERN, Geneva, 2020. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2020-020>. URL: <http://cds.cern.ch/record/2725258>.
 - [15] H. Baer and X. Tata. *Weak scale supersymmetry: From superfields to scattering events*. Cambridge University Press, 5 2006.
 - [16] John C. Baez. Getting to the bottom of noether’s theorem, 2022. [arXiv:2006.14741](https://arxiv.org/abs/2006.14741).
 - [17] R. Ballou. Latest results of the osqar photon regeneration experiment for axion-like particle search, 2014. [arXiv:1410.2566](https://arxiv.org/abs/1410.2566).
 - [18] D Benjamin, P Calafiura, T Childers, K De, W Guan, T Maeno, P Nilsson, V Tsulaia, P Van Gemmeren, T Wenaus, and on behalf of the ATLAS Collaboration. Production experience with the atlas event service. *Journal of Physics: Conference Series*, 898(6):062002, oct 2017. URL: <https://dx.doi.org/10.1088/1742-6596/898/6/062002>, doi:10.1088/1742-6596/898/6/062002.
 - [19] Christian Bierlich, Smita Chakraborty, Nishita Desai, Leif Gellersen, Ilkka Helenius, Philip Ilten, Leif Lönnblad, Stephen Mrenna, Stefan Prestel, Christian T. Preuss, Torbjörn Sjöstrand, Peter Skands, Marius Uthmeim, and Rob Verheyen. A comprehensive guide to the physics and usage of pythia 8.3, 2022. [arXiv:2203.11601](https://arxiv.org/abs/2203.11601).
 - [20] Ahmet Bingul. The ATLAS TRT and its performance at LHC. *J. Phys. Conf. Ser.*, 347:012025, 2012. doi:10.1088/1742-6596/347/1/012025.
 - [21] G.C. Branco, P.M. Ferreira, L. Lavoura, M.N. Rebelo, Marc Sher, and João P. Silva. Theory and phenomenology of two-higgs-doublet models. *Physics Reports*, 516(1-2):1–102, jul 2012. URL: <https://doi.org/10.1016%2Fj.physrep.2012.02.002>, doi:10.1016/j.physrep.2012.02.002.
 - [22] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-kt jet clustering algorithm. *Journal of High Energy Physics*, 2008(04):063, apr 2008. URL: <https://dx.doi.org/10.1088/1126-6708/2008/04/063>, doi:10.1088/1126-6708/2008/04/063.
 - [23] ATLAS Canada. ATLAS CANADA Forward Calorimeter — atlas-canada.ca. <https://www.atlas-canada.ca/fcal2.html>.
 - [24] Jean-Luc Caron. CERN Aerial view. Vue aérienne du CERN. <http://cds.cern.ch/record/841506>, 1986. URL: <http://cds.cern.ch/record/841506>.
 - [25] CERN. Accélérateur linéaire 4 — home.cern. <https://home.cern/fr/science/accelerators/linear-accelerator-4>.
 - [26] CERN. CAST — home.cern. <https://home.cern/fr/science/experiments/cast>.
 - [27] CERN. CMS recherche la durée de vie du boson de Higgs — home.cern. <https://home.cern/fr/news/news/physics/cms-homes-higgs-bosons-lifetime>.
 - [28] CERN. Cryogenics: Low temperatures, high performance — home.cern. <https://www.home.cern/science/engineering/cryogenics-low-temperatures-high-performance>.
 - [29] CERN. Grand collisionneur de hadrons — home.cern. <https://home.cern/fr/science/accelerators/large-hadron-collider>.
 - [30] CERN. How did we discover the Higgs boson? — home.cern. <https://home.cern/science/physics/higgs-boson/how>.

- [31] CERN. Internet prehistory at CERN — home.cern. <https://home.cern/news/opinion/computing/internet-prehistory-cern>.
- [32] CERN. La naissance du web | CERN — home.cern. <https://home.cern/fr/science/computing/birth-web>.
- [33] CERN. La supersymétrie — home.web.cern.ch. <https://home.web.cern.ch/fr/science/physics/supersymmetry>.
- [34] CERN. Le Booster du Synchrotron à protons — home.cern. <https://home.cern/fr/science/accelerators/proton-synchrotron-booster>.
- [35] CERN. Le complexe d'accélérateurs | CERN — home.cern. <https://home.cern/fr/science/accelerators/accelerator-complex>.
- [36] CERN. Le sps - le supersynchrotron à protons — home.cern. <https://home.cern/fr/science/accelerators/super-proton-synchrotron>.
- [37] CERN. LHC the guide FAQ | CERN — home.cern. <https://home.cern/resources/brochure/knowledge-sharing/lhc-facts-and-figures>.
- [38] CERN. LHC upgrades during LS2 — home.cern. <https://home.cern/press/2022/lhc-upgrades-during-ls2>.
- [39] CERN. LuminosityPublicResultsRun2 It; AtlasPublic It; TWiki — twiki.cern.ch. <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2>.
- [40] CERN. Nos États membres — home.cern. <https://home.cern/fr/about/who-we-are/our-governance/member-states>.
- [41] CERN. OSQAR — home.cern. <https://home.cern/fr/science/osqar>.
- [42] CERN. The Proton Synchrotron — home.cern. <https://home.cern/fr/science/accelerators/proton-synchrotron>.
- [43] CERN. The waltz of the LHC magnets has begun — home.cern. <https://home.cern/news/news/accelerators/waltz-lhc-magnets-has-begun>.
- [44] CERN. Interim summary report on the analysis of the 19 september 2008 incident at the lhc. https://edms.cern.ch/ui/file/973073/1/Report_on_080919_incident_at_LHC__2_.pdf, 2008.
- [45] CERN. Cavités radiofréquence. <https://cds.cern.ch/record/1998797>, 2014.
- [46] Serguei Chatrchyan et al. Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC. *Phys. Lett. B*, 716:30–61, 2012. arXiv:1207.7235, doi:10.1016/j.physletb.2012.08.021.
- [47] S. Y. Choi, M. Drees, A. Freitas, and P. M. Zerwas. Testing the majorana nature of gluinos and neutralinos. *Physical Review D*, 78(9), nov 2008. URL: <https://doi.org/10.1103/PhysRevD.78.095007>, doi:10.1103/PhysRevD.78.095007.
- [48] Georgios Choudalakis. On hypothesis testing, trials factor, hypertests and the bumhunter. <https://doi.org/10.48550/arXiv.1101.0390>, 2011. arXiv:1101.0390.
- [49] ATLAS Collaboration. Measurement of the higgs boson mass in the $h \rightarrow zz^* \rightarrow 4\ell$ decay channel using 139 fb^{-1} of $\sqrt{s} = 13 \text{ tev}$ pp collisions recorded by the atlas detector at the lhc, 2022. arXiv:2207.00320.
- [50] The ALICE Collaboration. The alice experiment at the cern lhc. *Journal of Instrumentation*, 3(08):S08002, aug 2008. URL: <https://dx.doi.org/10.1088/1748-0221/3/08/S08002>, doi:10.1088/1748-0221/3/08/S08002.
- [51] The ATLAS Collaboration. Detector & Technology — atlas.cern. <https://atlas.cern/Discover/Detector>.
- [52] The ATLAS Collaboration. The Collaboration — atlas.cern. <https://atlas.cern/Discover/Collaboration>.

- [53] The ATLAS Collaboration. Trigger and Data Acquisition System — atlas.cern. <https://atlas.cern/Discover/Detector/Trigger-DAQ>.
- [54] The ATLAS Collaboration. The atlas experiment at the cern large hadron collider. *Journal of Instrumentation*, 3(08):S08003, aug 2008. URL: <https://dx.doi.org/10.1088/1748-0221/3/08/S08003>, doi:10.1088/1748-0221/3/08/S08003.
- [55] The ATLAS Collaboration. The atlas experiment at the cern large hadron collider. *Journal of Instrumentation*, 3(08):S08003, aug 2008. URL: <https://dx.doi.org/10.1088/1748-0221/3/08/S08003>, doi:10.1088/1748-0221/3/08/S08003.
- [56] The Atlas Collaboration. Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc. *Physics Letters B*, 716(1):1–29, 2012. URL: <https://www.sciencedirect.com/science/article/pii/S037026931200857X>, doi:<https://doi.org/10.1016/j.physletb.2012.08.020>.
- [57] The CMS Collaboration. Collaboration | CMS Experiment — cms.cern. <https://cms.cern/collaboration>.
- [58] The CMS Collaboration. The cms experiment at the cern lhc. *Journal of Instrumentation*, 3(08):S08004, aug 2008. URL: <https://dx.doi.org/10.1088/1748-0221/3/08/S08004>, doi:10.1088/1748-0221/3/08/S08004.
- [59] The CMS Collaboration. A measurement of the higgs boson mass in the diphoton decay channel. *Physics Letters B*, 805:135425, 2020. URL: <https://www.sciencedirect.com/science/article/pii/S037026932030229X>, doi:<https://doi.org/10.1016/j.physletb.2020.135425>.
- [60] The LHCb Collaboration. The lhcb detector at the lhc. *Journal of Instrumentation*, 3(08):S08005, aug 2008. URL: <https://dx.doi.org/10.1088/1748-0221/3/08/S08005>, doi:10.1088/1748-0221/3/08/S08005.
- [61] The T2K Public Website Committee. The T2K Experiment — t2k-experiment.org. <https://t2k-experiment.org/t2k/>.
- [62] Wikimedia Commons. File:standard model of elementary particles.svg — wikimedia commons, the free media repository, 2019. https://commons.wikimedia.org/w/index.php?title=File:Standard_Model_of_Elementary_Particles.svg&oldid=341423479.
- [63] Lucas B.V. de Amorim, George D.C. Cavalcanti, and Rafael M.O. Cruz. The choice of scaling technique matters for classification performance. *Applied Soft Computing*, 133:109924, January 2023. URL: <http://dx.doi.org/10.1016/j.asoc.2022.109924>, doi:10.1016/j.asoc.2022.109924.
- [64] J. de Favereau, , C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi. DELPHES 3: a modular framework for fast simulation of a generic collider experiment. *Journal of High Energy Physics*, 2014(2), feb 2014. URL: <https://doi.org/10.1007/2Fjhep02%282014%29057>, doi:10.1007/jhep02(2014)057.
- [65] J. G. de Swart, G. Bertone, and J. van Dongen. How dark matter came to matter. *Nature Astronomy*, 1(3):0059, Mar 2017. doi:10.1038/s41550-017-0059.
- [66] Arden Dertat. Applied Deep Learning - Part 3: Autoencoders — towardsdatascience.com. <https://towardsdatascience.com/applied-deep-learning-part-3-autoencoders-1c083af4d798>.
- [67] Yu.L. Dokshitzer, G.D. Leder, S. Moretti, and B.R. Webber. Better jet clustering algorithms. *Journal of High Energy Physics*, 1997(08):001, sep 1997. URL: <https://dx.doi.org/10.1088/1126-6708/1997/08/001>, doi:10.1088/1126-6708/1997/08/001.
- [68] Claude Duhr. 2HDM x2013; FeynRules — feynrules.irmp.ucl.ac.be. <https://feynrules.irmp.ucl.ac.be/wiki/2HDM>.

- [69] John R. Ellis. Limits of the Standard Model. Technical report, CERN, Geneva, 2002. URL: <https://cds.cern.ch/record/591472>.
- [70] G. Aad et al. The performance of the jet trigger for the ATLAS detector during 2011 data taking. *The European Physical Journal C*, 76(10), sep 2016. URL: <https://doi.org/10.1140/epjc%2Fs10052-016-4325-0>, doi:10.1140/epjc/s10052-016-4325-0.
- [71] M. Aaboud et al. Jet reconstruction and performance using particle flow with the ATLAS detector. *The European Physical Journal C*, 77(7), jul 2017. URL: <https://doi.org/10.1140/epjc%2Fs10052-017-5031-2>, doi:10.1140/epjc/s10052-017-5031-2.
- [72] M. Aaboud et al. Performance of top-quark and W -boson tagging with ATLAS in run 2 of the LHC. *The European Physical Journal C*, 79(5), apr 2019. URL: <https://doi.org/10.1140/epjc%2Fs10052-019-6847-8>, doi:10.1140/epjc/s10052-019-6847-8.
- [73] Lyndon Evans and Philip Bryant. Lhc machine. *Journal of Instrumentation*, 3(08):S08001, aug 2008. URL: <https://dx.doi.org/10.1088/1748-0221/3/08/S08001>, doi:10.1088/1748-0221/3/08/S08001.
- [74] Stack Exchange. KL divergence between two univariate Gaussians — stats.stackexchange.com. <https://stats.stackexchange.com/questions/7440/kl-divergence-between-two-univariate-gaussians>.
- [75] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>.
- [76] Nicolas Greiner, Kyoungchul Kong, Jong-Chul Park, Seong Chan Park, and Jan-Christopher Winter. Model-independent production of a top-philic resonance at the LHC. *Journal of High Energy Physics*, 2015(4), apr 2015. URL: [https://doi.org/10.1007/jhep04\(2015\)029](https://doi.org/10.1007/jhep04(2015)029), doi:10.1007/jhep04(2015)029.
- [77] David Griffiths. *Introduction to elementary particles*. Wiley-VCH Verlag, Weinheim, Germany, 2nd edition, August 2008.
- [78] Jiuxiang Gu, Zhenhua Wang, Jason Kuen, Lianyang Ma, Amir Shahroudy, Bing Shuai, Ting Liu, Xingxing Wang, Li Wang, Gang Wang, Jianfei Cai, and Tsuhan Chen. Recent advances in convolutional neural networks, 2017. [arXiv:1512.07108](https://arxiv.org/abs/1512.07108).
- [79] Ken Hoffman. Machine Learning: How to Prevent Overfitting — medium.com. <https://medium.com/swlh/machine-learning-how-to-prevent-overfitting-fdf759cc00a9>.
- [80] Ryan Holbrook. Overfitting and Underfitting — kaggle.com. <https://www.kaggle.com/code/ryanholbrook/overfitting-and-underfitting/tutorial>.
- [81] <https://www.facebook.com/bbcnews>. Google AI defeats human Go champion — bbc.com. <https://www.bbc.com/news/technology-40042581>.
- [82] Renu Khandelwal. How to solve randomness in an artificial neural network? — towardsdatascience.com. <https://towardsdatascience.com/how-to-solve-randomness-in-an-artificial-neural-network-3>
- [83] Avinash Khare. The november J/ψ revolution: Twenty-five years later, 1999. [arXiv:hep-ph/9910468](https://arxiv.org/abs/hep-ph/9910468).
- [84] R. Khatiwada. Axion dark matter experiment: Detailed design and operations. *Review of Scientific Instruments*, 92(12):124502, dec 2021. URL: <https://doi.org/10.1063/1.50037857>, doi:10.1063/1.50037857.
- [85] J. et al. Kim. Top-Philic-Zprime x2013; FeynRules — feynrules.irmp.ucl.ac.be. <https://feynrules.irmp.ucl.ac.be/wiki/Top-Philic-Zprime>.
- [86] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017. [arXiv:1412.6980](https://arxiv.org/abs/1412.6980).

- [87] Samuel Klein and Tobias Golling. Decorrelation with conditional normalizing flows, 2022. [arXiv:2211.02486](https://arxiv.org/abs/2211.02486).
- [88] Roman Kogler et al. Jet Substructure at the Large Hadron Collider: Experimental Review. *Rev. Mod. Phys.*, 91(4):045003, 2019. [arXiv:1803.06991](https://arxiv.org/abs/1803.06991), doi:10.1103/RevModPhys.91.045003.
- [89] Sofia Kostoglou. The CAST experiment at CERN. Master’s thesis, Natl. Tech. U., Athens, 6 2016.
- [90] Bharath Krishnamurthy. The Evolution of Chess AI — builtin.com. <https://builtin.com/artificial-intelligence/chess-ai>.
- [91] David Krohn, Jesse Thaler, and Lian-Tao Wang. Jet Trimming. *JHEP*, 02:084, 2010. [arXiv:0912.1342](https://arxiv.org/abs/0912.1342), doi:10.1007/JHEP02(2010)084.
- [92] Fabian Kuger. Signal formation processes in micromegas detectors and quality control for large size detector construction for the atlas new small wheel. 08 2017. URL: https://www.researchgate.net/publication/318981598_Signal_Formation_Processes_in_Micromegas_Detectors_and_Quality_Control_for_large_size_Detector_Construction_for_the_ATLAS_New_Small_Wheel.
- [93] Brookhaven National Laboratory. Brookhaven National Laboratory x2014; a passion for discovery — bnl.gov. <https://www.bnl.gov/world/>.
- [94] J-P. Lenoir. Probabilités et statistiques. https://www.imo.universite-paris-saclay.fr/~pierre.pansu/web_ifips/Tests.pdf.
- [95] Huafeng Liu, Liping Jing, Jingxuan Wen, Pengyu Xu, Jiaqi Wang, Jian Yu, and Michael K. Ng. Interpretable deep generative recommendation models. *Journal of Machine Learning Research*, 22(202):1–54, 2021. URL: <http://jmlr.org/papers/v22/20-1098.html>.
- [96] David London. GPP : David London — lps.umontreal.ca. http://www.lps.umontreal.ca/~london/SMbeamer_website.pdf, 2020.
- [97] Scott Mayer McKinney, Marcin Sieniek, and Varun Godbole. International evaluation of an ai system for breast cancer screening. *Nature*, 577(7788):89–94, Jan 2020. doi:10.1038/s41586-019-1799-6.
- [98] Issa Mohamadi. A mathematical proof for the existence of a possible source for dark energy, 2017. [arXiv:1704.04430](https://arxiv.org/abs/1704.04430).
- [99] Rabindra N. Mohapatra. Gauge Theories of the Strong, Weak, and Electromagnetic Interactions. *Physics Today*, 67(6):50–50, 06 2014. [arXiv:https://pubs.aip.org/physicstoday/article-pdf/67/6/50/10099524/50_1_online.pdf](https://pubs.aip.org/physicstoday/article-pdf/67/6/50/10099524/50_1_online.pdf), doi:10.1063/PT.3.2421.
- [100] NASA. Dark Energy, Dark Matter | Science Mission Directorate — science.nasa.gov. <https://science.nasa.gov/astrophysics/focus-areas/what-is-dark-energy>.
- [101] Joao Pequeno. Computer Generated image of the ATLAS calorimeter. <https://cds.cern.ch/record/1095927>, 2008.
- [102] Joao Pequeno. Computer generated image of the ATLAS inner detector. <https://cds.cern.ch/record/1095926>, 2008.
- [103] Joao Pequeno. Computer generated image of the ATLAS Muons subsystem. <http://cds.cern.ch/record/1095929>, 2008.
- [104] K. B. Petersen and M. S. Pedersen. The matrix cookbook, October 2008. Version 20081110. URL: <http://www2.imm.dtu.dk/pubdb/p.php?3274>.
- [105] T. N. Pham. Ckm matrix elements, 2011. [arXiv:1110.6050](https://arxiv.org/abs/1110.6050).
- [106] Martin Popel, Marketa Tomkova, Jakub Tomek, Łukasz Kaiser, Jakob Uszkoreit, Ondřej Bojar, and Zdeněk Žabokrtský. Transforming machine translation: a deep learning system reaches news translation quality comparable to human professionals. *Nature Communications*, 11(1):4381, Sep 2020. doi:10.1038/s41467-020-18073-9.

- [107] Paul Prasse, Christoph Sawade, Niels L. Wehr, and Tobias Scheffer. Learning to identify concise regular expressions that describe email campaigns. *Journal of Machine Learning Research*, 16(112):3687–3720, 2015. URL: <http://jmlr.org/papers/v16/prasse15a.html>.
- [108] H. Primakoff. Photo-production of neutral mesons in nuclear electric fields and the mean life of the neutral meson. *Phys. Rev.*, 81:899–899, Mar 1951. URL: <https://link.aps.org/doi/10.1103/PhysRev.81.899>, doi:10.1103/PhysRev.81.899.
- [109] Jonathan Pumplin, Daniel Robert Stump, Joey Huston, Hung-Liang Lai, Pavel Nadolsky, and Wu-Ki Tung. New generation of parton distributions with uncertainties from global QCD analysis. *Journal of High Energy Physics*, 2002(07):012–012, Jul 2002. URL: <https://doi.org/10.1088/1126-6708/2002/07/012>, doi:10.1088/1126-6708/2002/07/012.
- [110] pypi.org. pyBumpHunter — pypi.org. <https://pypi.org/project/pyBumpHunter/>.
- [111] Gavin P. Salam. Towards jetography. *The European Physical Journal C*, 67(3):637–686, Jun 2010. doi:10.1140/epjc/s10052-010-1314-6.
- [112] Irhum Shafkat. Intuitively Understanding Variational Autoencoders — towardsdatascience.com. <https://towardsdatascience.com/intuitively-understanding-variational-autoencoders-1bfe67eb5daf>.
- [113] Aditya Sharma. Variational Autoencoder in TensorFlow — learnopencv.com. <https://learnopencv.com/variational-autoencoder-in-tensorflow/>.
- [114] Alex Sherstinsky. Fundamentals of recurrent neural network (rnn) and long short-term memory (lstm) network. *Physica D: Nonlinear Phenomena*, 404:132306, March 2020. URL: <http://dx.doi.org/10.1016/j.physd.2019.132306>, doi:10.1016/j.physd.2019.132306.
- [115] SNO. The SNO+ Experiment — snoplus.phy.queensu.ca. <https://snoplus.phy.queensu.ca/>. [Accessed 28-02-2024].
- [116] Stanford. The Stanford Linear Accelerator Center — slac.stanford.edu. <https://www.slac.stanford.edu/gen/grad/GradHandbook/slac.html>.
- [117] SuperNova Early Warning System. Neutrinos — snews.bnl.gov. <https://snews.bnl.gov/popsci/neutrino.html>.
- [118] Mark Thomson. *Modern Particle Physics*. Cambridge University Press, 2013. doi:10.1017/CB09781139525367.
- [119] Super User. sezione d’urto tt(bar) — web.infn.it. https://web.infn.it/cms_bologna/index.php/attivita-bologna/analisi/sezione-d-urto-tt-bar.
- [120] D. R. Wardrope. Boosted Heavy Particles and Jet Substructure — cds.cern.ch. <http://cds.cern.ch/record/1604479/?ln=fr>.
- [121] Steven Weinberg. A New Light Boson? *Phys. Rev. Lett.*, 40:223–226, 1978. doi:10.1103/PhysRevLett.40.223.
- [122] Maciej Wielgosz, Andrzej Skoczeń, and Matej Mertik. Recurrent neural networks for anomaly detection in the post-mortem time series of lhc superconducting magnets, 2017. arXiv:1702.00833.
- [123] Wikipedia. Convention de sommation d’Einstein — fr.wikipedia.org. https://fr.wikipedia.org/wiki/Convention_de_sommation_d'Einstein.
- [124] Wikipedia. Feynman diagram — en.wikipedia.org. https://en.wikipedia.org/wiki/Feynman_diagram.
- [125] Wikipedia. Kullback–Leibler divergence — en.wikipedia.org. https://en.wikipedia.org/wiki/Kullback%E2%80%93Leibler_divergence.
- [126] Wikipedia. Rayonnement continu de freinage — fr.wikipedia.org. https://fr.wikipedia.org/wiki/Rayonnement_continu_de_freinage.

- [127] Wikipedia. Théorème de Noether (physique) — fr.wikipedia.org. [https://fr.wikipedia.org/wiki/Théorème_de_Noether_\(physique\)](https://fr.wikipedia.org/wiki/Théorème_de_Noether_(physique)).
- [128] Wikipedia. Timeline of particle discoveries — en.wikipedia.org. https://en.wikipedia.org/wiki/Timeline_of_particle_discoveries.
- [129] Wikipedia. Vera Rubin — fr.wikipedia.org. https://fr.wikipedia.org/wiki/Vera_Rubin.
- [130] R. L. *et al.* Workman. Review of Particle Physics. *PTEP*, 2022:083C01, 2022. doi:10.1093/ptep/ptac097.
- [131] S Xella. Physics objects reconstruction in the ATLAS experiment. Technical report, CERN, Geneva, 2013. URL: <https://cds.cern.ch/record/1519113>.
- [132] Chao Xu, Hong Tao, Jing Zhang, Dewen Hu, and Chenping Hou. Label distribution changing learning with sample space expanding. *Journal of Machine Learning Research*, 24(36):1–48, 2023. URL: <http://jmlr.org/papers/v24/22-0210.html>.
- [133] Jiacheng Xu and Greg Durrett. Spherical latent spaces for stable variational autoencoders, 2018. arXiv:1808.10805.
- [134] Yanqing Zhang, Xuan Bi, Niansheng Tang, and Annie Qu. Dynamic tensor recommender systems. *Journal of Machine Learning Research*, 22(65):1–35, 2021. URL: <http://jmlr.org/papers/v22/19-792.html>.