Sorbonne Université

École doctorale des Sciences de la Terre et de l'Environnement et Physique de l'Univers de Paris (ED 560)

Laboratoire de Physique Nucléaire et des Hautes Énergies, Paris – Équipe ATLAS

Étalonnage des jets et mesures précises de sections efficaces de production de jets avec les données de l'expérience ATLAS

Par Louis Ginabat

Thèse de doctorat de physique des particules expérimentale

Dirigée par Mélissa Ridel
Co-encadrée par Bogdan Malaescu

Présentée et soutenue publiquement le 25 septembre 2023

Devant un jury composé de :

Sergey	Barsuk,	rapporteur de thèse
Suzanne	Gascon-Shotkin,	rapportrice de thèse
Matt	Leblanc,	membre
Bogdan	Malaescu,	co-encadrant de thèse
Mélissa	Ridel,	directrice de thèse
Luca	Scotto Lavina,	président du jury

CERN-THESIS-2023-391
25/09/2023

Except where otherwise noted, this work is licensed under
http://creativecommons.org/licenses/by-nc-nd/3.0/

Résumé

Le modèle standard (MS) de la physique des particules décrit les particules élémentaires qui constituent la matière, ainsi que trois des quatre forces fondamentales. Il n'est presque pas remis en cause par les données expérimentales, ce qui n'indique pas de brèche évidente où la Nouvelle Physique pourrait se situer. Cette thèse se concentre sur une stratégie pour tester le MS, la comparaison de ses prédictions avec des mesures de précision, et sur un aspect de la section Chromo-Dynamique Quantique (QCD) du MS : les *jets*. Le travail présenté ici utilise les données collectées par l'expérience ATLAS, située au CERN, à partir de collisions pp à 13 TeV produites par le Grand Collisionneur de Hadrons (LHC) pendant la période de prise de données du Run 2.

Les *jets* sont l'une des principales observables au LHC en raison de leur section efficace de production élevée, présents dans la plupart des analyses soit comme signal, soit comme bruit de fond. Il s'agit de gerbes collimatées de hadrons issus de l'hadronisation d'un quark ou d'un gluon. Les hadrons produisent des traces dans les trajectographes d'ATLAS et des dépôts d'énergie dans ses calorimètres, qui doivent ensuite être regroupés pour former des « *jets* reconstruits ». Les *jets* de particules et les *jets* reconstruits présentent des différences dans leurs observables, comme leur quantité de mouvement transverse p_T, en raison d'effets de détecteur qui doivent donc être corrigés pour permettre des comparaisons entre la théorie et l'expérience. Deux stratégies de correction sont discutées dans cette thèse : l'étalonnage et la déconvolution, pour corriger les effets de détecteurs respectivement dans l'échelle et la résolution des observables.

J'ai travaillé sur l'inter-étalonnage en pseudo-rapidité η, une des étapes de la chaîne de corrections de l'échelle d'énergie des jets (JES). Cet étalonnage *in situ* rend la réponse du détecteur homogène sur toute la plage en η du détecteur, en exploitant la conservation de p_T dans les systèmes *dijets*. J'ai produit l'étalonnage pour la période de prise de données 2018 pour la première fois, en effectuant toutes sortes de vérifications pour m'assurer de sa qualité. J'ai également travaillé sur l'implémentation « absolue » de cette méthode, qui vise à déconvoluer les effets de physique des effets de détecteur, afin de mieux les comprendre séparément. Cela a permis de réduire de 50 % l'incertitude de modélisation qui était dominante dans la tranche de 25 à 40 GeV. Cela m'a également permis d'effectuer une vérification importante au niveau des particules, à savoir que les incertitudes conçues pour couvrir les effets physiques ne sont pas sous-estimées.

En outre, j'ai contribué à deux mesures de sections efficaces de production de *jets*, des mesures de précision doublement différentielles qui comptent les *jets* dans des tranches de p_T et de rapidité $|y|$. La section efficace inclusive compte tous les *jets* produits au-delà d'un certain seuil en p_T, tandis que la section efficace de *jets* dominants n'en compte qu'un par événement : celui qui a le p_T le plus élevé. Les effets de la résolution de l'énergie des *jets* (JER) sont corrigés par une méthode de déconvolution itérative stabilisée dynamiquement (IDS). Les incertitudes systématiques sur la JES et la JER ont été propagées à travers la déconvolution, utilisant la méthode *Bootstrap* pour évaluer les incertitudes statistiques ainsi que les corrélations. J'ai amélioré l'évaluation du biais lié à la sensibilité de la méthode de déconvolution aux différences de forme entre les données et le MC, et j'ai évalué l'impact de la non-gaussianité de la JER avec une nouvelle incertitude. Les résultats de la section efficace inclusive de production de *jets* sont montrés et comparés aux prédictions théoriques, manifestant un désaccord important qui est traité par de multiples stratégies, comme l'utilisation d'incertitudes statistiques non lissées de l'inter-étalonnage en η.

Abstract

The Standard Model (SM) of particle physics describes the elementary particles that constitute matter, as well as three of the four fundamental forces. It remains almost unchallenged by experimental data, thus not indicating any clear breach where New Physics could be located. This thesis focuses on one strategy to test the SM, the comparison of its predictions to precision measurements, and on one aspect of the Quantum Chromo-Dynamics (QCD) section of the SM: jets. The work presented here uses data collected by the ATLAS experiment, located at CERN, from 13 TeV pp collisions produced by the Large Hadron Collider (LHC) during its Run 2 data-taking period.

Jets are one of the main observables at the LHC due to their high production cross-section. They are collimated sprays of hadrons, stemming from the hadronization of a quark or gluon. The hadrons produce tracks in the tracking systems of ATLAS and energy deposits in its calorimeters, which then need to be clustered together to form "reconstructed jets". Particle jets and reconstructed jets show differences in their observables, like their transverse momentum p_T, due to detector effects which therefore need to be corrected for, to enable comparisons between theory and experiment. Two such strategies are discussed in this thesis: calibration and unfolding, to correct for the detector effects respectively in the scale and the resolution.

I have worked on the η-intercalibration, one of the steps of the corrections chain of the Jet Energy Scale (JES). This *in situ* calibration makes the detector response homogeneous across the whole pseudo-rapidity η range of the detector, by exploiting the transverse momentum p_T conservation in dijet systems. I produced the calibration for the 2018 data-taking period for the first time, making all sorts of cross-checks to ensure its quality. I also worked on the "absolute" implementation of this method, that aims at disentangling the physics effects from the detector effects, to better understand them separately. This led to the reduction by 50 % of the modeling uncertainty that was dominant in the 25 to 40 GeV bin. It also allowed me to make the important verification at particle level that the uncertainties designed to cover for physics effects are not underestimated.

Furthermore, I contributed to the inclusive and leading jet production cross-section measurements, two double-differential precision measurements that count jets in bins of p_T and rapidity $|y|$. While the inclusive measurement counts every single jet produced above a certain p_T threshold, the leading jet measurement only counts one per event: the one with the highest p_T. The Jet Energy Resolution (JER) effects are corrected for by the Iterative Dynamically Stabilized unfolding (IDS) method. Systematic uncertainties on the JES and JER have been propagated through the unfolding, using the Bootstrap method to evaluate the statistical uncertainties as well as correlations. I have improved the evaluation of the bias related to the sensitivity of the unfolding method to the shape differences between data and MC, and I evaluated the impact of the non-Gaussianity of the JER in a new uncertainty. Results of the inclusive jet production cross-section are shown and compared to theoretical predictions in this thesis, exhibiting an important disagreement that is being addressed by multiple strategies, like the use of non-smoothed statistical uncertainties of the η-intercalibration.

Synopsis

Le Modèle Standard (MS) de la physique des particules est l'édifice théorique qui décrit les constituants élémentaires de la matières et leurs interactions à travers trois forces fondamentales : la force électromagnétique, la force nucléaire forte et la force nucléaire faible. Construit en parallèle des progrès expérimentaux fulgurants du siècle passé, s'appuyant sur le cadre théorique fourni par la théorie quantique des champs, il décrit surprenamment bien les phénomènes naturels aux plus petites échelles de longueur et de durée, ou de manière équivalente aux plus grandes échelles d'énergie. En dépit de son succès, le MS est incomplet. Entre autres difficultés, il ne décrit pas la quatrième force fondamentale qu'est la gravité, n'explique pas la matière noire et l'énergie sombre, demeure silencieux quant à l'origine de l'asymétrie matière-antimatière, ne fournit pas d'explication pour ses 26 paramètres libres, et laisse un certain nombre de problèmes de naturalité ouverts. L'incomplétion théorique du MS invite à la recherche de physique allant au-delà, tant théoriquement qu'expérimentalement. Néanmoins, l'absence de remise en cause expérimentale laisse ouvert un champ des possibles immense, limité seulement par l'imagination, et les théories tentant de dépasser le MS sont foisonnantes. L'importance de la recherche expérimentale en physique des particules est donc cruciale.

À cette fin, le CERN, l'organisation européenne pour la recherche nucléaire, situé à la frontière franco-suisse, a construit depuis 1954 une infrastructure d'accélérateurs de particules, afin de produire des collisions de particules à des énergies toujours plus élevées pour sonder la matière à des distances toujours plus courtes. Le grand collisionneur de hadrons (LHC) en est un : cet instrument, mesurant 27 km de circonférence, accélère des protons ou des ions lourds à des vitesses proches de la vitesse de la lumière, en s'appuyant sur les accélérateurs de particules déjà construits auparavant. En 2010, il a battu le record du monde des collisions artificielles les plus énergétiques, atteignant 7 TeV dans le centre de masse, puis 8 TeV en 2012, 13 TeV de 2015 à 2016, une période appelée le Run 2, et enfin $13,6$ TeV depuis 2022. Les collisions sont produites dans des détecteurs de particules, l'un d'entre eux étant ATLAS, qui est l'expérience à laquelle je contribue. Il s'agit d'un détecteur généraliste, pouvant mesurer les propriétés de presque toutes les particules du MS émises lors des collisions qui ont un temps de vie suffisant pour atteindre le détecteur sans se désintégrer : électrons, photons, muons, hadrons (particules composées de quarks et gluons), à l'exception des neutrinos. Cette thèse se concentre sur une stratégie pour tester le MS, la comparaison de ses prédictions avec des mesures de précisions réalisées à partir des collisions proton-proton enregistrées durant le Run 2, et sur un aspect du MS: les jets.

Un quark ou un gluon émis suite à une collision n'atteint pas le détecteur : il va s'hadroniser afin de ne laisser que des particules composites possédant une charge de couleur nulle, appelées les hadrons. Ces particules ont en commun un même boost de Lorentz et forment donc des gerbes hadroniques coniques appelées des jets, qui atteignent le détecteur. Les hadrons laissent des traces dans les trajectographes et des dépôts d'énergie dans les calorimètres électromagnétiques et hadroniques, qu'il faut ensuite regrouper afin de reconstruire les propriétés du quark ou gluon initial, ce que l'on appelle un « jet reconstruit » par opposition au « jet vrai » qui est constitué de particules. Il existe plusieurs façons de grouper les particules ou les dépôts énergétiques ; en l'occurrence, cette thèse se focalise sur les jets de petit rayon où les dépôts énergétiques sont rassemblés avec l'algorithme PFlow, et où les particules vraies ou les groupes de dépôts d'énergie sont réunis en jets avec l'algorithme anti-k_t de rayon $R = 0.4$. Les jets reconstruits et les jets vrais possèdent des différences dans leurs observables, comme leur quantité de mouvement transverse p_T, du fait des effets induits par le détecteur, et il est donc nécessaire de corriger ces effets systématiques afin de comparer les résultats expérimentaux aux prédictions théoriques. Deux stratégies de corrections sont discutées dans cette thèse : l'étalonnage et la déconvolution, pour corriger respectivement l'échelle et la résolution des observables, mis en place respectivement dans l'inter-étalonnage des jets en pseudo-rapidité et dans les mesures de sections efficaces de production de jets.

Inter-étalonnage des jets en η

L'étalonnage consiste en une estimation de l'écart systématique entre la valeur vraie de l'observable et la valeur observée, suivie d'une correction de cet écart, et d'une quantification de l'incertitude sur l'erreur résiduelle après

rectification. Dans le cas de l'échelle d'énergie des jets (JES) pour les jets de petit rayon, l'étalonnage est réalisé dans ATLAS en plusieurs étapes. En premier lieu, la dépendance au bruit d'empilement est réduite ; il s'agit du bruit de fond lié aux collisions de protons qui ne sont pas la collision d'intérêt et qui se sont produites en même temps ou peu de temps auparavant ou après. Puis vient une étape de correction des effets de détecteur majeurs en s'appuyant sur des jeux de données simulés par des techniques de Monte Carlo (MC). Enfin, la JES est étalonnée avec des méthodes *in situ*, c'est-à-dire qui s'appuient sur les données expérimentales. Une de ces méthodes est l'inter-étalonnage en η.

L'inter-étalonnage en η a pour but d'homogénéiser la réponse du détecteur entre les différentes régions en pseudo-rapidité η (latitude par rapport au point d'interaction, vis-à-vis de l'axe du faisceau de protons). Une région du détecteur est sondée, et étalonnée relativement à la réponse dans la région centrale du détecteur, qui est la région de référence. Pour ce faire, une égalité physique est exploitée : la conservation de la quantité de mouvement transverse, qui impose que pour des événements contenant exclusivement deux jets dont la quantité de mouvement transverse p_T est parfaitement bien mesurée (aucun effet de détecteur), l'égalité suivante est respectée : $p_T^{\text{ref}} = p_T^{\text{sonde}}$. En réalité, à cause des effets de résolution du détecteur (du calorimètre principalement), cette égalité n'est attendue qu'en moyenne : $\left\langle \frac{p_T^{\text{ref}}}{p_T^{\text{sonde}}} \right\rangle = 1$. S'il existe un effet de détecteur qui doit être corrigé lors de l'étalonnage, alors cette quantité n'est plus égale à 1. On l'appelle le coefficient d'inter-étalonnage $c = \left\langle \frac{p_T^{\text{ref}}}{p_T^{\text{sonde}}} \right\rangle$. Néanmoins, ce coefficient seul ne peut pas être utilisé comme coefficient de correction, car il ne contient pas seulement des effets de détecteur (qui doivent être corrigés), mais également des effets de physique ; s'il était utilisé tel quel, il altèrerait la physique et induirait donc un biais. En effet, les événements contiennent généralement plus que deux jets, il existe toujours une radiation additionnelle, souvent sous la forme d'un troisième jet ou de radiations en dehors du cône du jet (OOC), qui rend l'équilibre entre les deux jets dominants imparfait. Afin que l'étalonnage ne corrige que les effets de détecteur et ne contienne pas d'effets de physique, un rapport est pris entre les coefficients d'inter-étalonnage calculés dans les données expérimentales et dans le MC :

$$\mathcal{C} = \frac{c_{\text{data}}}{c_{\text{MC}}} \tag{1}$$

De cette façon, les effets de physique, simulés et contenus dans c_{MC}, annulent ceux contenus dans c_{data}, de sorte que l'étalonnage $\mathcal{C}$ ne corrige bien que les effets de détecteur attendus.

L'étalonnage utilisé dans ATLAS est calculé avec le MC au niveau reconstruit (reco) au dénominateur, ce qui est appelé « l'implémentation relative »:

$$\mathcal{C}^{\text{relatif}} = \frac{c_{\text{data}}}{c_{\text{MC reco}}} \tag{2}$$

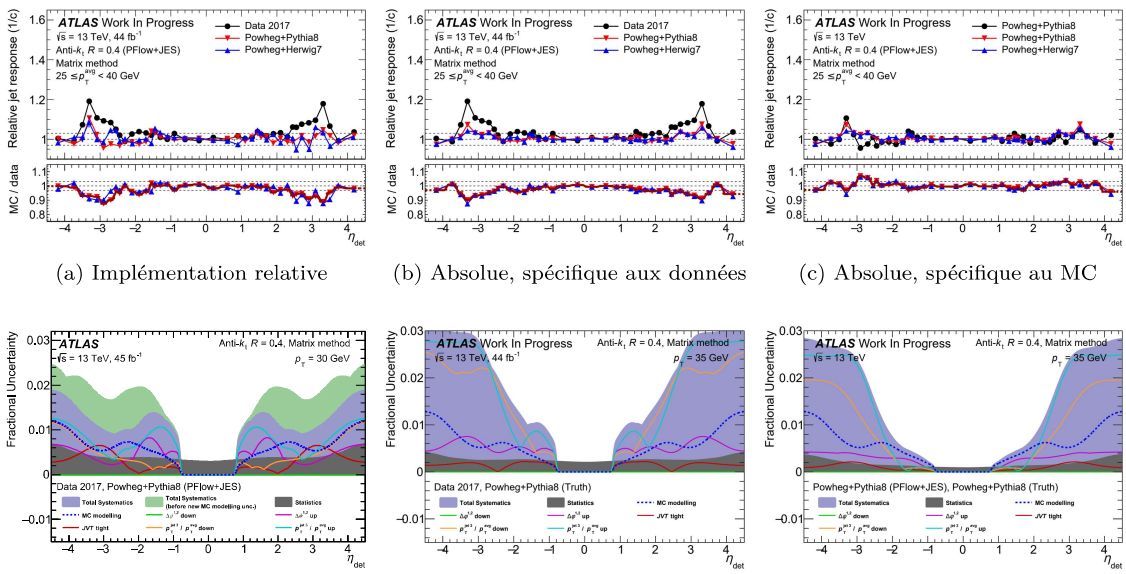
Mais comme $c_{\text{MC reco}}$ convolue à la fois des effets de physique et de détecteur (comme c_{data}), il est malaisé de déterminer si une variation dans le coefficient d'inter-étalonnage provient de l'un ou de l'autre. L'implémentation « absolue » vise à résoudre cet écueil en déconvoluant les deux effets, afin de pouvoir les étudier indépendamment. La solution est d'utiliser non pas un étalonnage mais deux, un pour les données et un pour le MC, où chacun est défini en utilisant la quantité au niveau reco au numérateur et le MC au niveau vrai au dénominateur :

$$\begin{cases} \mathcal{C}^{\text{absolu}}_{\text{spécifique aux données}} = \dfrac{c_{\text{data}}}{c_{\text{MC vrai}}} \\ \mathcal{C}^{\text{absolu}}_{\text{spécifique au MC}} = \dfrac{c_{\text{MC reco}}}{c_{\text{MC vrai}}} \end{cases} \tag{3}$$

L'intérêt de l'implémentation absolue réside dans l'usage du MC au niveau vrai, qui ne contient que des effets de physique. Ainsi, les effets de détecteur et de physique sont déconvolués : les effets de physique sont dans $c_{\text{MC vrai}}$, les effets de détecteur sont dans $\mathcal{C}^{\text{absolu}}_{\text{spécifique aux données}}$ et les effets de détecteur simulés sont dans $\mathcal{C}^{\text{absolu}}_{\text{spécifique au MC}}$. Cela permet d'étudier les différents effets indépendamment, pour mieux les comprendre.

L'étalonnage obtenu avec les deux approches, l'implémentation relative et l'implémentation absolue, est montré dans la région $25 < \mathrm{p_T^{avg}} < 40$ GeV dans la Fig. 1, ainsi que les incertitudes associées. L'objectif est atteint, les effets de physique et de détecteur sont déconvolués, et deux étalonnages sont obtenus pour l'implémentation absolue, un pour les données et un pour le MC. Notez que les incertitudes pour l'approche absolue peuvent sembler plus grandes, mais ces figures ne montrent pas les importantes corrélations induites par la physique ; il est attendu que, lorsqu'appliquée à une analyse de physique en tenant compte des corrélations, l'implémentation absolue ait des incertitudes similaires à celles de l'implémentation relative.

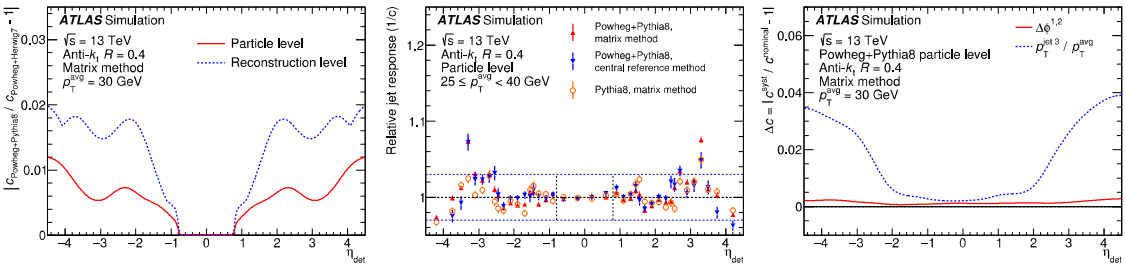
L'implémentation absolue a déjà permis des progrès grâce à la déconvolution des effets de physique et de détecteur, notamment dans la compréhension des effets qui contribuent aux incertitudes. Il y a plusieurs sources d'incertitudes pour l'inter-étalonnage en η : les incertitudes statistiques, liées au nombre d'événements par tranche



(a) Implémentation relative (b) Absolue, spécifique aux données (c) Absolue, spécifique au MC

(d) Implémentation relative (e) Absolue, spécifique aux données (f) Absolue, spécifique au MC

Figure 1: Comparaison des implémentations relative et absolue de l'inter-étalonnage en η. **(a-c)** montrent les coefficients d'inter-étalonnage dans les parties supérieures et les étalonnages dans les parties inférieures, où les points noirs sont les quantités reco (données ou Powheg+Pythia8), les triangles rouges sont les générateurs MC nominaux (Powheg+Pythia8), les triangles bleus sont les générateurs MC alternatifs (Powheg+Herwig7), et la courbe rouge est l'étalonnage lissé. **(d-f)** montrent les incertitudes relatives associées, où les bandes grises sont les incertitudes statistiques, la bande violette est la somme en quadrature des incertitudes systématiques symétrisées. Notez que pour l'implémentation absolue, les incertitudes spécifiques aux données et au MC sont fortement corrélées physiquement. L'incertitude de modélisation des MC (courbe en pointillés bleus) est obtenue en comparant la calibration obtenue avec Powheg+Pythia8 et Powheg+Herwig7. La bande verte dans **(d)** montre les incertitudes systématiques totales en utilisant le niveau reco pour calculer l'incertitude de modélisation MC, tandis que la bande violette utilise le niveau vrai pour cette incertitude : j'ai choisi la seconde option parce qu'elle permet une réduction de 50 % de l'incertitude de modélisation MC, qui était dominante en deçà de 40 GeV, tout en évitant un possible double-comptage des effets de détecteur avec les incertitudes de saveur.



(a) $p_T = 30$ GeV (b) $25 \leq \mathrm{p_T^{avg}} < 40$ GeV (c) $p_T^{\mathrm{avg}} = 30$ GeV

Figure 2: **(a)** Incertitude de modélisation MC obtenue soit au niveau vrai (courbe rouge) soit au niveau reco (courbe bleue) pour Powheg+Pythia8 (pour le jeu de données MC16d, qui correspond aux conditions de la période 2017). Le niveau vrai est bien plus petit dans la région $p_T < 40$ GeV et évite tout possible double-comptage des effets de détecteur avec les incertitudes de saveur. L'incertitude de modélisation MC au niveau vrai est donc adoptée. **(b)** Coefficients d'inter-étalonnage obtenus au niveau vrai pour Powheg+Pythia8 (pour le jeu de données MC16d). Une structure non triviale est observée dans la région avant. **(c)** Incertitudes systématiques obtenues au niveau vrai seulement avec les coefficients d'inter-étalonnage : $\Delta c = \left|\frac{c^{\mathrm{alt}}}{c^{\mathrm{nom}}} - 1\right|$, utilisant le même découpage que l'étalonnage, les mêmes paramètres de lissage, et symétrisés de la même façon. **(b,c)** Les effets de physique observés dans les coefficients d'inter-étalonnage au niveau vrai sont du même ordre de grandeur que les incertitudes systématiques de sélection conçues pour couvrir ces effets de physique, prouvant ainsi que les incertitudes systématiques ne sont pas sous-estimées.

de l'espace des phases, l'incertitude de modélisation MC et les incertitudes liées à la variation des critères de sélection des événements.

L'échantillon d'événements est simulé avec le générateur MC Powheg+Pythia8, qui utilise un modèle de douche partonique dipolaire à hiérarchie en p_T, et le modèle d'hadronisation de chaînes de Lund. Mais d'autres modèles existent, et donc une incertitude est évaluée pour prendre en compte le fait qu'il n'y a pas vraiment de raison de faire confiance à un modèle plutôt qu'à un autre. Pour ce faire, l'étalonnage est obtenu avec un autre générateur MC, Powheg+Herwig7, qui utilise des modèles de douche partonique et d'hadronisation différents (respectivement une douche à hiérarchie angulaire et une hadronisation par groupement dans l'espace des phases). La différence entre l'étalonnage obtenu avec les deux MC différents fournit l'incertitude de modélisation MC, montrée en pointillés bleus dans la Fig. 1, ainsi que dans la Fig. 2a. L'amélioration récente consiste à calculer l'incertitude de modélisation MC non pas avec le MC au niveau reco mais au niveau vrai. En effet, la différence dans les effets de détecteurs induite par le changement de générateur MC est déjà prise en compte dans une autre incertitude de la JES, l'incertitude de saveurs, ce qui fait qu'elle est comptée deux fois si on utilise le MC au niveau reco. L'incertitude déterminée avec le MC au niveau vrai n'inclut donc que la différence d'effets de physique provenant de la différence de MC. Ce changement a permis la division par deux de l'incertitude de modélisation MC dans la région $25 < \mathrm{p_T^{avg}} < 40$ GeV, comme montré dans la Fig. 2a, qui était l'incertitude dominante dans cette région.

Afin de mitiger les effets de physique, les événements choisis doivent être de « bons » événements dijet, c'est-à-dire qu'ils ne doivent pas être trop pollués par l'impact des radiations additionnelles telles que le troisième jet, les radiations en dehors du cône du jet, ou encore le bruit d'empilement. Les événements qui participent à l'élaboration de l'inter-étalonnage en η sont donc choisis précautionneusement pour minimiser l'impact des effets de physique : ils doivent passer certains critères de sélection. Néanmoins, même les critères les plus stricts ne permettraient pas d'avoir un échantillon d'événements purement dijet, et il faut également conserver une certaine puissance statistique en ne supprimant pas trop d'événements. Des effets de physique subsistent donc (et c'est pour cela qu'il faut faire le rapport des coefficients d'intercalibration des données et du MC), et des incertitudes sont conçues pour couvrir leurs effets résiduels potentiels. Ces incertitudes sont obtenues en variant les critères de sélection des événements, ce qui permet d'obtenir une nouvelle calibration qui est alors comparée à la calibration nominale. Ces incertitudes « de sélection » sont montrées dans la Fig. 1. Le progrès apporté par l'implémentation absolue a consisté en la vérification que les effets de physique, qui sont mesurables dans le coefficient d'intercalibration au niveau vrai montré dans la Fig. 2b, sont du même ordre de grandeur que les incertitudes conçues pour les couvrir, montrées dans la Fig. 2c. Cela permet d'affirmer que les incertitudes de sélection ne sont pas sous-estimées.

Le travail réalisé lors de ce doctorat a également consisté en un grand nombre de vérifications que la méthode d'obtention de l'inter-étalonnage en η fonctionne bien comme attendu. Cela a notamment conduit à une compréhension plus profonde de l'usage de l'asymétrie $\mathcal{A}$, une grandeur intermédiaire indispensable pour cette méthode, et à la justification mathématique de son usage et de sa relation avec le coefficient d'intercalibration c.

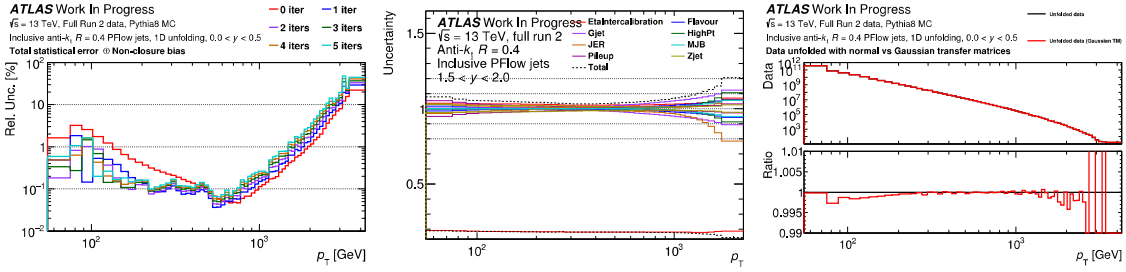
Mesures de sections efficaces de productions de jets

Une des prédictions historiques du Modèle Standard de la physique des particules est la section efficace σ inclusive de production de jets. Il s'agit d'une évaluation de la probabilité de produire des jets lors des collisions. D'un point de vue expérimental, cela consiste en un comptage des jets détectés dans ATLAS. Cette mesure est effectuée de manière doublement différentielle, en quantité de mouvement transverse p_T et en rapidité absolue $|y|$:

$$\frac{\mathrm{d}^2\sigma}{\mathrm{d}p_T\ \mathrm{d}|y|} = \frac{N}{\mathcal{L}\Delta p_T \Delta|y|} \tag{4}$$

où N est le nombre de jets détectés dans la région de l'espace des phases $(p_T, |y|)$ de volume $\Delta p_T \times \Delta|y|$, et où $\mathcal{L}$ est la luminosité enregistrée par ATLAS : il s'agit d'une normalisation qui quantifie le nombre de collisions de protons ayant été réalisées.

Une autre mesure de section efficace a été proposée récemment : la mesure de section efficace de production de jets dominants, également doublement différentielle en p_T et $|y|$. Dans ce cas, seulement un jet est compté par événement : le jet dont la quantité de mouvement transverse p_T est la plus élevée. Cette analyse comporte une difficulté : l'observable est sensible aux rayonnements à basse énergie ; en effet, l'émission d'un jet de basse énergie par le jet dominant peut faire passer celui-ci de dominant à sous-dominant. Il en est de même des effets de résolution du détecteur, qui peuvent faire qu'un jet d'une certaine énergie est mesuré à une énergie différente, et ce malgré la calibration de la JES (qui corrige le biais systématique, mais pas la dispersion statistique liée à la résolution en énergie des calorimètres). Il existe donc des migrations entre les jets dominants, sous-dominants, et sous-sous-dominants (les migrations avec les ordres supérieurs sont négligeables). Lors de la déconvolution des effets



(a) $|y| < 0.5$ (b) $1.5 < |y| < 2.0$ (c)

Figure 3: Incertitudes pour la région $|y| < 0.5$ de la mesure de section efficace inclusive de production de jets. **(a)** Somme en quadrature des incertitudes statistiques et du biais de sensitivité de la méthode de déconvolution aux différences de forme entre les données et le MC, déconvoluée en utilisant PYTHIA8. Le nombre d'itérations choisi est 1. **(b)** Incertitudes systématiques de JES et de JER propagées à travers la déconvolution, sommées en quadrature par catégorie, déconvoluées en utilisant PYTHIA8. Notez que l'incertitude d'inter-étalonnage en η est incorrecte et sera produite à nouveau. **(c)** Amplitude des effets non-gaussiens du MC. La partie supérieure montre les sections efficaces inclusives, déconvoluées soit en utilisant la matrice de transfert normale, soit en utilisant la matrice de transfert gaussienne, et la partie inférieure montre le rapport des deux, qui est l'incertitude de non-gaussianité (après soustraction de 1).

de détecteur, en plus de prendre en compte les migrations en p_T liées à la résolution, il faut également prendre en compte les migrations d'ordre, ainsi que les migrations en rapidité.

J'ai pu contribuer à ces deux analyses. Pour la mesure inclusive, j'ai effectué la déconvolution des effets de résolution du détecteur et la propagation des incertitudes. Pour la mesure de production de jets dominants, j'ai recommencé à une étape moins avancée l'analyse déjà amorcée par Robert HANKACHE : j'ai étendu l'analyse à la région $3.0 < |y| < 4.5$, j'ai amélioré la compréhension des migrations entre les différents ordres de jets et les migrations en rapidité, et j'ai produit les histogrammes nécessaires pour la déconvolution (spectres de section efficace, matrice de repliement, efficacités) ; mais la déconvolution ainsi que la propagation des incertitudes restent à terminer : de nombreux éléments sont déjà en place, les difficultés sont toutes en passe d'être résolues. Ainsi, je me concentrerai sur l'analyse inclusive dans le reste de ce synopsis.

La mesure de section efficace inclusive de production de jets consiste en la production des histogrammes contenant le nombre d'événements détectés dans chaque région de l'espace des phases, et en la correction des migrations dues à la résolution en énergie grâce à une technique de déconvolution. Les spectres avec résolution convoluée et déconvoluée sont reliés par une matrice de passage dite « matrice de repliement », notée M, obtenue grâce aux simulations MC pour lesquelles le niveau vrai et le niveau reconstruit (reco) d'un même événement sont connus. Les distributions MC vraie t et reco r sont reliées par $r = M \times t$. Ainsi, pour obtenir le spectre de données avec les effets de détecteur déconvolués u à partir du spectre de données d, il suffit en principe de réaliser l'opération $u = M^{-1} \times d$. Cette opération n'est pas idéale car les valeurs propres les plus petites de la matrice de repliement sont très sensibles aux fluctuations statistiques, mais des méthodes alternatives existent. La méthode employée est une méthode de déconvolution itérative, stabilisée dynamiquement (IDS), où le nombre d'itérations est une régularisation.

Cette méthode IDS, comme toute méthode de déconvolution, est sensible aux différences de forme entre les spectres des données et du MC, qui induisent un biais méthodologique. Ce biais est réduit par le nombre d'itérations de la méthode IDS, et est estimé de la façon suivante. Le spectre MC au niveau vrai est repondéré et convolué aux effets de résolution par la matrice de repliement, de sorte à avoir un spectre dont le niveau reco est semblable au spectre des données, et dont le niveau vrai associé est connu. Le spectre reco obtenu est alors déconvolué puis comparé au spectre vrai associé (le spectre repondéré), et la différence est une mesure du biais. J'ai évalué ce biais pour la première fois dans la région $3.0 < |y| < 4.5$, ce qui a nécessité quelques améliorations par rapport aux analyses précédentes.

La propagation des incertitudes statistiques à travers la déconvolution est également nécessaire. Celle-ci est réalisée avec la méthode de ré-échantillonage *bootstrap*, qui consiste à créer des distributions équivalentes statistiquement à la distribution originale en répliquant chacun des événements un certain nombre de fois suivant une distribution poissonienne, à les propager à travers la déconvolution, puis à mesurer la dispersion des répliques.

Lorsque le nombre d'itérations de la méthode IDS augmente, le biais de sensibilité aux différences de forme diminue, mais les incertitudes statistiques augmentent, il s'agit donc d'un compromis. Leur somme en quadrature est montrée pour la région $|y| < 0.5$ dans la Fig. 3a. Le nombre d'itérations optimal est identifié à 1.

Les incertitudes sur l'étalonnage de l'échelle d'énergie des jets (JES) et sur la résolution en énergie des jets (JER) doivent être propagées à travers la déconvolution. La correction appliquée aux jets est modifiée, soit pour les jets des données soit pour les jets du MC, et la déconvolution est effectuée à nouveau sur les nouveaux spectres et avec la nouvelle matrice de repliement, puis le spectre déconvolué est comparé au spectre nominal. Une incertitude statistique est également estimée pour chacune de ces incertitudes systématiques au moyen de la méthode *bootstrap*. Ainsi, chacune des 2432 sources d'incertitudes est propagée individuellement, avec des répliques *bootstrap*. Ceci rend l'analyse particulièrement lourde : elle nécessite beaucoup de mémoire de stockage, beaucoup de RAM, et un assez grand nombre de CPUs. Il a donc été nécessaire d'optimiser l'analyse du point de vue computationnel, soit en réduisant le nombre de répliques *bootstrap* lorsque c'est possible, notamment pour les incertitudes de statistique de l'inter-étalonnage en η, soit en optimisant la librairie *bootstrap* elle-même : j'ai pu améliorer sa vitesse d'exécution d'un facteur trois, une amélioration utile à tous ses usagers. Les incertitudes systématiques sont montrées dans la Fig. 3b.

J'ai également évalué une nouvelle incertitude pour la première fois dans une analyse de section efficace dans ATLAS. Il s'agit d'une estimation de l'effet causé par les non-gaussianités présentes dans la simulation MC, qui n'est pas mesuré *in situ*. En effet, la JER, mesurée *in situ*, est estimée comme étant la première gaussienne d'une distribution. La matrice de repliement donne accès aux effets non gaussiens, et prend donc en compte ces effets pour le résultat nominal. Ainsi, en déconvoluant les spectres avec une matrice de repliement construite à partie de gaussiennes, on peut estimer cet effet, qui est montré dans la Fig. 3c, et y associer une incertitude.

Une fois les spectres déconvolués obtenus, ainsi que les estimations des incertitudes, il est possible de les comparer aux prédictions théoriques. L'accord entre données expérimentales et prédiction théorique est correct lorsqu'on regarde seulement une tranche de l'espace des phases, mais n'est pas bon lorsqu'on calcule un $\frac{\chi^2}{N_{\text{dofs}}}$ global. Ce problème est connu de longue date, et avant de conclure à un désaccord entre théorie et expérience, plusieurs méthodes sont tentées afin de réduire le χ^2. Notamment, les incertitudes statistiques de l'inter-étalonnage en η doivent être prises en compte indépendamment, d'où le très grand nombre d'incertitudes systématiques à propager. Nous avons également montré récemment qu'utiliser un inter-étalonnage en η non lissé réduit significativement le χ^2. D'autres approches ont été tentées, notamment concernant le schéma de corrélation des incertitudes sur les prédictions théoriques. Ces progrès ne sont pas suffisants pour obtenir un $\frac{\chi^2}{N_{\text{dofs}}}$ raisonnable, mais il reste des idées à essayer pour continuer d'améliorer ce χ^2.

À Papa et Pabousou.

After sleeping through a hundred million centuries, we have finally opened our eyes on a sumptuous planet, sparkling with color, bountiful with life. Within decades we must close our eyes again. Isn't it a noble, an enlightened way of spending our brief time in the sun, to work at understanding the universe and how we have come to wake up in it?

Nightwish, in *The Greatest Show on Earth*, 2015

Remerciements

Heureux qui, comme Ulysse, a fait un beau voyage.

Joachim du Bellay, *Les Regrets*, 1558

Me voilà au terme d'une étape majeure de mon odyssée démarrée il y a neuf ans (déjà) ! Et quelle étape ! Un doctorat ! Bien loin de l'autoroute de mes études précédentes, qui n'étaient pourtant déjà pas une sinécure, loin s'en faut, il s'est agi tantôt d'un sentier aux contours flous et d'un azimut incertain, tantôt d'un fleuve impétueux et tumultueux, avec tout de même une boussole et quelques accalmies deçà delà. Je rejoins Ulysse en cela, mais pas seulement : quoiqu'il termina son périple seul, ainsi que je le fus lorsque je me trouvais face à ma page blanche pendant ces nuits estivales, ce n'est pas sans son équipage qu'il parvint à traverser les épreuves et arriver à bon port. Redevable, je le suis également, à bien des égards, et il m'échoit maintenant de remercier ici toutes celles et tous ceux sans qui cette épopée eût certainement pris un tournant bien différent.

En premier lieu, je souhaite bien entendu remercier ceux sans qui l'aventure n'aurait même pas commencé, Bogdan Malaescu et Mélissa Ridel, qui m'ont accepté en doctorat il y a de ça déjà trois ans et demi. Plus que des études, le doctorat est un apprentissage, et je remercie Bogdan non seulement pour les discussions de physique que nous avons eues, mais également pour tous ses conseils qu'il m'a donnés, que ce soient des méthodes de travail, pour ma vie personnelle ou ma carrière. Bogdan s'est toujours rendu très disponible pour moi, et m'a soit accompagné au quotidien soit laissé de la liberté, chaque fois que le besoin s'en faisait sentir, et ce toujours avec la volonté de faire de moi un chercheur. Je lui suis reconnaissant de m'avoir enseigné la rigueur scientifique et l'exigence de la qualité qui est nécessaire en recherche ; je pense que ce n'est pas donné à tout le monde. Je lui suis également reconnaissant d'avoir été compréhensif lorsque ma vie privée a dû prendre le pas sur ma vie professionnelle (par exemple en me laissant le temps d'organiser à nouveau mon mariage, qui a été reporté à cause de la pandémie de Covid-19).

Je remercie Sergey Barsuk, Suzanne Gascon-Shotkin, Matt Leblanc et Luco Scotto-Lavina pour avoir accepté d'être membres de mon jury de thèse.

Je remercie l'équipe enseignante de l'UE « modélisation numérique en Python », et principalement Pacôme Delva qui m'a donné beaucoup de latitudes pour expérimenter la pédagogie ; quoique cela m'a demandé beaucoup de travail, j'ai pu avoir une expérience très complète de l'enseignement à l'université : (ré)écriture du cours, TD et TP en classe avec les élèves, barème d'évaluation, notation des copies, suivi et évaluation des projets, suivi des élèves. J'en profite pour remercier les élèves qui se sont investis dans mon cours, et avec qui il a été particulièrement agréable d'échanger après les cours (parfois plus d'une heure, au mépris de leurs pauses déjeuner !) Je conserve précieusement votre *mail* de remerciements qui m'a fait chaud au cœur !

J'en profite pour remercier ces professeurs excellents qui, au cours de ma scolarité, m'ont transmis leur passion : Cécile Mucignat (mythologie, latin et grec), Mme Barotte (mathématiques), M. Dumonteuil (mathématiques), Thibaud Naulet (physique), Matthieu Émond (chimie), Christophe Imbert (physique), Rémi Le Roux (chimie organique), Olivier Dauchot (physique statistique), Marios Petropoulos (relativité générale). Chaque fois, vous avez cru en moi avant que je ne le fasse moi-même, vous m'avez fait confiance ou simplement donné une chance ; à ce titre, je remercie également Geneviève Fleury de m'avoir accepté en stage en 2017.

Je remercie les groupes et sous-groupes d'ATLAS avec lesquels j'ai collaboré, notamment lors des réunions hebdomadaires : JetETmiss, *in situ*, jetXsec. Il y a toujours eu une excellente ambiance de travail, toujours encourageante, et j'ai appris beaucoup des retours et commentaires que l'on me prodiguait. En fait, je me suis senti accueilli, encouragé à participer et à prendre des responsabilités, et c'était très motivant de savoir que d'autres gens attendaient mon travail, s'y intéressaient. Je remercie tout particulièrement Tancredi Carli pour son suivi attentif et ses explications qui m'ont permis de me lancer dans le projet de sections efficaces sans perdre de temps, Davide Melini pour avoir pris en charge les dernières étapes de l'inter-étalonnage en η des jets à grand rayon, Jonathan Bossio pour son aide sur différentes questions techniques et pour son goût du travail bien fait (il n'y a pas que les résultats qui comptent, la qualité du code aussi est cruciale), Daniel Camarero Muñoz et Matt Leblanc pour leurs encouragements, Reina Camacho pour les grandes discussions (à HEP2023 et au CERN) qui m'ont

vraiment aidé dans mes réflexions professionnelles et dans ma vie personnelle, ainsi que pour son accueil au CERN, et finalement toute l'équipe JetETmiss avec qui j'ai vite sympathisé lors de mon bref séjour au CERN. Je remercie le groupe *in situ* de m'avoir fait confiance pour animer une session lors de l'atelier HCW.

Je remercie Robert Hankache, précédent doctorant de Bogdan et Mélissa, qui a travaillé sur l'inter-étalonnage en η et sur l'analyse de la section efficace de production de jets dominants : il m'a expliqué beaucoup de choses sur la physique et les aspects techniques, en visioconférence ou lors de mes deux semaines à Manchester, où Robert m'a accueilli avec beaucoup d'attentions. Je le remercie également pour les longues conversations que nous avons échangées, et pour la découverte de la dégustation de whiskys !

Je remercie le personnel administratif et technique du LPNHE, ainsi que le groupe ATLAS du LPNHE, notamment pour leur aide technique à distance durant le Covid-19, pour la réactivité des procédures administratives, et pour m'avoir permis d'aller en missions à Manchester et au CERN, et en conférence à HEP2023 au Chili. Je remercie tout spécialement Karine Marquois pour son aide patiente avec le service après-vente de DELL, les cinq fois où j'ai eu des pièces défectueuses.

Je remercie Sébastien Bongard, mon parrain au LPNHE, qui m'a donné un grand nombre de conseils personnels, qui a toujours été là pour discuter dans les moments les plus difficiles. J'espère que tu m'autorises cette citation apocryphe qui m'a marquée : « le doctorat, ce n'est pas seulement un marathon, c'est Compostelle : il faut tenir le chemin mais surtout il faut savoir ce pour quoi on le fait ». Tu as toujours su poser les bonnes questions, trouver les mots et les métaphores adéquates à toutes les situations, et aider mon discernement par la profondeur de nos discussions, tout en restant simple et concret. Je te remercie pour ta considération, ton temps et ta sagesse.

Ce doctorat m'a permis de faire de belles rencontres avec des personnes extérieures au LPNHE, et avec qui j'ai eu d'excellentes conversations (de physique entre autres) et de nombreux conseils personnels. Notamment, je remercie tout particulièrement Mark Owen et Eva Santos pour leurs encouragements à poursuivre la voie universitaire, rencontrés à HEP2023, tout comme Iaroslava Bezshyiko avec qui j'ai passé une semaine de vacances extraordinaires dans l'Atacama. C'est également grâce à cette conférence que j'ai eu l'occasion de discuter avec Reina Camacho, du LPNHE et de JetETmiss, qui m'a tant aidé notamment à avoir confiance en moi ou en mon travail ; je ne saurais trop te remercier pour ces galères au Chili où tu m'as prêté main forte : cette première journée à Valparaíso où nous n'arrivions pas à payer en carte bleue et où nous avons tourné dans la ville pour trouver un distributeur, et cet avion retour que j'ai bien failli manquer à deux minutes près... Avec le recul, ça fait des histoires à raconter !

Il est impératif que je remercie Valentina Raskina le plus chaleureusement possible. Nous avons partagé le bureau 12-22-213 pendant ces trois années, avons vécu nos joies et nos peines ensemble. Nos déjeuners au sushi rue Linné, à refaire le monde, me manquent déjà. Merci pour tes encouragements, ton aide et ton soutien, particulièrement ces quelques derniers mois, c'est vraiment grâce à toi que je suis allé jusqu'au bout. Merci pour ton accueil au CERN, merci pour les soirées à l'escalade, pour ces danses mémorables, pour la passion du chocolat, pour les soirées au Lapin Agile, pour les nombreuses découvertes de restaurants. Those were hard days, but I will forever cherish the memory of every smile that made them worth living. Je remercie Raoul Follereau de pallier l'imperfection de mes mots :

> « Un sourire ne coûte rien et produit beaucoup, [...]
> Il ne dure qu'un instant, mais son souvenir est parfois éternel. »

J'adresse également mes plus sincères remerciements à Romain Bouquet, que ce soit pour les longues discussions ou pour les nombreuses aides techniques (je serais certainement encore en train de chercher ce bug avec le `JetCalibTool` et les `TStores` sans ton aide...)

Je remercie Anna Niemic, qui m'a vraiment énormément motivé avec des encouragements simples, notamment en me disant « ça serait chouette que ta prochaine participation à Space Apéro soit après ta soutenance, pour avoir l'avant et l'après », ou encore lors de la répétition de soutenance que tu m'as proposée.

Je remercie Line Delagrange, avec qui j'ai eu tant de discussions sur l'inter-étalonnage en η, à comprendre les subtilités conceptuelles et techniques de la méthode en long, en large et en travers. Nous sommes vétérans de ce code affreux, nous y avons survécu !

J'adresse mes remerciements les plus sincères à l'ensemble de l'équipe de doctorants et post-docs du LPNHE : vous êtes vraiment des gens formidables, toutes et tous, et je suis ravi d'avoir passé ces années à vos côtés ! Votre soutien m'a été inestimable ! Je garde d'incroyables souvenirs de nos soirées raclette/fondue/gaufres/crêpes, ou simplement houmous et guacamole, de karaokés, de danses... ou à la Bulle à Bière ! La « Ph.D. biennale » était une idée géniale, mémorable ! Merci à tous pour votre amitié ! J'espère n'oublier personne : en plus de celles et ceux cités précédemment, je remercie Alberto Rosales, Alessandro Scarabotto, André Malvezzi Lopes, Anthony Correia, Ariel Matalon, Artur Oudot, Christina Agapopoulou, Claire Dalmazzone, Claudia de Dominicis, Edmar Egidio Purcino de Souza, Enya van den Abeele, Fotis Giasemis, Giorgos Papadopoulos, Greeshmani Doddi, Jad Sardain, Julianna Stermer, Lata Panwar, Laura Boggia, Lavinia Russo, Leander Lacroix, Lucile Mellet, Luis Pascual, Luiz Balabram, Marion Guelfand, Michelangelo Traina, Mykola Khandoga, Pablo Correa, Paul Chabrillat, Reem Taibah, Renaud Amalric, Romain van den Broucke,

Sara Maleubre, Slava Trusov, Thierry Souverin, Tommaso Fulghesu, Ugo Pensec, Ulysse Virginet, Viet Nguyen, Vlada Yevarouskaya, Yajun He, Yongyu Pan. J'aimerais écrire un paragraphe pour chacun d'entre vous, comme je l'ai déjà fait pour ceux qui avez déjà soutenu !

Je remercie tous mes amis en dehors du LPNHE pour leur amitié et leur soutien, et tout particulièrement Anthony Guillen avec qui j'ai eu tant de plaisir à aller à l'escalade et à discuter de physique théorico-spéculative ! Je remercie également Agathe Chapuis, Paul-Adrien Blancquart, Sylvain Cam, Aurore Vullien, Marie-Lou Novène, Nina Du, Mickaël Paulon, Antoine Adeline, Angelica Savannet, Elina Gilbert, Arnaud Chaub, Michella Khoury Damaa, Pierre Glidic, Théo Lenavetier, ainsi que tous mes nombreux amis de Saint Joseph, de Ginette, de l'ESPCI ou de l'ICFP que je n'ai pas revus autant que j'aurais voulu ces dernières années mais qui m'ont fait l'honneur et le plaisir de venir à mon mariage !

Je remercie ma famille, qui m'a toujours soutenu, encouragé et motivé dans tous mes projets, a toujours pu se rendre disponible, que ce soit pour de l'aide pratique comme un déménagement ou l'organisation de mon mariage, ou pour discuter de tous aspects de la vie. J'ai la chance d'avoir une famille passionnée et passionnante, qui m'a offert des valeurs morales solides et cohérentes. Je remercie tout particulièrement mon père, qui m'a le premier transmis sa vision du monde et sa passion pour la technicité et pour la physique, son regard curieux et émerveillé sur la nature ; et mon grand-père, avec qui j'ai régulièrement des discussions pendant des week-ends entiers avec toujours autant de passion, à disserter sur tout ce qui existe. Je vous dois une approche de la science qui ne soit pas seulement utilitaire mais surtout qui véhicule une vision du monde ; une science qui ne soit pas prescriptive ou coercitive mais au contraire un espace d'imagination et de créativité intellectuelle ; une science comme un édifice immense, ancré dans une histoire humaine riche, où des ponts sont sans cesse créés entre les domaines ; une science comme une frontière admirablement bien située entre l'abstraction intellectuelle et l'intuition, sans jamais se départir de son rapport matériel et concret au réel.

Enfin, j'adresse mes remerciements à Anne-Caroline, mon épouse. Tu m'as apporté un soutien de chaque instant, as toujours été très compréhensive malgré mes heures de travail particulièrement tardives, mon acception de la notion de « matin », le bruit de mon ordinateur, et j'en passe... Tu étais avec moi pour chaque réussite, aussi petite fut-elle, tu as écouté tous mes longs monologues lorsque les choses n'allaient pas comme je le souhaitais, tu as compris mon indisponibilité générale à de nombreux moments de ce long projet de doctorat. Ces quelques mots extraits du poème magnifique *Que serais-je sans toi ?* de Louis Aragon, mis en musique et chanté par Jean Ferrat, expriment mieux mes sentiments que le mot « merci » :

« Que serais-je sans toi, qui vins à ma rencontre ?
Que serais-je sans toi qu'un cœur au bois dormant ?
Que cette heure arrêtée au cadran de la montre ?
Que serais-je sans toi que ce balbutiement ? »

Finalement, je souhaite également remercier Frédéric Chopin pour ses nocturnes, Juan José Mosalini pour son Alma de Tango, Maksim Mrvica pour sa rhapsodie croate, Nolwenn Arzel pour sa harpe celtique, Ferrat chantant Aragon. Je remercie JBX pour *Reflets d'Acide*, John Lang pour *Le Donjon de Naheulbeuk*, Pierre Christin et Jean-Claude Mézières pour *Valérian et Laureline*, Alain Ayroles et Jean-Luc Masbou pour *De Cape et de Crocs*. Au vu de la consommation immodérée que j'en ai faite, il me semble raisonnable de remercier le chocolat pour son existence, et plus spécifiquement les tablettes 70 % de chez Chapon, le noir truffé noir et le noir fourré à la pâte d'amande de chez Côte d'Or, les 70 %, 85 % et 70 % au coulis de chocolat de chez Lindt. Ces diverses œuvres artistiques ou culinaires m'ont accompagné de près sans jamais faillir !

Table of Contents

Résumé **i**

Abstract **iii**

Synopsis **v**

Remerciements **xvii**

Table of Contents **xxi**

Introduction **1**

1 The Theoretical Framework **3**
Introduction . . . 3
1.1 The Quantum Field Theory Framework . . . 4
1.1.1 The need for a quantum theory of relativistic fields . . . 4
1.1.2 Lagrangian formalism and symmetries . . . 5
1.1.3 Perturbative approach . . . 6
1.1.4 Renormalization . . . 8
1.2 The Standard Model of Particle Physics . . . 9
1.2.1 Symmetries of the Standard Model Lagrangian . . . 9
1.2.2 Field content of the Standard Model . . . 10
1.2.3 The Lagrangian of the Standard Model . . . 12
1.2.4 The zoo of hadrons . . . 14
1.2.5 The partonic model . . . 14
1.3 Jets . . . 17
1.3.1 Hadron production, and event generation in simulations . . . 18
1.3.1.1 The hard subprocess in a proton-proton collision . . . 19
1.3.1.2 Parton showers . . . 20
1.3.1.3 Hadronization . . . 21
1.3.1.4 Soft QCD: underlying event and collective effects . . . 23
1.3.1.5 Hadron decay . . . 24
1.3.1.6 Event generators . . . 24
1.3.1.7 Partonic predictions and corrections . . . 24
1.3.2 Jet definitions . . . 25
1.3.2.1 Requirements . . . 25
1.3.2.2 Jet algorithms . . . 27
1.4 Physics beyond the Standard Model . . . 31
1.4.1 Hints at physics beyond the Standard Model . . . 31
1.4.2 Search strategies for physics Beyond the Standard Model . . . 34

2 The ATLAS Experiment **37**
Introduction: CERN . . . 37
2.1 The Large Hadron Collider . . . 38
2.1.1 Operation periods . . . 38
2.1.2 The acceleration infrastructure . . . 39
2.1.3 The beam of the Large Hadron Collider . . . 42
2.2 The ATLAS detector . . . 45

2.2.1 System of coordinates . . . 47
2.2.2 Particle energy loss and shower development . . . 49
2.2.3 Magnet systems . . . 52
2.2.4 Inner detector and tracking system . . . 54
2.2.5 Electromagnetic and hadronic calorimeters . . . 56
2.2.6 Muon spectrometer . . . 60
2.2.7 Forward detectors . . . 62
2.2.8 Triggers and data acquisition . . . 64
2.2.9 Data processing and object reconstruction . . . 66

3 The Reconstruction and Calibration of Jets in ATLAS 71
Introduction . . . 71
3.1 Jet reconstruction . . . 72
3.1.1 Particle-like object reconstruction . . . 72
3.1.1.1 Topological clusters . . . 72
3.1.1.2 Particle Flow Objects . . . 75
3.1.1.3 Constituent-level pile-up rejection . . . 76
3.1.1.4 Unified Flow Objects . . . 78
3.1.2 Jet reconstruction . . . 80
3.1.2.1 Jet algorithm . . . 80
3.1.2.2 Large-R jet grooming . . . 80
3.2 Quality criteria . . . 81
3.2.1 Pile-up rejection . . . 81
3.2.2 Event-level selection . . . 82
3.2.3 Timing selection . . . 84
3.2.4 Jet-level selection . . . 84
3.3 Jet calibration . . . 87
3.3.1 The calibration chain of small-R jets . . . 88
3.3.1.1 Pile-up corrections . . . 89
3.3.1.2 Absolute MC-based calibration . . . 91
3.3.1.3 Global calibration . . . 92
3.3.1.4 Flavor uncertainties . . . 94
3.3.1.5 E/p calibration . . . 96
3.3.2 Residual *in situ* calibration . . . 98
3.3.2.1 η-intercalibration . . . 100
3.3.2.2 V+jet balance . . . 101
3.3.2.3 Multijet balance . . . 103
3.3.2.4 Combining the corrections and uncertainties . . . 103
3.3.3 Jet Energy Resolution . . . 105
3.3.3.1 Jet Energy Resolution measurement with dijet events . . . 105
3.3.3.2 The noise term of the Jet Energy Resolution . . . 107
3.3.3.3 Jet Energy Resolution combination and results . . . 108

4 Towards an Absolute η-intercalibration 109
Introduction to the η-intercalibration . . . 110
4.1 The absolute implementation for the *in situ* calibration . . . 113
4.1.1 Deconvolving physics effects and detector effects in the intercalibration coefficient . . . 113
4.1.2 Benefits of the absolute implementation . . . 114
4.1.2.1 Isolation of the different effects and meaning of the calibration . . . 114
4.1.2.2 Possible improvements in the uncertainties . . . 114
4.1.3 Dealing with the absolute implementation of the calibration in a physics analysis . . . 115
4.2 Detailed description of the η-intercalibration . . . 117
4.2.1 Event samples . . . 117
4.2.1.1 Selection criteria of events . . . 118
4.2.1.2 Triggers . . . 121
4.2.1.2.1 Triggers as event selectors for a given phase space region . . . 121
4.2.1.2.2 Determining the turn-on points of combinations of central and forward trigger chains with the emulation and bootstrapping methods . . . 122
4.2.1.2.3 Determining the turn-on points missing from the previous analysis with a fit 127

4.2.1.3 MC normalization . . . 128
4.2.1.4 Data normalization . . . 128
4.2.2 The binning . . . 129
4.2.2.1 The p_T^{avg} binning . . . 130
4.2.2.2 The η binning . . . 131
4.2.2.3 The asymmetry binning . . . 132
4.2.2.4 The final (p_T, η) binning . . . 132
4.2.3 From the asymmetry to the calibration: overview . . . 133
4.2.4 The asymmetry . . . 134
4.2.4.1 Determining the mean asymmetry in each bin . . . 136
4.2.5 The central reference method to obtain the intercalibration coefficient . . . 139
4.2.6 The matrix method to obtain the intercalibration coefficient . . . 140
4.2.6.1 Description . . . 140
4.2.6.2 The analytic solution . . . 142
4.2.6.3 Automatic rebinning in η . . . 144
4.2.6.4 Treatment of the central region . . . 145
4.2.6.5 Propagation of the statistical uncertainties . . . 145
4.2.7 From the intercalibration coefficients to the calibration . . . 146
4.2.7.1 Ratio of intercalibration coefficients . . . 146
4.2.7.2 From a p_T^{avg} binning to a p_T binning . . . 147
4.2.7.3 Smoothing . . . 147
4.2.8 Uncertainties . . . 150
4.2.8.1 Statistical uncertainties on the calibration . . . 150
4.2.8.2 Systematic uncertainties . . . 150
4.2.8.3 Binning and smoothing . . . 151
4.2.8.4 Symmetrization . . . 152
4.2.8.5 Combination of the systematic uncertainties . . . 153
4.2.8.6 Checking at truth level that systematic uncertainties originating from the event selection do account for the physics effects . . . 153
4.2.8.7 Using the truth level for the Monte Carlo modeling uncertainty . . . 155
4.3 Calibrations and their uncertainties: results . . . 156
4.3.1 Precision calibration with the relative implementation . . . 156
4.3.1.1 The calibration and its uncertainties . . . 156
4.3.1.2 Closure test . . . 159
4.3.2 Comparison of the relative and absolute implementations . . . 162
Conclusion . . . 165

5 Jet Production Cross-section Measurements 167
Introduction to the inclusive and leading jet production analyses . . . 168
5.1 Dataset preparation . . . 171
5.2 Unfolding . . . 172
5.2.1 The Iterative Dynamically Stabilized unfolding method . . . 172
5.2.2 Preparing histograms for the unfolding . . . 173
5.2.2.1 Choice of the binning . . . 173
5.2.2.2 The transfer matrix . . . 175
5.2.2.3 Choice between rapidity and pseudo-rapidity . . . 176
5.2.2.4 Matching efficiency and purity . . . 179
5.2.3 Application of the IDS unfolding method to the inclusive jet cross-section analysis . . . 179
5.2.3.1 First spectra . . . 179
5.2.3.2 Statistical uncertainties and correlations with the bootstrap method . . . 180
5.2.3.3 Bias related to the sensitivity of the unfolded method to the data/MC shape differences 183
5.2.3.4 Choice of the number of iterations . . . 187
5.3 Experimental uncertainties . . . 189
5.3.1 Propagation of the uncertainties on the JES and the JER . . . 189
5.3.2 Non-Gaussianity of the JER . . . 192
5.3.3 Other uncertainties . . . 193
5.4 Technical optimizations . . . 194
5.5 Results . . . 195
5.5.1 Unfolded data and uncertainties . . . 195

5.5.2 Comparison to theory predictions . . . 201
Conclusion and prospects . . . 203

Conclusion and Perspectives **205**

Bibliography **207**

A Appendix for the η-intercalibration **215**
A.1 Brief description of the code for the η-intercalibration . . . 215
A.2 Differences for the large-R jets . . . 215
A.3 Internal tags of the datasets used, for users within ATLAS . . . 217
A.4 Computation of the barycenters of the (p_T^{avg}, η) bins . . . 218
A.5 Analytic solution for the matrix method when combining η bins . . . 219
A.6 The whole calibration . . . 224
A.7 Precision calibration for large-R jets . . . 224

B Appendix for the Cross-section Measurements **233**
B.1 Detailed list of all the JES and JER uncertainties . . . 233

Introduction

What is matter made of? Such a common, and yet highly rich question. Greek philosophers 2500 years ago were already speculating on the answer to this question, debating between the continuum and the atomic hypotheses, though of course not having any tool at the time to get an answer. A new era in the never-ending quest to understand the secrets of Nature began over a century ago, when scientists tried for the first time to accelerate particles and make them collide [1]. It began with the CROOKES tube, that allowed THOMSON in 1897 to understand that the cathode rays are composed of negatively charged particles, 1800 times lighter than the hydrogen atom and yet constituents of the atoms, later called electrons, that were accelerated to $60,000$ km/s, *i.e.* 20 % of the speed of light. The same approach and basic ingredients – electric field to accelerate, magnetic field to deviate, vacuum, interactions to detect – have been used afterwards in countless experiments, among which the Large Hadron Collider (LHC), based at CERN, and its detectors such as ATLAS. The LHC produces proton collisions at the highest center-of-mass energies ever reached: during its first data-taking campaign, from 2010 to 2012, it produced collisions at 7 TeV and 8 TeV; and during its second run, from 2015 to 2018, it reached the energy of 13 TeV. During Run 2, the ATLAS experiment, to which I contribute, recorded the impressive integrated luminosity of 140 fb^{-1}.

Going hand in hand with theory, these experiments progressively led to the discovery of a plethora of particles and to a much deeper understanding of the fundamental forces, nowadays classified and understood in tremendous detail in a unified theory known as the Standard Model (SM) of particle physics. This monument is based on a quantum theory of fields, and describes with only 26 free parameters the rich phenomenology of three forces – the electromagnetic force, the strong and weak nuclear forces – 61 fields and their properties as well as the associated fundamental particles, and hundreds of composite particles. The SM explains matter and interactions on no less than 12 orders of magnitude, from the electronvolt to the teraelectronvolt energy scales, with barely any contradiction from experiments. And yet, despite being incredibly successful, the SM is not the end of the story, as there still are gaps to be filled, that maybe are breaches awaiting to make a dent in. The hints that new Beyond the Standard Model (BSM) physics exists come from multiple areas: the SM does not incorporate gravity and does not provide an explanation for dark matter and dark energy, nor for the 26 free parameters; current knowledge of neutrinos is limited, the matter-antimatter asymmetry of the universe is unexplained, *etc.* However, these hints do not indicate any clear way where to search for, and modern-day experiments like ATLAS either directly search for predictions of BSM models, or test every aspect of the SM with the highest precision achievable.

This thesis focuses on the precision approach, and on one particular observable: jets. Jets are ubiquitous at the LHC: they have the highest production cross-section by far compared to any other process, and are therefore used by most physics analyses, either as signal or as background. They are collimated sprays of hadrons, stemming from the hadronization of a high-energy quark or gluon emitted by the proton-proton collision. The particles emitted in a conical shape enter the ATLAS detector: they produce tracks in the tracking systems and energy deposits in the calorimeters, which have to be clustered together to reconstruct the jet and obtain the properties of the initial quark or gluon. Jets built from particles or from detector inputs exhibit differences in their observables, like their transverse momentum p_T, due to detector effects. Corrections are therefore necessary to enable comparison between experimental results and theoretical predictions. Two strategies are discussed in this thesis: calibration and unfolding, to correct for the detector effects respectively in the scale and the resolution of the transverse momentum of jets.

On the calibration side, I worked on the η-intercalibration, which is part of the calibration chain of the Jet Energy Scale (JES). The first steps in the JES are the removal of the effects of additional proton-proton collisions from the same bunch crossing, an effect known as "pile-up". This is crucial as there were on average 33.7 collisions per bunch crossing during the Run 2 data-taking period. The next steps are based on events simulated using Monte Carlo (MC) simulation techniques, to capture the main features of the detector effects, and to improve the resolution. Finally, *in situ* techniques correct for the residual detector effects, measured directly from data events. There are multiple such *in situ* techniques, one of which is the η-intercalibration. This step aims at making the detector response homogeneous across the whole pseudo-rapidity η range of the detector, by exploiting the transverse momentum conservation in dijet systems. I produced a precision calibration for the whole Run 2 data-taking period,

and worked on the improvement of the method with an "absolute implementation", that aims at a better separation of physics and detector effects in the η-intercalibration, leading to improvements in the systematic uncertainties of this method.

On the physics analysis side, I contributed to two closely related precision measurements: the inclusive and leading jet production cross-section measurements. Those are double-differential measurements that count jets in bins of transverse momentum p_T and absolute rapidity $|y|$. While the inclusive measurement counts every single jet produced above a certain p_T threshold, the leading jet measurement only counts one per event: the one with the highest p_T. The Jet Energy Resolution (JER) effects are corrected for by the Iterative Dynamically Stabilized unfolding (IDS) method. After careful evaluation of uncertainties of various origins, results for the inclusive jet cross-section are shown and compared to theoretical predictions.

This thesis is organized as follows:

- Chapter 1 presents the theoretical framework. Phenomenological aspects of Quantum Field Theory (QFT) are explained, as QFT is the foundation of the Standard Model of particle physics, that is then presented. Jets are then described in more detail, from their formation to their reconstruction. A glimpse at physics Beyond the Standard Model (BSM) is shown, to motivate the continued experimental efforts.
- Chapter 2 presents the experimental aspects. The Large Hadron Collider (LHC) is introduced. Then the ATLAS experiment is presented: its structure, and the data-taking operations and conditions of the Run 2 period.
- Chapter 3 presents in detail the treatment of small-radius jets in ATLAS: their reconstruction, selection and calibration.
- Chapter 4 presents my work on the η-intercalibration. The absolute implementation approach is introduced, then the η-intercalibration is presented and discussed in full details, leading to the presentation and discussion of the results.
- Chapter 5 presents my contributions to the jet production cross-sections. The event selection is introduced, then the unfolding procedure is detailed along with the estimation of the statistical uncertainties and the evaluation of the sensitivity of the unfolded result to the particle-level distribution in the Monte Carlo simulation. Then the propagation of the systematic uncertainties is explained. Finally, results are shown, compared to theoretical predictions and discussed.

Chapter 1

The Theoretical Framework

> Les formules algébriques ne sont que des échafaudages analogues à ceux qui ont servi à construire un palais admirablement conçu : que les chiffres tombent, et le palais d'Uranie resplendit dans l'azur, offrant aux yeux émerveillés toute sa grandeur et toute sa magnificence !
>
> Camille Flammarion, *Astronomie populaire*, 1890

Introduction

The Standard Model (SM) of particle physics is one of those successful theories of physics that have withstood the test of time. It is truly a monument of modern physics, as it achieved the long-standing goal of combining phenomena that appear to have little in common at first glance: light, electricity, magnetism, particles, atoms, nuclear reactions. It is at the crossroad of two of the most successful theories of the early XXth century, quantum mechanics and special relativity, and provides a solid basis for nuclear physics, light-matter interactions, astrophysics, cosmology, and all macroscopic emergent phenomena like chemistry, thereby connecting the infinitely small, infinitely big and infinitely complex. The SM is based on interacting fields, thus also owing a lot to classical mechanics, and it revolves around the notion of symmetry through Noether's theorem. Despite its mathematical complexity, the SM has proven to be useful in making predictions that are not contradicted with experiment up to an unprecedented degree in the history of science, fruitful to engineering with dozens of applications, and successful in providing an intelligible representation of nature. Beautiful.

Yet, the SM is incomplete. It realizes the tour-de-force to include three fundamental forces of nature in one consistent theory: electromagnetism, the weak and strong nuclear forces; but it does not include the fourth one, gravity. It provides a coherent understanding for all particles (whether fundamental or composite) observed on Earth to date, explaining their properties, dynamics and interactions; but it does not explain dark matter and dark energy, required for astrophysical and cosmological observations. Despite its extensive use of advanced mathematics, it lacks internal coherence: the theoretical foundations of the SM do not meet modern-day's standards. And it is not fully satisfying from an epistemological point of view, as it leaves a few fundamental questions unanswered.

Here lies one of the biggest problems of modern-day particle physics: in spite of its incomplete character, the SM still holds. It is expected to fail, for all those reasons, but it really doesn't, thus not paving the way to any clear directions where physicists could engulf their efforts for the search for what stands beyond. What then? Well, searching everywhere, all-round. Maybe not truly all-round, as there are way too many possibles of what physics Beyond the Standard Model (BSM) could be compared to the efforts that experimental analyses require. "Well-motivated" observables are chosen: those that are the most easily accessible from an experimental point of view, or those that are the most likely to corroborate or falsify a large number of theories. "Jets" are one such observables, that pertain to both categories.

Jets stem from high-energy particle collisions like the ones produced at LHC, and result from the hadronization of an outgoing quark or gluon. They are collimated sprays of hadrons, emitted in a conical shape, that reach the detector. Jets are essential probes for physics. They provide privileged access to phenomena of Quantum Chromo-Dynamics (QCD), the section of the SM that deals with the strong nuclear force, and to many models of BSM physics. They are ubiquitous at the LHC, their production cross-section is several orders of magnitude above any other process, making jets used by most analyses either as signal or as background.

The understanding of jets is however not straight-forward, with this chapter only introducing the theoretical aspects, leaving the experimental ones for Chapter 3. A quark or a gluon radiates soft or collinear gluons, that themselves can split into quark-antiquark pairs, until reaching the non-perturbative threshold. This comes with two issues: calculations are only done up to a certain order in perturbation theory, thus the quantum corrections due

to higher order terms are always missing; and there is currently no mathematical tool to treat the hadronization step (and non-perturbative phenomena in general) from first principles. In both cases, QCD-inspired models are used, that introduce *ad hoc* parameters, trying to capture as best as possible the main effects, and still providing a phenomenological (yet incomplete) understanding of the underlying physics. To complicate things further, there is no unique definition for jets. Particles emitted are clustered together using a jet algorithm, but there is no unique way to do so: they can be clustered if they fall into a cone with different possible choices for the cone size, or clustered sequentially with different possible orders for the clustering sequences.

This chapter is organized as follows. In Section 1.1, the essential elements of Quantum Field Theory used to build the Standard Model are recalled, with phenomenological interpretation. In Section 1.2, the Standard Model is presented. In Section 1.3, jet production and jet clustering are introduced. In Section 1.4, a selected overview of hints at physics Beyond the Standard Model is presented.

1.1 The Quantum Field Theory Framework

This section is meant as a phenomenological overview of the essential building blocks that constitute the theory of the Standard Model (SM), to provide the reader with a second order transition between textbooks and the rest of the material in this thesis.[1] A reader already familiar with Quantum Field Theory (QFT) can skip it altogether, or parts of it.

1.1.1 The need for a quantum theory of relativistic fields

Two big successes of physics of the early XXth century are special relativity and quantum mechanics. They are not just theories in the sense that they are coherent theses that allow to accurately describe some natural phenomena, they are paradigms: frameworks, toolboxes for the expression of theories, that convey visions of our world.

- The special relativity paradigm aims at making theories independent of the choice of space and time reference frame (like the Galilean relativity) while preserving the constraint that there exists a maximum speed. Its key ingredients are tensors, that allow for the intrinsic expression of physical objects while preserving covariance of the components with respect to changes of reference frame, and in general expressions that are invariant under the action of the Poincaré group. Space and time are not independent, they form an object called space-time. The action is expressed using a Lagrangian. Mass is equivalent to energy.
- The quantum paradigm deals with the notions of superposition of states, quantized quantities and the indetermination principle. It uses three key ingredients: wave functions, that are oscillations in fields, delocalized into the entirety of space; operators associated to physical observables, that allow to make measurements; and the Schrödinger equation, that describes the time evolution of a system. Space and time are disconnected here: the Hamiltonian is the time evolution operator, that assumes a time-foliation. Operators do not necessarily commute.

Making those two frameworks compatible was essential to understand some experiments based on electrodynamics, like the photoelectric effect, shown to be well understood by a quantum version of electrodynamics by Einstein in 1905 (this being the kick-off of the quantum theory), the Zeeman effect, the Stern-Gerlach experiments, or properties of matter in general like the structure of the atom and of the nucleus, light-matter interactions, magnetism, superconductivity, superfluidity, radioactivity, LASER, *etc.* But making special relativity and quantum mechanics compatible was everything but obvious and only converged between the 1940s and the 1950s to give birth to Quantum Electro-Dynamics (QED), a quantum theory of the electromagnetic field and electrically charged fermions. Furthermore, it was not clear at all at the time whether nuclear physics could be integrated into the picture. The unification only came in the 1960s and 1970s, with the theory of the strong force (Quantum Chromo-Dynamics, QCD) on the one hand, and the unification of the electromagnetic and weak forces into the electroweak force, along with the Higgs mechanism of spontaneous symmetry breaking.

To exemplify the type of conceptual issues that quickly arise when thinking about special relativity and quantum mechanics at the same time, let's show that changing the reference frame can make a massive object acquire more kinetic energy than the threshold of object creation. Let M be a point-like object of mass m, energy E and momentum $\vec{p}$. According to special relativity, the following relation holds true:

$$E^2 = \vec{p}^2 c^2 + m^2 c^4 \tag{1.1}$$

with $c = 299\ 792\ 458$ m/s being the maximum speed (also called "speed of light in the vacuum"). In the comoving frame A, the object is at rest: $\vec{p_A} = \vec{0}$, and therefore $E_A = mc^2$. Let B be a reference frame moving in uniform

[1] This section is mostly based on the lectures of second year of masters degree at ICFP (ENS Ulm and Sorbonne University, Paris) by Amir-Kian Kashani-Poor, and on [2].

rectilinear translation at the speed $\vec{v} = (v, 0, 0)$ with respect to frame A. In B, the object is not at rest: $\vec{p_B} \neq \vec{0}$, therefore $E_B \neq E_A$. More precisely, $E_B = \gamma E_A = \gamma mc^2$, with $\gamma = \frac{1}{\sqrt{1-\left(\frac{v}{c}\right)^2}}$ the LORENTZ factor. For $\frac{v}{c} \approx 1$, one has $E_B > 2mc$, which is the threshold of object creation (a one-object state could tunnel to a two-object state). Therefore one needs a quantum theory that describes the creation and annihilation of objects (particles in the context of particle physics). The number of particles is not fixed.

Another technical difficulty comes from the LORENTZ invariance: physical laws should be the same in any inertial frame. For instance, $E_A^2 - \vec{p}_A^2 = m^2$ in a frame (using natural units $c = \hbar = 1$), then in another frame where $\vec{p}_B \neq \vec{p}_A$, energy must change to preserve $E_B^2 - \vec{p}_B^2 = m^2$. The LORENTZ group acts on $(E_A, \vec{p}_A)$ to yield $(E_B, \vec{p}_B)$ such that $E_A^2 - \vec{p}_A^2 = E_B^2 - \vec{p}_B^2$. When thinking of this in the context of quantum mechanics, this means that the LORENTZ group must act on the HILBERT space.

Those two concepts are pivotal. The first issue gives a taste of one way to quantize the theory, called the "canonical quantization": with the familiar creation $a^\dagger$ and annihilation a operators of quantum mechanics. Quantization can also be done with path integrals, a different formulation relying more on the wave nature of matter, very analogous to HUYGENS' principle. The second issue exemplifies the central idea of symmetry: the theory is invariant under transforms, independent of the mathematical choice of the numbers used to represent nature, and therefore this points towards the use of covariant objects (tensors) and representations of symmetry groups.

To give an example, the simplest field one can write is a scalar field of mass m:

$$\phi(x) = \int \frac{\mathrm{d}^3 p}{(2\pi)^3} \frac{1}{\sqrt{2\omega_p}} \left(a_p(t) e^{-ip \,\cdot\, x} + a_p^\dagger(t) e^{ip \,\cdot\, x}\right) \tag{1.2}$$

with $x = \left(x^0 = t, x^1, x^2, x^3\right)$, $p = \left(p^0 = \omega_p = \sqrt{m^2 + \vec{p}^2}, p^1, p^2, p^3\right)$, $p \cdot x = \eta_{\mu\nu} p^\mu x^\nu = p^0 x^0 - p^i x^i$ (with EINSTEIN's convention of implicit summation over repeated indexes). The field shows $a_p(t)$ and $a_p^\dagger(t)$, the annihilation and creation operators of a particle (with mass m, momentum p at time t). $\int \frac{\mathrm{d}^3 p}{(2\pi)^3} \frac{1}{\sqrt{2\omega_p}}$, $p \cdot x$ are LORENTZ-invariant. Superposition is also explicit in the integral over the momenta.

1.1.2 Lagrangian formalism and symmetries

The Lagrangian approach of classical mechanics can be adapted to Quantum Field Theory (QFT), following DIRAC's principles of quantization: fields become operators acting on states, POISSON brackets $\{\,\cdot\,,\,\cdot\,\}$ are replaced by commutators $\frac{1}{i}[\,\cdot\,,\,\cdot\,]$, and ordering issues have to be dealt with. The Lagrangian allows to derive the equations of motion satisfied by the fields and to show the couplings of the fields at once; its role is also central due to NOETHER's theorem.

NOETHER's theorem states that if the Lagrangian has a continuous symmetry, then there exists a physical quantity associated with that symmetry that is conserved in time. Metaphorically, it means that constraints in the playing field prevent some movement possibilities for the players such that a quantity never changes; the arena influences the dynamics of the fighters inside it. The conserved quantity is called the NOETHER charge. The mathematical theory of representations of LIE algebras allows to provide objects that naturally respect the symmetries. Representations are therefore the building blocks of fields.

One symmetry group of fundamental interactions, according to experiments, is the group of symmetries of spacetime, *i.e.* invariance by translations, rotations and boosts. More specifically, only the connected component of this group that contains the identity is needed. This group is the "proper orthochronous POINCARÉ group", written

$$\mathcal{P}_0 = \mathbb{R}^4 \rtimes SO^\uparrow(1,3) \tag{1.3}$$

with group multiplication $(a, S) \circ (b, T) = (a + S \times b, S \times T)$. First, this group has ten generators and therefore implies ten conservation laws by NOETHER's theorem: one for energy, three for momentum, three for the angular momentum, three for the velocity of the center of mass. Second, its representations are associated to a quantum number, the spin, that dictates a lot of the properties of the fields. Only fundamental fields with spin 0, $\frac{1}{2}$ and 1 are known (though hadrons with higher spins are known). Fields with integer spin are called bosonic fields, they obey the BOSE-EINSTEIN statistics, while fields of half-integer spin are called fermionic fields and obey the FERMI-DIRAC statistics. Representations of spin 0 are called "scalars", representations of spin $\frac{1}{2}$ are called "spinors", representations of spin 1 are called "vectors". Scalars and vectors (of the SM) obey the KLEIN-GORDON equation of motion, while spinors (of the SM) obey the DIRAC equation of motion.

The second type of symmetry that can be part of the Lagrangian is the gauge groups. A symmetry in a system essentially means that there are multiple mathematical choices to describe exactly the same physical object, a choice that is therefore meaningless. The symmetry being local, any two particles can have different gauge choices, but

when moving from one point to another one, for instance to describe the interaction of two particles, the gauge choice has to be the same for them both. Derivatives in the Lagrangian therefore have to be covariant not only with the choice of reference frame, but also with the choices of gauges. The gauge choice is mediated by the "gauge bosons", which intervenes in the derivatives to make them covariant. Therefore, to each generator of a gauge group corresponds a gauge boson. The gauge symmetries play another important role: they are associated with NOETHER charges that are quantized.

Basically, the first steps for the formulation of a quantum field theory are:

1. The symmetry group structure is defined.
2. The fields are chosen, they are attributed a representation structure of the symmetry group, and quantum numbers. Some fields are gauge fields, some are matter fields.
3. The couplings between the fields are decided, providing the theory with a Lagrangian. Some consistency rules are to be respected: covariance of the derivatives, renormalizability (see later), causality, unitarity.

If some consistency rules are not respected, it does not necessarily mean that the theory is useless or not predictive, only that its domain of applicability is restricted. To make things more concrete, the full symmetry group of the SM Lagrangian is provided in Section 1.2.1, the fields and their representation structure are listed in Section 1.2.2 and Fig. 1.2, and the full SM Lagrangian is given in Section 1.2.3, from which one can derive the couplings of Fig. 1.3.

1.1.3 Perturbative approach

Due to the complexity of the SM Lagrangian (see Section 1.2.3), not many mathematical methods are known to make physical predictions. The main one is the perturbative approach to compute scattering amplitudes, which is particularly useful as it perfectly suits the scattering experiments where a beam of particles collides against another beam (like at the LHC) or against a fixed target. In such an experiment, the scattering probability is given by

$$\mathcal{P}_{i\to f} = \frac{\text{Number of scattered particles of type } f}{\text{Number of incoming particles of type } i} = \frac{\sigma_{i\to f}}{A} \tag{1.4}$$

where A is the area of a cross-section of the beam perpendicular to the beam axis, and $\sigma_{i\to f}$ is the so called "scattering cross-section", that also has units of an area, that is a measure of the interaction strength to produce a given final state f from the initial state i (a more rigorous definition is given in Chapter 5, Eq. (5.2) based on beam luminosity). Quantum mechanically, the scattering probability is given by

$$\mathcal{P}_{i\to f} = \frac{|\langle i|S|f\rangle|^2}{\langle i|i\rangle\,\langle f|f\rangle} \tag{1.5}$$

where S is simply called the "S-matrix", from which trivial factors are extracted by convention:

$$S = 1 + i(2\pi)^4\delta^4\left(\sum_{\text{in}} p_i - \sum_{\text{out}} p_j\right)\mathcal{M} \tag{1.6}$$

A "matrix element" is by definition $\langle f|\mathcal{M}|i\rangle$.[2]

In perturbation theory, the Lagrangian is divided into two parts, the free Lagrangian in which fields do not interact (usually kinematic terms), and the interaction part of the Lagrangian which has to be small:

$$\mathcal{L}(\phi_1,\phi_2,\dots) = \mathcal{L}_0(\phi_1,\phi_2,\dots) + \mathcal{L}_{\text{int}}(\phi_1,\phi_2,\dots) \tag{1.7}$$

The goal being to relate solutions involving free fields (*i.e.* fields which evolution is governed by $\mathcal{L}_0$) to solutions of the full interacting Lagrangian. This is done using multiple steps.

First, the LEHMANN-SYMANZIK-ZIMMERMANN (LSZ) formula relates S-matrix elements for n asymptotic momentum eigenstates to an expression involving the quantum fields $\phi(x)$ (given here for the case of scalar fields):

$${}_{\text{out}}\langle f|S|i\rangle_{\text{in}} = \left[i\int \mathrm{d}^4x_i e^{-ip_1x_1}\left(\Box_1+m^2\right)\right]\cdots\left[i\int \mathrm{d}^4x_n e^{ip_nx_n}\left(\Box_n+m^2\right)\right]{}_{\text{out}}\langle\Omega|T\left\{\phi(x_1)\cdots\phi(x_n)\right\}|\Omega\rangle_{\text{in}} \tag{1.8}$$

with the $-i$ in the exponent applying for initial states and the $+i$ for final states, $\Box_j = \eta^{\mu\nu}\frac{\partial}{\partial x_j^\mu}\frac{\partial}{\partial x_j^\nu}$, $T\{\cdots\}$ the time-ordered product and $|\Omega\rangle$ the ground state (or vacuum) of the interacting theory. Basically, this means that the $\Box + m^2$ operator has to be applied to the time-ordered product of fields, and the result is FOURIER-transformed.

[2]Note that a lot of technicalities are not mentioned here for the sake of simplicity, like the definition of the space of states (FOCK's space), "in and out states", *etc.*

Second, the time-ordered product of two free fields is expressed and called their FEYNMAN propagator:

$$D_F(x_1, x_2) = \langle 0|T\left\{\phi_0(x_1)\phi_0(x_2)\right\}|0\rangle = \lim_{\varepsilon\to 0}\int \frac{\mathrm{d}^4k}{(2\pi)^4}\frac{i}{k^2 - m^2 + i\varepsilon}e^{ik(x_1-x_2)} \tag{1.9}$$

with $|0\rangle$ the vacuum state in the free theory.

Third, the interacting fields are connected to the free fields by using the perturbativity hypothesis:

$$\langle\Omega|T\left\{\phi(x_1)\cdots\phi(x_n)\right\}|\Omega\rangle = \frac{\left\langle 0\middle|T\left\{\phi_0(x_1)\cdots\phi_0(x_n)e^{i\int \mathrm{d}^4x\mathcal{L}_{\mathrm{int}}[\phi_0]}\right\}\middle|0\right\rangle}{\left\langle 0\middle|T\left\{e^{i\int \mathrm{d}^4x\mathcal{L}_{\mathrm{int}}[\phi_0]}\right\}\middle|0\right\rangle} \tag{1.10}$$

where $T\left\{e^{i\int \mathrm{d}^4x\mathcal{L}_{\mathrm{int}}[\phi_0]}\right\}$ is a DYSON series, which is a notation for the perturbative expansion:

$$T\exp\left[i\int \mathrm{d}^4x\mathcal{L}_{\mathrm{int}}[\phi_0]\right] = 1 + i\int \mathrm{d}^4x' T\left\{\mathcal{L}_{\mathrm{int}}[\phi_0](x')\right\} + \cdots + \frac{i^n}{n!}\int \mathrm{d}^4x_1'\cdots\int \mathrm{d}^4x_n' T\left\{\prod_{i=1}^{n}\mathcal{L}_{\mathrm{int}}[\phi_0](x_i')\right\} \tag{1.11}$$

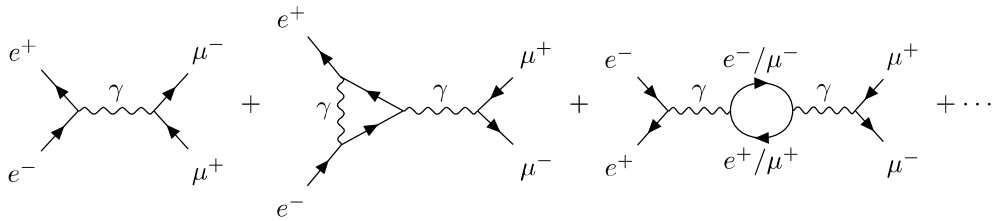
Fourth, the time-ordered product of free fields is related to FEYNMAN propagators with WICK's theorem:

$$T\left\{\phi_0(x_1)\cdots\phi_0(x_n)\right\} = :\phi_0(x_1)\cdots\phi_0(x_n) + \text{all possible contractions}: \tag{1.12}$$

where $:\cdots:$ denotes the normal ordering: all the creation operators $a_p^\dagger$ are on the left of all the annihilation operators a_p, and a contraction means taking two fields $\phi_0(x_i)$ and $\phi_0(x_j)$ from anywhere in the series and replacing them with a factor of $D_F(x_i, x_j)\mathbb{1}$, for each pair of fields (including "one contraction", "two contractions", *etc.*) This theorem is particularly useful because the remaining normal ordered products vanish in the vacuum matrix elements, and all that remains is the FEYNMAN propagators.

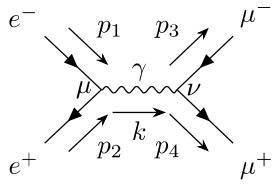
With all those mathematical prowess, S-matrix elements are finally calculable! We are left with products of FEYNMAN propagators, *i.e.* integrals over space-time.

Fortunately, the application of all those theorems can be automatized, and a diagrammatic representation with simple rules exists to find out the terms at different orders in perturbation theory and to provide an intuitive representation of the phenomenon. These are called FEYNMAN diagrams. Here is an example for the process $e^+e^- \to \mu^+\mu^-$ in QED:

$i\mathcal{M} =$ 

(1.13)

with only a few examples drawn (tree level and two second order diagrams). The physical quantity that is measured is $\frac{1}{4}\sum_{\text{spins}}|\mathcal{M}|^2$, which shows that FEYNMAN diagrams interfere between each other, which can also be interpreted as a quantum superposition of multiple processes. Each diagram can be computed independently; for instance, the tree-level one is equal to:

$i\mathcal{M}_1 =$ 

$$= (-ie)^2\bar{v}(p_2)\gamma^\mu u(p_1)\frac{-i\left[\eta_{\mu\nu} - (1-\xi)\frac{k_\mu k_\nu}{k^2}\right]}{k^2 + i\varepsilon}\bar{u}(p_3)\gamma^\nu v_\mu(p_4) \tag{1.14}$$

which, staying at tree-level and after calculation, gives

$$\frac{1}{4}\sum_{\text{spins}}|\mathcal{M}|^2 = 2\frac{e^4}{s^2}\left[u^2 + t^2 + 4s\left(m_e^2 + m_\mu^2\right) - 2\left(m_e^2 + m_\mu^2\right)^2\right] \tag{1.15}$$

with the MANDELSTAM variables $s = (p_1 + p_2)^2$, $t = (p_1 - p_3)^2$ and $u = (p_1 - p_4)^2$.

1.1.4 Renormalization

The example of the FEYNMAN diagram shown above works well, it is calculable without any issue, but it is not so simple in general, especially for diagrams that comprise loops like the two other diagrams. The loops allow for unconstrained momenta, that are off the mass shell (*i.e.* $q^2 \neq m^2$) and therefore the integral must run over all possibilities up to infinity. For instance, the second diagram receives, among other terms, the following contribution from the photon-electron loop:

$$-ie^3 \int \frac{\mathrm{d}^4 q}{(2\pi)^4} \gamma^\mu \frac{i\left(\gamma^\alpha (p_1 - q)_\alpha + m\right)}{(p_1 - q)^2 - m^2 + i\varepsilon} \gamma^\rho \frac{i\left(\gamma^\beta (p_2 - q)_\beta + m\right)}{(p_2 - q)^2 - m^2 + i\varepsilon} \gamma^\nu \frac{-i\eta_{\mu\nu}}{q^2 + i\varepsilon} \tag{1.16}$$

When q is large, it dominates:

$$e^3 \gamma^\mu \gamma^\alpha \gamma^\rho \gamma^\beta \gamma_\mu \int \frac{\mathrm{d}^4 q}{(2\pi)^4} \frac{q_\alpha q_\beta}{(p_1 - q)^2 (p_2 - q)^2 q^2} \sim \int \frac{q^5 \, \mathrm{d}q}{q^6} \tag{1.17}$$

This integral is UV-divergent and infinite. There are two methods to deal with such infinities: regularization and renormalization.[3]

The regularization approach introduces a cutoff, an energy scale above which the integral is not performed, making the result cutoff-dependent, and then takes the limit where the cutoff is sent to infinity. Phenomenologically, it amounts to saying that the theory is incomplete in the UV region, and that integrating in this region makes no sense as we don't know what lies in there. There are multiple techniques. The dimensional regularization of 'T HOOFT and VELTMAN, carries the integral in a space with a fictitious fractional number of dimensions. The PAULI-VILLARS regularization adds fictitious particles to the theory, at very large masses.

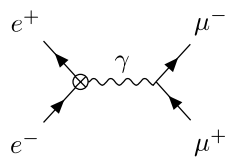
The other approach is renormalization. The diagram discussed above is in fact not measurable by itself experimentally. As long as we are always computing measurable quantities, the answer should come out finite. Other quantities that appear naturally in the theory are in fact not measurable either, like the electric charge and mass of the electron, as well as the normalization of the fields themselves. Those are bare quantities, that do not take into account the contribution of virtual-particle loop effects to the physical constants themselves.

The idea of renormalization is to compute the bare propagators, and then to re-scale the bare quantities (fields, masses, couplings) to obtain physical quantities, such that the infinities cancel out. For instance, in QED, one can parameterize the scaling factors as $e_0 = Z_e e_R$, $m_0 = Z_m m_R$, $\psi^0(x) = \sqrt{Z_2}\psi^R(x)$ and $A^0_\mu = \sqrt{Z_3} A^R_\mu$, with 0 for the bare quantities and R for the renormalized ones. The important point here is that in QED there are only four quantities that can be re-scaled, and it turns out that it is enough to cancel all the possible infinities of QED, even though there are a lot more than four correlation functions that can be computed! There is a choice of the expression for the normalization coefficients, which is called a subtraction scheme choice; this choice should not matter because physical results are independent of the scheme, and because in the end what renormalization does is $\infty - \infty$ regardless of finite additive terms. There are three common choices: the Minimal Subtraction scheme (MS), in which the $\delta_i = Z_i - 1$ have no finite part; the modified Minimal Subtraction scheme ($\overline{\mathrm{MS}}$) that is slightly more convenient for some computations and therefore more used; and the on-shell subtraction scheme that chooses the physical mass m_R as the pole of the total propagator (the one with all orders in perturbation theory), called the "pole mass". Equivalently, the Lagrangian can directly be expressed in terms of renormalized quantities, thus providing extra FEYNMAN rules known as "counter-terms", as can be seen for the QED Lagrangian:

$$\mathcal{L} = -\frac{1}{4}\left(\partial_\mu A^0_\nu - \partial_\nu A^0_\mu\right)^2 + \bar{\psi}^0 \left(i\partial\!\!\!/ - e_0 A\!\!\!/^0 - m_0\right)\psi^0 \tag{1.18}$$

$$= -\frac{1}{4} Z_3 \left(\partial_\mu A^R_\nu - \partial_\nu A^R_\mu\right)^2 + i Z_2 \bar{\psi}_R \partial\!\!\!/ \psi_R - Z_2 Z_m m_R \bar{\psi}_R \psi_R - e_R Z_e Z_2 \sqrt{Z_3} \bar{\psi}_R A\!\!\!/_R \psi_R \tag{1.19}$$

Renormalization is essentially isolating the finite and physical part from the bare quantities, while the remaining part of the bare quantities (the counter-terms) cancel out the infinities arising from loops. For instance, the second diagram in Eq. (1.13) has to be summed with the following one for infinities to cancel out:



(1.20)

[3]Regularization can be used jointly with renormalization: renormalization basically makes $\infty - \infty$, which is not well defined, so regularization can be used to get rid of those infinities too.

Renormalization introduces an unphysical scale (usually expressed in energy or mass), the "renormalization scale" or "subtraction point" μ. The physical predictions of the theory are independent of the choice of the scale, at every order, leading to an important constraint: the renormalization group equation. The scale is usually chosen close to the energies and momenta exchanged in the interaction. This choice essentially affects how much of the result comes from the diagrams without loops and how much comes from the finite part of the diagrams with loops. When changing the renormalization scale, the physical constants effectively change, describing the changes in the behavior or the field theory under the changes in the energies invoked in the interaction. This scale-dependence is encoded in a "β-function" $\beta(g) = \mu \frac{\mathrm{d}g}{\mathrm{d}\mu}$.

In the case of QED, it is expressed as

$$\beta(e) = \mu \frac{\mathrm{d}e}{\mathrm{d}\mu} = \frac{e^3}{12\pi^2} \tag{1.21}$$

This β function being positive, the coupling decreases when the energy decreases, meaning that the theory is free in the infrared. The coupling increases when the energy increases, and even becomes infinite at a finite energy called the LANDAU pole, where perturbation theory for QED breaks (estimated as 10^{286} eV). Phenomenologically, a bare electron is surrounded by a cloud of virtual particles (photons, positrons, *etc*) that interact with the initial electron. The electron-system behaves as if it had a different mass and charge, and is called the "dressed electron". The cloud of virtual particles screens the charge of the bare electrons, making the coupling weaker at large space scales (small energies).

In the case of QCD (with three families of DIRAC fermions in the fundamental representation of $SU(3)$ and no colored scalar field, as in the SM), it is expressed as

$$\beta(g_s) = -\frac{7}{16\pi^2} g_s^3 \tag{1.22}$$

This β function is negative, meaning that it is asymptotically free in the UV, but not at low energies. The solution to the 1-loop renormalization group equation (expressed by the β function, and using $\alpha_s = \frac{g_S^2}{4\pi}$) is

$$\alpha_s(\mu) = \frac{2\pi}{7} \frac{1}{\ln \frac{\mu}{\Lambda_{\mathrm{QCD}}}} \tag{1.23}$$

where Λ_{QCD} is the LANDAU pole of QCD, with a value of $\Lambda_{\mathrm{QCD}} = 332 \pm 17$ MeV (in the $\overline{\mathrm{MS}}$ scheme, at 4 loops in the running of α_s).[3] This explains the phenomenon of "color confinement": color-charged particles (quarks and gluons) cannot be isolated and create bound states called hadrons that are color-neutral. Phenomenologically, hadrons are composed of "valence quarks" that are surrounded, like electrons, by a cloud of virtual particles (quarks, gluons, *etc*) that is called the "sea". Near this energy Λ_{QCD} and below, perturbative theory is not valid for QCD. This "running" of the strong coupling constant α_s is verified experimentally by measuring it at different energies, as shown in Fig. 1.1.

1.2 The Standard Model of Particle Physics

Based on the elements of Quantum Field Theory (QFT) presented in the previous section, the Standard Model (SM) of particle physics can be presented.[4]

1.2.1 Symmetries of the Standard Model Lagrangian

The SM Lagrangian possesses a few types of symmetries: some postulated for its construction, some "accidental". The fields are representations of the postulated symmetries. Each symmetry is associated to a conserved charge by NOETHER's theorem, see Section 1.1.2.

The proper orthochronous POINCARÉ group $\mathcal{P}_0 = \mathbb{R}^4 \rtimes SO^{\uparrow}(1,3)$, already discussed, is a global symmetry that expresses invariance under translations, rotations and LORENTZ boosts. It leads to the conservation of energy, momentum, angular momentum, velocity of the center of mass.

The gauge group $SU(3)_C \times SU(2)_L \times U(1)_Y$ of local symmetries leads to the conservation of color charge C, weak isospin Y and weak hypercharge Y_W respectively. The $SU(2)_L \times U(1)_Y$ is spontaneously broken by the HIGGS mechanism: the HIGGS field, which is an $SU(2)$ doublet $\phi = \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}$ acquires a vacuum expectation value through spontaneous symmetry breaking: $\langle \phi \rangle = \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} \end{pmatrix} \neq 0$ (in unitary gauge). The remaining symmetry is $U(1)_{\mathrm{QED}}$, with

[4]This section is mostly based on the lectures of second year of masters degree at ICFP (ENS Ulm and Sorbonne University, Paris) by Stéphane LAVIGNAC, and on [2].

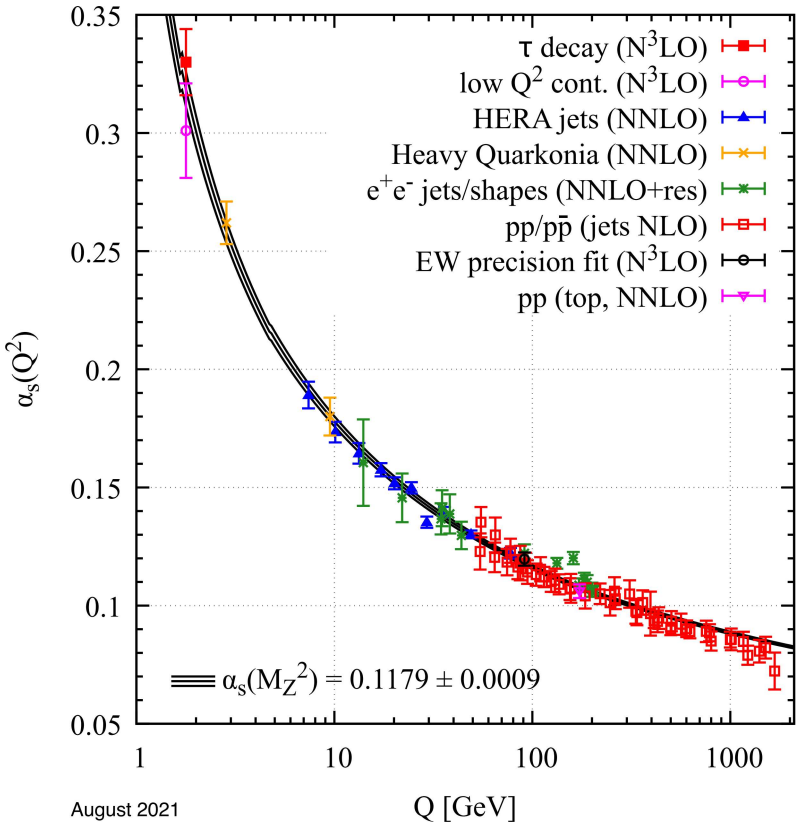


Figure 1.1: Summary of measurements of α_S as a function of the energy scale Q. Reproduced from [3].

the electric charge Q being the associated conserved charge, defined as $Q = Y_3 + \frac{1}{2}Y_W$, with Y_3 the third component of Y. Therefore the remaining gauge symmetry of the SM is $SU(3)_C \times U(1)_{\text{QED}}$.

The SM Lagrangian shown in Section 1.2.3 is also accidentally invariant under a few more global $U(1)$ symmetries: the baryon phase, leading to the conservation of the baryon number $B = \frac{1}{3}(n_q - n_{\bar{q}})$, with n_q the number of quarks and $n_{\bar{q}}$ the number of antiquarks; and three lepton phases, one for each lepton family, leading to the conservation of the three associated lepton numbers L_e, L_μ and L_τ, with $L_\ell = n_\ell - n_{\bar{\ell}}$.

1.2.2 Field content of the Standard Model

For the Lagrangian to be expressed, the field content of the SM has to be given a representation structure and quantum charges.

Bosons

Bosons are integer-spin fields. Gauge fields are necessarily in the adjoint representation of the associated gauge group: B is the gauge field associated to $U(1)$, W^1, W^2 and W^3 are the gauge fields associated to $SU(2)_L$ and G^α with $\alpha \in [\![1, 8]\!]$ are the gauge fields associated to $SU(3)$. Gauge fields have spin 1. Gluons carry two color charges.

The Higgs boson is an $SU(2)_L$ doublet of complex scalar fields, transforming in the fundamental representation of $SU(2)_L$, with $Y_W = \frac{1}{2}$. In the SM Lagrangian, all fields are massless, some of which acquire a mass dynamically through the Higgs mechanism with the electro-weak spontaneous symmetry breaking. The three Goldstone bosons associated to the symmetry breaking, which are three of the four real fields of ϕ, are "eaten" by the weak bosons and combine with $W^{1,2,3}$ to give the weak bosons W^+, W^- and Z^0, and represent their longitudinal polarization, thus making these bosons massive. W^3 and B also combine to give the photon γ, the gauge boson of the remaining $U(1)_{\text{QED}}$ symmetry, that remains massless. The fourth real scalar field of ϕ is the Higgs boson h, which has spin 0, no electric charge and no color charge.

Fermions

All fermions of the SM have spin $\frac{1}{2}$, and they come in three families that have the exact same properties except for their masses. They all come with their anti-matter counterpart, obtained by application of the CP operator, and are Dirac spinors, except maybe for neutrinos that could be Majorana fermions and be their own anti-particles. All

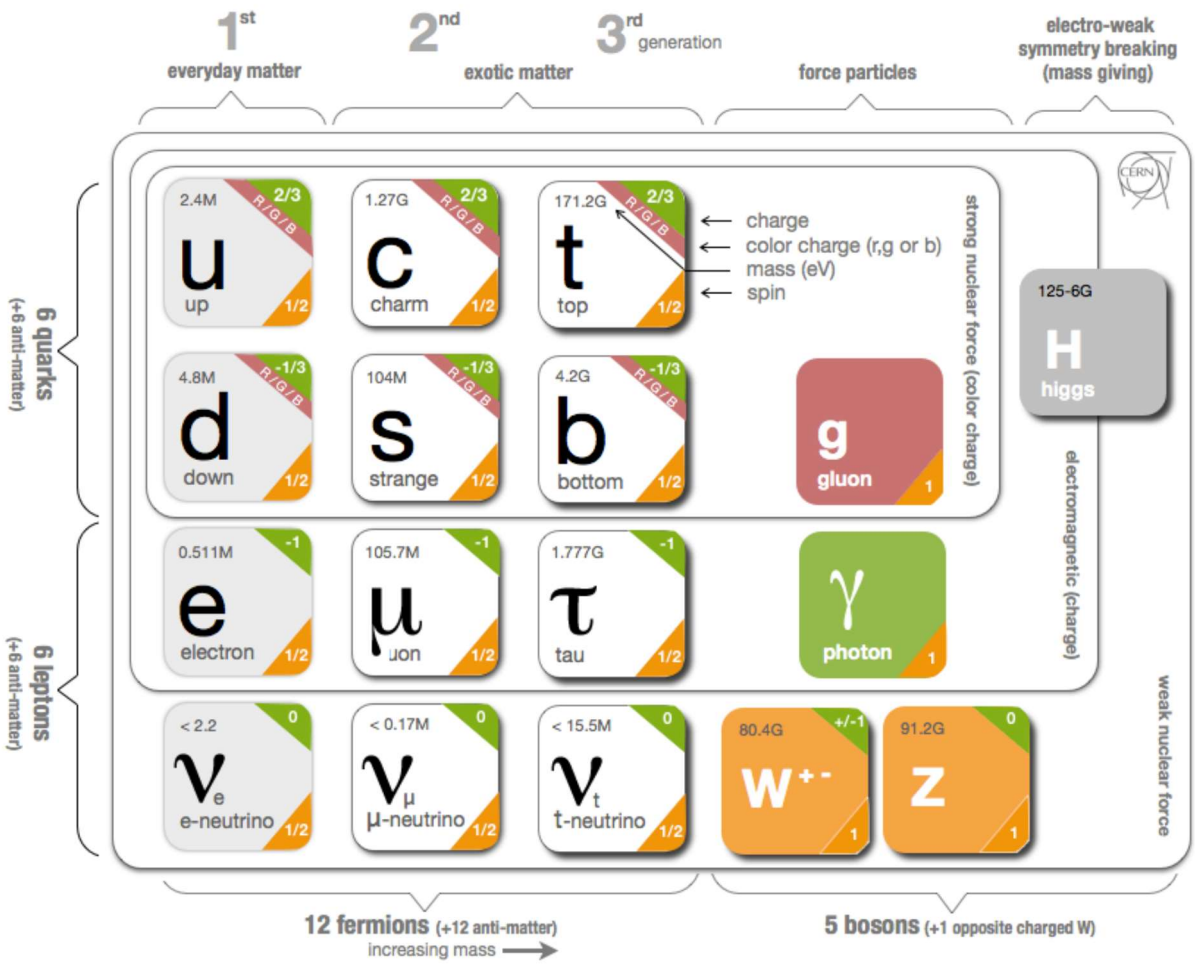


Figure 1.2: Summary of particles composing the SM and their properties (spin, mass, electric charge, color charge). The boxes indicate which fermions can interact with the given gauge boson. Reproduced from [4].

fermions interact with the HIGGS field (through the YUKAWA couplings described hereafter), and therefore acquire a mass through electro-weak symmetry breaking.

Fermions are divided into two categories: quarks, that are charged under $SU(3)_C$, and leptons that are not. Quarks have a color charge, and transform in the fundamental representation of $SU(3)_C$, as well as a fractional electric charge $\frac{2}{3}$ for the up u, charm c and top t quarks, and $-\frac{1}{3}$ for the down d, strange s and bottom b quarks. Leptons have integer electric charges: -1 for the electron e, muon μ and tau τ, and 0 for the associated neutrinos, the electron neutrino ν_e, the muon neutrino ν_μ and the tau neutrino ν_τ.

Fermions come in two chiralities: left-handed and right-handed ones. The left-handed fermions combine into an $SU(2)$ doublet that transforms in the fundamental representation of $SU(2)$, while right-handed fermions do not combine. This leaves the following fields: $Q = \begin{pmatrix} u_L \\ d_L \end{pmatrix}$ with $Y_W = \frac{1}{6}$, u_R with $Y_W = \frac{2}{3}$, d_R with $Y_W = -\frac{1}{3}$, $L = \begin{pmatrix} e_L \\ \nu_{eL} \end{pmatrix}$ with $Y_W = -\frac{1}{2}$, e_R with $Y_W = -1$, ν_{eR} (if right-handed neutrinos exist) with $Y_W = 0$, and all the other fields obtained from the two other generations and the anti-matter counterparts.

The zoo of fundamental particles

This makes a total of 61 fields in the SM (assuming the existence of right-handed neutrinos): 2 quarks and 2 leptons per generation, multiplied by 3 for the number of generations, multiplied by 2 for the chirality and again by 2 for the antimatter content, plus the photon γ, 8 gluons, the W^+, W^- and Z^0 weak bosons, and the HIGGS field h. Their properties are summarized in Fig. 1.2.

1.2.3 The Lagrangian of the Standard Model

The SM Lagrangian is split into multiple parts:

$$\mathcal{L} = \mathcal{L}_{\text{gauge}} + \mathcal{L}_{\text{QCD}} + \mathcal{L}_{\text{EW}} + \mathcal{L}_{\text{Higgs}} + \mathcal{L}_{\text{Yukawa}} \tag{1.24}$$

The strong coupling constant associated to $SU(3)_C$ is g_S, the coupling associated to $SU(2)_L$ is g and the coupling associated to $U(1)_Y$ is g'.

The gauge part is:

$$\mathcal{L}_{\text{gauge}} = -\frac{1}{4}\sum_a G^a_{\mu\nu}G^{a\mu\nu} - \frac{1}{4}B^{\mu\nu}B_{\mu\nu} - \frac{1}{4}\sum_i W^{i\mu\nu}W^i_{\mu\nu} \tag{1.25}$$

with the field strengths:

$$G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu + g_S f^{abc} G^b_\mu G^c_\nu \tag{1.26}$$

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu \tag{1.27}$$

$$W^i_{\mu\nu} = \partial_\mu W^i_\nu - \partial_\nu W^i_\mu + g\varepsilon^{ijk} W^j_\mu W^k_\nu \tag{1.28}$$

with f^{abc} the structure constants of $SU(3)$ and ε^{ijk} the fully antisymmetric tensor.

The coupling term in the strong sector is:

$$\mathcal{L}_{\text{QCD}} = \sum_{q\in\text{quarks}} \bar{q} i\gamma^\mu D_\mu q \tag{1.29}$$

with the covariant derivative

$$D_\mu q = \left(\partial_\mu - i g_S G^a_\mu T^a\right) q \tag{1.30}$$

with T^a the generators of $SU(3)$.

The coupling term in the electro-weak sector is:

$$\mathcal{L}_{\text{EW}} = \sum_{\psi\in\text{fermion representations}} \bar{\psi} i\gamma^\mu D_\mu \psi \tag{1.31}$$

with the covariant derivatives

$$D_\mu\psi = \left(\partial_\mu - igW^i_\mu \frac{\sigma^i}{2} - ig' Y_{W,\psi} B_\mu \mathbb{1}_2\right)\psi \qquad \text{for } SU(2) \text{ doublets} \tag{1.32}$$

$$D_\mu\psi = \left(\partial_\mu - ig' Y_{W,\psi} B_\mu\right) \qquad \text{for } SU(2) \text{ singlets} \tag{1.33}$$

with σ^i the generators of $SU(2)$ and $\mathbb{1}_2$ the 2×2 identity matrix.

The Higgs sector is:

$$\mathcal{L}_{\text{Higgs}} = \left(D_\mu\phi\right)^\dagger \left(D^\mu\phi\right) - \left(\kappa\phi^\dagger\phi + \lambda\left(\phi^\dagger\phi\right)^2\right) \tag{1.34}$$

with the covariant derivative

$$D_\mu\phi = \partial_\mu\phi + igW^i_\mu\frac{\sigma^i}{2}\phi - i\frac{g'}{2}B_\mu\phi \tag{1.35}$$

The Yukawa sector is:

$$\mathcal{L}_{\text{Yukawa}} = -y_u\bar{Q}\tilde{\phi}u_R - y_d\bar{Q}\phi d_R - y_e\bar{L}\phi e_R + h.c. \tag{1.36}$$

with $\tilde{\phi} = i\sigma^2\phi^*$, σ^2 the second Pauli matrix and $h.c.$ the Hermitian conjugate.

A rearrangement of the terms after electro-weak symmetry breaking is necessary to make the electro-weak bosons appear clearly. After calculation, it provides all the Feynman diagram vertices shown in Fig. 1.3.

The flavor (or "interaction") eigenstates of the weak interaction are different from the mass eigenstates. Matrices allow to move from the mass eigenstates to the flavor eigenstates: the Cabibbo-Kobayashi-Maskawa (CKM) matrix for quarks, and the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix for neutrinos. For instance with quarks (with apostrophes for the flavor states), it writes

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \tag{1.37}$$

These 3×3 matrices are strongly constrained by unitarity, leaving 9 degrees of freedom, 5 of which can be absorbed as phases of the fermion fields, leaving 4 parameters. Three parameters describe the linear mixing in terms of

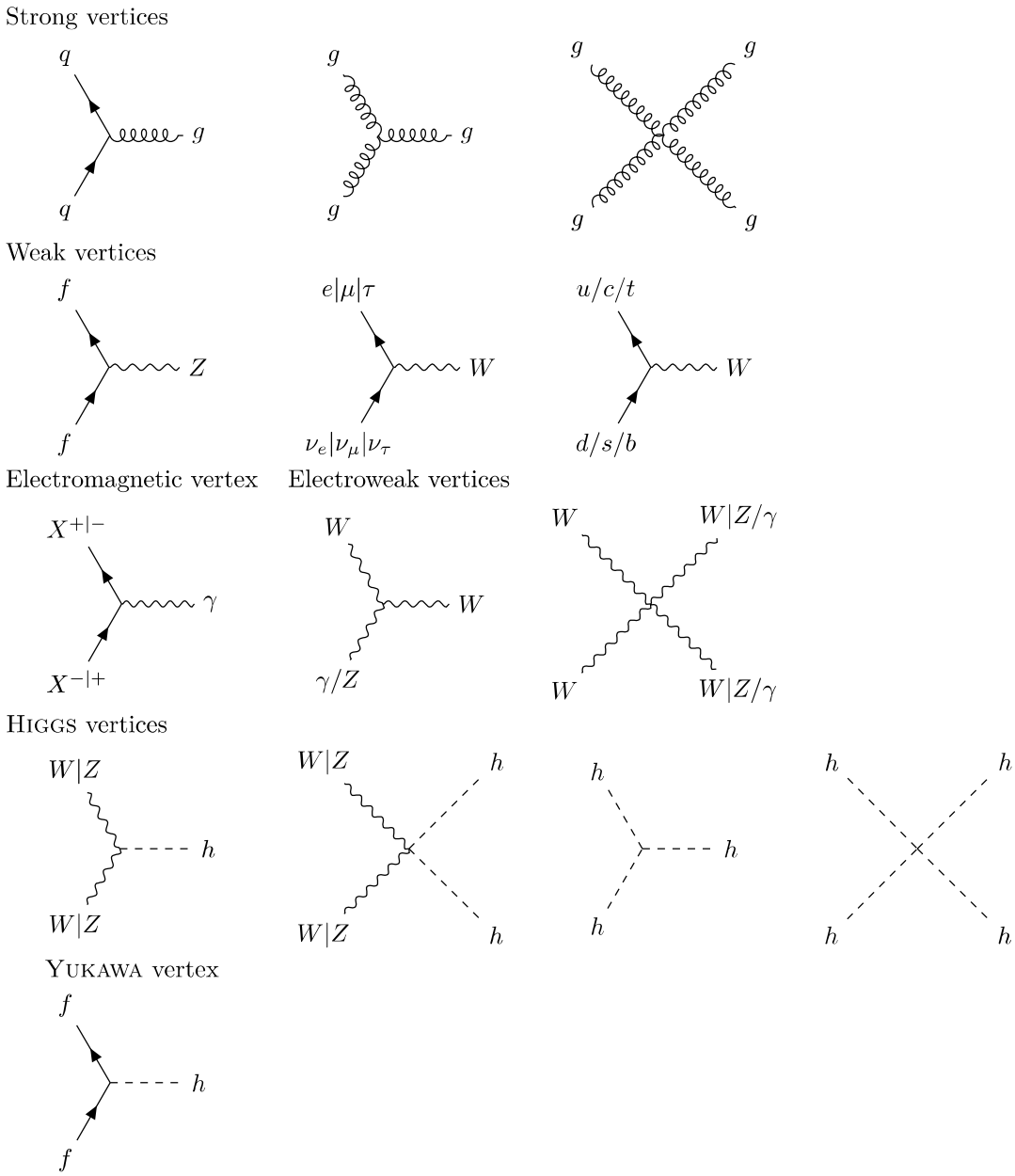


Figure 1.3: All possible interaction vertices for Feynman diagrams in the SM. q is any quark, g is a gluon, X is any charged fermion, γ is a photon, f is any fermion.

rotations, θ_{12}, θ_{23} and θ_{13}, and one phase δ describes the amount of CP violation that the mixing induces. With the cosines and sines of the angles denoted c_{jk} and s_{jk} respectively, it writes

$$V = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & 0 & s_{13}e^{i\delta} \\ 0 & 1 & 0 \\ -s_{13}e^{i\delta} & 0 & c_{13} \end{pmatrix} \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{1.38}$$

1.2.4 The zoo of hadrons

Free quarks or gluons do not exist in nature at low energy, because of color confinement, see Section 1.1.4. They form bound states called "hadrons": "mesons" are composed of two quarks, and "baryons" are composed of three quarks. For example for baryons, a color singlet is formed from three colored quarks using the totally antisymmetric tensor as $B = \varepsilon_{ijk} q^i q^j q^k$, with $q^i \in \{u, d, c, s, b\}$ (because t quarks decay through weak interaction before they can hadronize).

When neglecting the masses of the quarks, the QCD Lagrangian exhibits an extra $SU(N_f)_L \times SU(N_f)_R$ symmetry, that is only approximate since in reality quarks have a mass. This approximation is decent for the u, d and s quarks, since their masses are orders of magnitude lower than the masses of the other quarks, so there is an approximate $SU(3)_L \times SU(3)_R$. This symmetry is useful to classify the bound states that involve u, d and s quarks, and to predict their properties; historically, it even predicted the existence of the Ω^- baryon that was then discovered later. Due to the QCD vacuum containing non-vanishing condensates, this $SU(3)_L \times SU(3)_R$ symmetry is broken to a remaining $SU(3)$ symmetry. This symmetry says that a rotation in flavor space brings a hadron state to a superposition of hadronic states. The space of superpositions of hadronic states is a vector space, which basis corresponds to different particle species. To understand how mesons and baryons transform under this $SU(3)$ symmetry, representation theory is needed. For mesons, a state $M = \varepsilon_{ij} q^i q^j$, with each quark in the fundamental representation $\underset{\sim}{3}$ of $SU(3)$, transforms under the tensor product of the representations: $\underset{\sim}{3} \otimes \underset{\sim}{3} = \underset{\sim}{8} \oplus \underset{\sim}{1}$, and therefore there is a meson octet and a meson singlet. For baryons, they transform under $\underset{\sim}{3} \otimes \underset{\sim}{3} \otimes \underset{\sim}{3} = \underset{\sim}{10} \oplus \underset{\sim}{8} \oplus \underset{\sim}{8} \oplus \underset{\sim}{1}$ and therefore there is a baryon decuplet, a baryon octet and a baryon singlet. This is known as the Gell-Mann-Lévy non-linear σ model, or the "eightfold way". This model also allows to know the quark content of the different particles. For instance the three pions are $\pi^+ = |u\bar{d}\rangle$, $\pi^0 = \frac{1}{\sqrt{2}}\left(|u\bar{u}\rangle - |d\bar{d}\rangle\right)$ and $\pi^- = |d\bar{u}\rangle$. The quark content of the proton and the neutron is respectively $p = |uud\rangle$ and $n = |udd\rangle$. When extended to c quarks, the symmetry is badly broken due to this quark being much heavier, but the classification is still useful, and is shown in Fig. 1.4a for mesons and in Fig. 1.4b for baryons. In these plots, the "eightfold" way is seen in the plane with charm number $C = 0$.

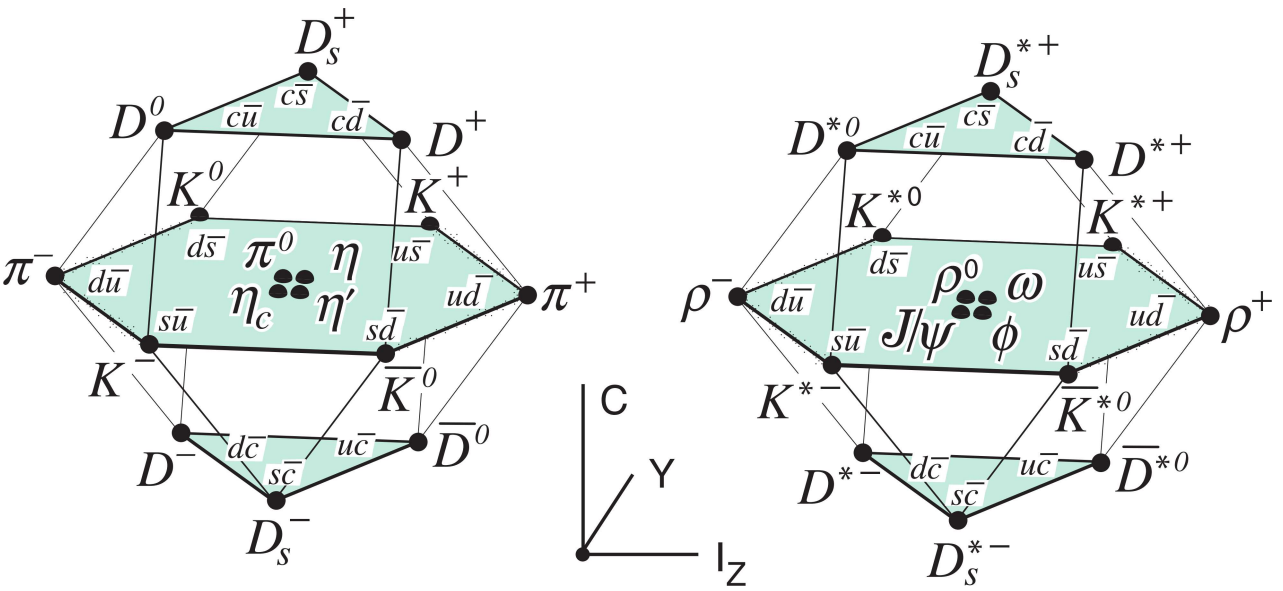
Hadrons can also contain b quarks, making a lot of new possibilities. There can be resonances with the same quark content at higher masses. There can also be composite states involving more than three quarks: four quarks for tetraquarks, five quarks for pentaquarks; in theory, even more is possible though not observed yet. Hybrid mesons are also a possibility, with real gluons, like $|qqg\rangle$. Finally, bound states only involving gluons are also predicted and called "glueballs", though not observed yet for sure, like $|gg\rangle$ or $|ggg\rangle$. This makes a plethora of possible composite particles, listed in [3].

Not all those particles are stable, actually most of them decay to other particles sooner or later. For instance one particle that is important for jets is the neutral pion π^0, that decays to two photons 98.8 % of the time. The proton is stable (and the free neutron is not).

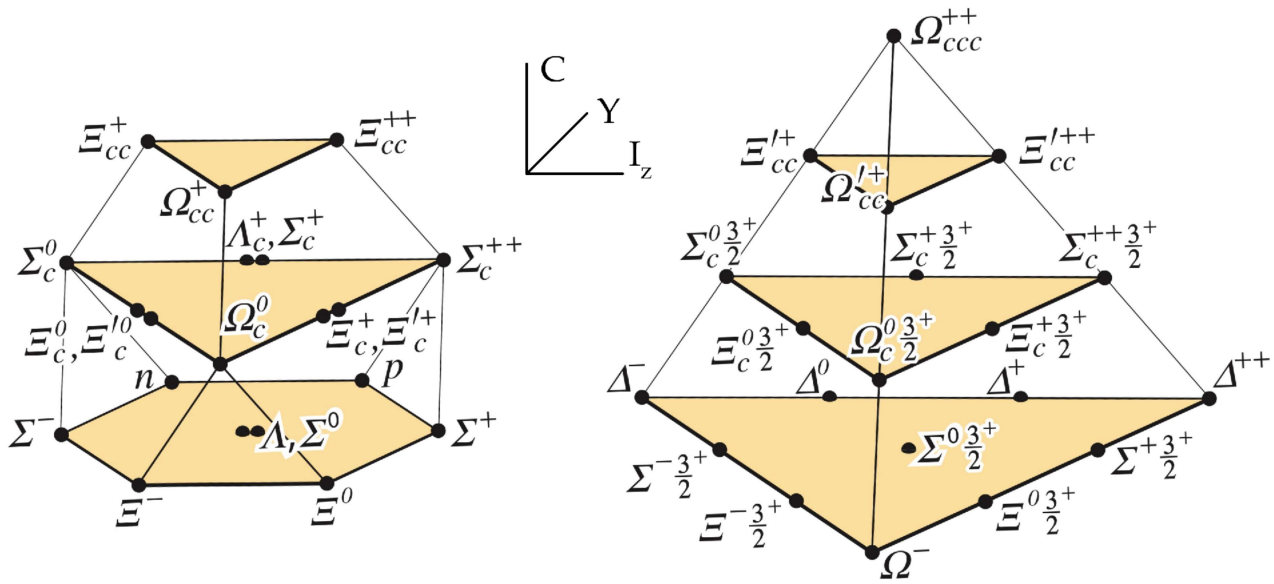
1.2.5 The partonic model

Quarks and gluons associate to form hadrons of different kinds, due to color confinement: for instance, a proton is made of two up quarks and one down quark, as mentioned. However, when probing the structure of the proton, for instance by the scattering of an electron onto the constituents of the proton (an experiment called the Deep Inelastic Scattering, or DIS), one realizes that the proton is not just composed of three quarks, the electron probe can scatter onto a strange quark for instance, especially with highly energetic probes. The three quarks which quantum numbers add up to make those of the proton in the σ-model are called the "valence quarks", but the content of the hadron is a lot richer, and cannot be predicted from perturbative QCD. The other quarks and gluons constitute the quark and gluon "sea". Figure 1.5a gives an intuitive representation of the content of the proton: gluons and other quarks than the valence quark can form through loops. A probe can therefore scatter not only on the valence quarks represented in Fig. 1.5b but also on the quarks from the sea as in Fig. 1.5c, and on sea gluons as well provided that the probe has a color charge.

Feynman proposed the "partonic" model, in which a hadron is composed of "partons" that are essentially free, with "Parton Distribution Functions" (PDFs) that describe the partonic content of the hadron (valence quarks, gluons, higher mass quarks, photons, and any particle of the standard model even though the following discussion will be restricted to quarks and gluons), where each parton carries a fraction ξ of the hadron's momentum, this



(a)



(b)

Figure 1.4: The extended "eightfold" classification of hadrons composed of u, d, s and c quarks, organized using their isospin I_z, charm C and hypercharge Y. **(a)** Classification of the spin-0 and spin-1 mesons. The nonets of light mesons (u, d, s) occupy the central planes. **(b)** Classification of the spin-$\frac{1}{2}$ and spin-$\frac{3}{2}$ baryons. The nonet and decuplet of light mesons occupy the bottom planes. Reproduced from [3].

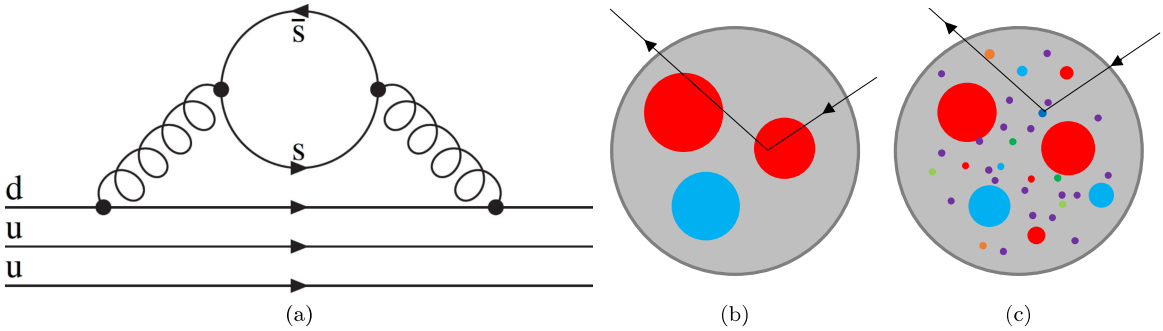


Figure 1.5: **(a)** Illustration of the content of a proton: the valence quarks can radiate a gluon that itself can split into a $q\bar{q}$; thus the partonic content of the proton is not just the valence quarks. **(b,c)** Representation of a DIS probe (electron for instance) that scatters onto **(b)** one of the valence quarks of the proton or **(c)** one of the sea quarks.

fraction not being directly measurable. When an interaction occurs between a hadron and any high-energy point-like particle, the energy scale of the hard scatter process Q is large compared to Λ_{QCD}: the timescale of the interaction is then $\sim \frac{1}{Q} \ll \frac{1}{\Lambda_{\text{QCD}}}$. The incoming particle therefore probes the structure of the hadron before it has time to fluctuate into another state, therefore the partons' wavefunctions are treated as decoherent, which allows for the treatment of PDFs as classical probabilities and leads to the factorization of the hard-scatter process H (calculable in perturbation theory) and the hadron structure f: $\sigma = f \otimes H$. For instance, in the case of DIS, the cross-section for $e^- P^+ \to e^- X$ is given by the cross-sections of $e^- p_i \to e^- X$, where p_i is a parton with momentum $p_i^\mu = \xi P^\mu$ (P^μ being the momentum of the proton), integrated over ξ and summed over i:

$$\sigma(e^- P^+ \to e^- X) = \sum_i \int_0^1 \mathrm{d}\xi f_i(\xi) \sigma(e^- p_i \to e^- X) \tag{1.39}$$

Note that this factorization is only proven in some specific regimes in DIS (see [2] for the proof using operator product expansions). Still, PDFs are used in all particle collisions as they are shown to agree very well with experimental data.

The PDFs are not really probability density functions, as they are not required to sum to 1, their sum rules come from conservation laws: conservation of quantum numbers and momentum. For instance, the content of the proton in terms of down quarks fluctuates due to virtual $d\bar{d}$ pairs, but due to the conservation of the down-quark number (in QED and QCD), we have

$$\int \mathrm{d}\xi \left[f_d(\xi) - f_{\bar{d}}(\xi) \right] = 1 \tag{1.40}$$

where f_d and $f_{\bar{d}}$ are the PDFs for the content in down quark and antiquark respectively in the proton. Similarly, because the proton has up-quark number of 2 and strange quark number of 0:

$$\int \mathrm{d}\xi \left[f_u(\xi) - f_{\bar{u}}(\xi) \right] = 2 \tag{1.41}$$

$$\int \mathrm{d}\xi \left[f_s(\xi) - f_{\bar{s}}(\xi) \right] = 0 \tag{1.42}$$

The PDFs for the charm and bottom quark numbers follow the rule of that of the strange quark. There is no gluon quantum number, so there is no such sum rule for gluons. Finally, the PDFs have to obey the conservation of momentum:

$$\sum_i \int \mathrm{d}\xi \left[\xi f_i(\xi) \right] = 1 \tag{1.43}$$

In the computation of any process involving PDFs at NLO, infra-red poles appear, that go away with renormalization of the PDFs and the introduction of a physical scale: μ_F, the factorization scale. This is further motivated by the measured PDFs that are observed to have a slight (logarithmic) dependency on Q^2. The running of the PDFs is given by the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) set of equations of the form:

$$\mu_F \frac{\mathrm{d}}{\mathrm{d}\mu_F} \begin{pmatrix} f_i(x, \mu_F) \\ f_g(x, \mu_F) \end{pmatrix} = \sum_j \frac{\alpha_S}{\pi} \int_x^1 \frac{\mathrm{d}\xi}{\xi} \begin{pmatrix} P_{q_i q_j}\left(\frac{x}{\xi}\right) & P_{q_i g}\left(\frac{x}{\xi}\right) \\ P_{g q_j}\left(\frac{x}{\xi}\right) & P_{gg}\left(\frac{x}{\xi}\right) \end{pmatrix} \begin{pmatrix} f_i(\xi, \mu_F) \\ f_g(\xi, \mu_F) \end{pmatrix} \tag{1.44}$$

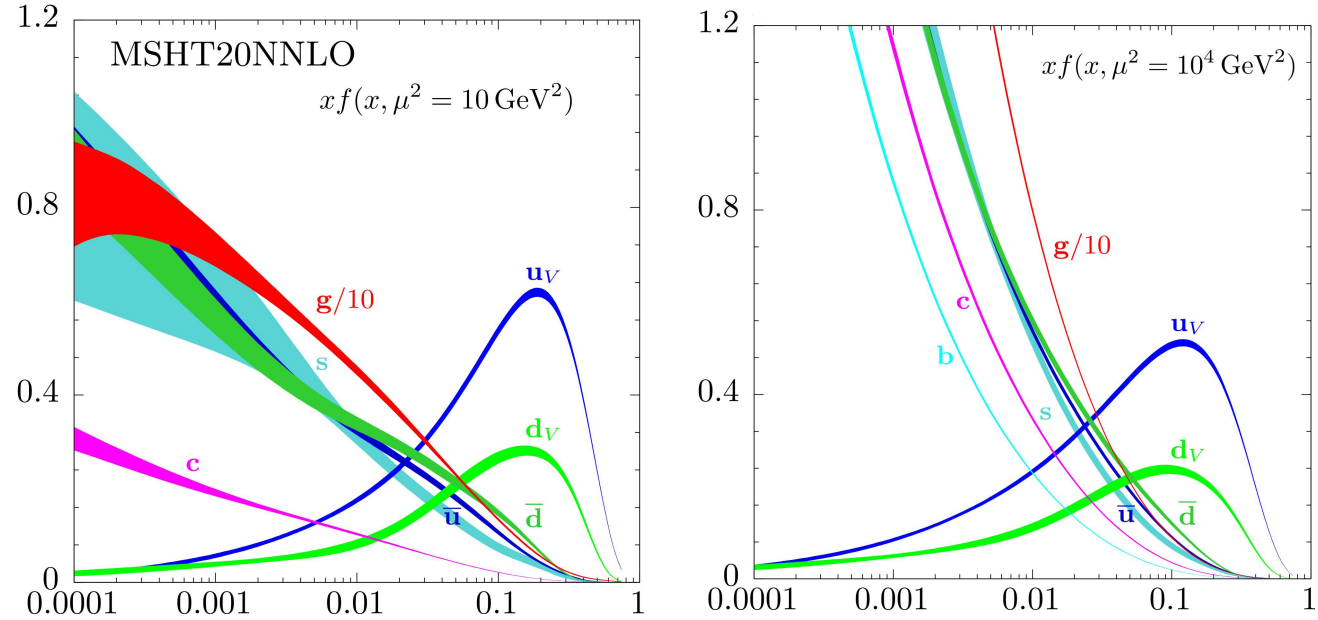


Figure 1.6: $xf(x, \mu_F^2)$ as a function of x, with f_i the Parton Distribution Functions (PDFs), for two values of the factorization scale μ_F: **(left)** $\mu_\mathrm{F}^2 = 10\ \mathrm{GeV}^2$ and **(right)** $\mu_\mathrm{F}^2 = 10^4\ \mathrm{GeV}^2$. Reproduced from [3].

with P_{XY} the DGLAP "splitting functions", which leading order expression can be found in [2] for instance, that describe the probability for a given parton Y to split into X and a gluon.

This running of the PDFs on μ_F means that probing the hadron's content at a certain time or length scale lets the fast or short-distance fluctuations of the sea of quarks and gluons, that happen on lower time or length scales, average out. The PDFs are measured at a certain energy scale, then the running allows to predict their form at all other energy scales. They are shown in Fig. 1.6 for two values of the factorization scale.

Numerically, it turns out that $\int \mathrm{d}\xi\xi\,(f_u(\xi) + f_d(\xi)) \approx 0.38$. Thus, only around 38 % of the proton momentum is contained in the up and down valence quarks, whereas the gluon content of the proton ranges from 35 % to 50 % depending on the factorization scale, the remainder of the proton's momentum being in the sea quarks.

1.3 Jets

In high-energy particle collisions, many different particles can be produced, as described in Sections 1.2.2 and 1.2.4, whether they are fundamental or composite particles: photons, electrons, muons, weak bosons, quarks, gluons, all kinds of hadrons. Some of these particles reach the detector, like photons, electrons or muons, but most of the rest don't and decay into other particles, either because they are short-lived or due to color-confinement. This is the case for quarks and gluons: they radiate other quarks and gluons, a process called showering, before combining into hadrons, a non-perturbative process called hadronization. Since the initial quarks and gluons are boosted objects, their decay products are collimated and thus emitted in a cone in (y, ϕ) space. This provides a natural way to cluster the particles into objects called "jets", from which properties of the parton that originated these particles can be retrieved, such as its kinematics or its quantum numbers (like its flavor). Jets are one of the most basic observables in QCD, providing portals both to perturbative and non-perturbative QCD, for instance through energy flows inside jets and their substructure, correlations inside or between jets, hadronic decays, production cross-sections, *etc.* Jets are also crucial for the partonic structure of decays of massive particles, such as t quarks, τ leptons, W, Z and H bosons. At high-energy hadron colliders such as the LHC, jets have the dominant production cross-section by several orders of magnitude compared to any other final-state observable, as shown in Fig. 1.7, making jets an essential source of signal or noise for most analyses.

Jets have the difficult but essential task of bringing together theory, phenomenology and experiment by providing a common definition for partons at fixed order in perturbation theory, clusters of stable particles, and clusters of energy deposits in the detector. There is no universal method for the clustering of hadrons: for instance, when a quark radiates a gluon, for what range of kinematics should the gluon (or its decay products) be part of the quark jet? Should the clustering sequence start from low-energy or high-energy particles? From shallow angles or wide angles? A choice has to be made depending on the physics that is studied. The clustering algorithms to produce jets are called "jet algorithms", and multiple of them exist for different uses: studying the kinematics or the quantum

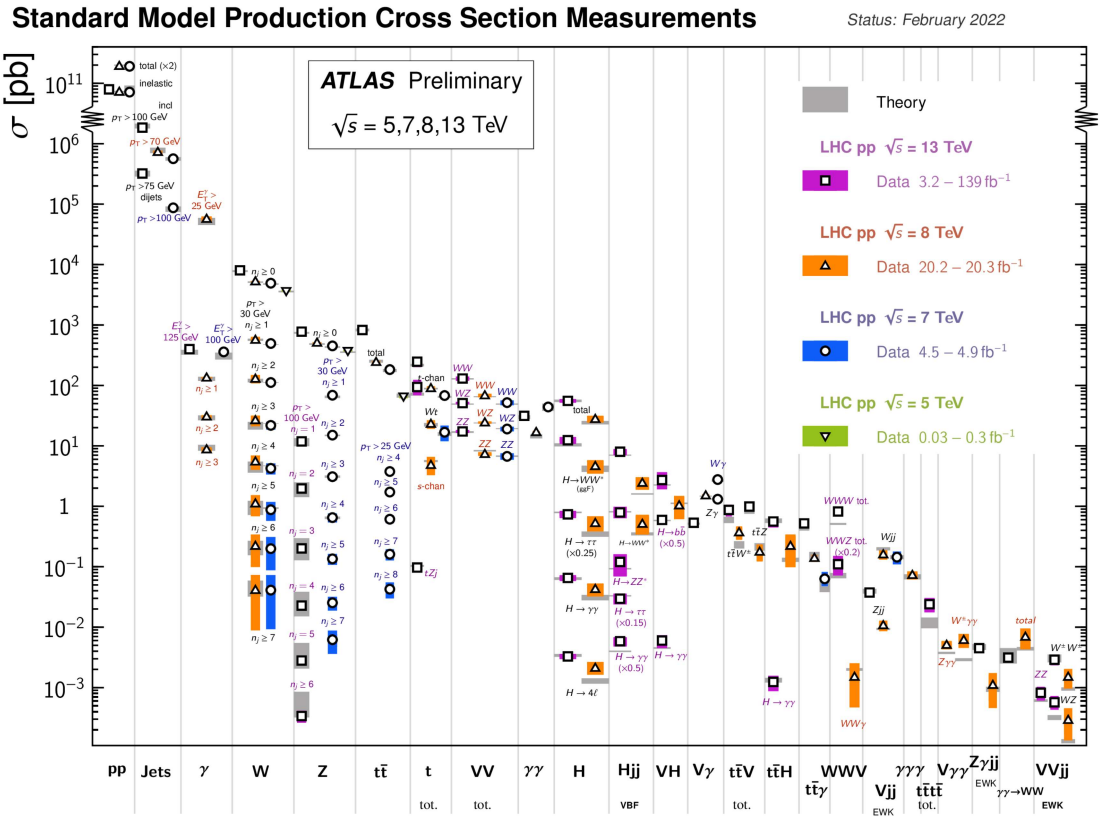


Figure 1.7: Summary of several Standard Model total production cross-section measurements. Reproduced from [5].

numbers of the original partons from the hard-scatter subprocess or from the showering, studying the substructure and energy flows, *etc.*

The current understanding of hadron formation through perturbative and non-perturbative QCD will be discussed in Section 1.3.1, followed by a description of the jet algorithms in Section 1.3.2.

1.3.1 Hadron production, and event generation in simulations

The modern understanding of the process that leads from a typical high-energy proton-proton (pp) collision at LHC and its hard parton interaction to a collection of particles that reach the detector is depicted in Fig. 1.8.[5]

Currently, the full description of pp collisions cannot be performed from first-principles alone:

- The description of the content of the proton requires the parton model along with the Parton Density Functions (PDFs), that are not proven to factorize from the matrix element, that require a physical scale μ_F and that are evaluated based on experimental data;
- The matrix element is computed only to a fixed order in perturbation theory, often at LO or NLO, the state of the art rarely exceeding N^{2}LO. This only works if perturbative QCD is applicable in the first place, *i.e.* if α_S is small, *i.e.* if the momentum transfer is large enough;
- Non-perturbative effects, that currently cannot be treated mathematically, play a crucial role in going from colored partons to hadrons (color-neutral bound states).

Yet, we are not completely ignorant of the processes that play a role, and a variety of QCD-inspired semi-classical models are used to describe the production of hadrons from a pp collision, that are in good agreement with experimental data. The main ingredient being a (hypothetical) factorization of the "hardness" scales (whether it be energy, virtuality, length, time, angles): each scale is resolved by a dedicated model and hands the resulting objects to the model of the next scale. This can be understood as successive wave-function collapses, with loss of quantum coherence between the objects at each stage. Those models need to be tweaked to match data: an event-generator such as PYTHIA uses $\mathcal{O}(100)$ parameters [7].

This section describes the phenomena in decreasing degrees of hardness.

[5] This section is mostly based on [2, 3, 6].

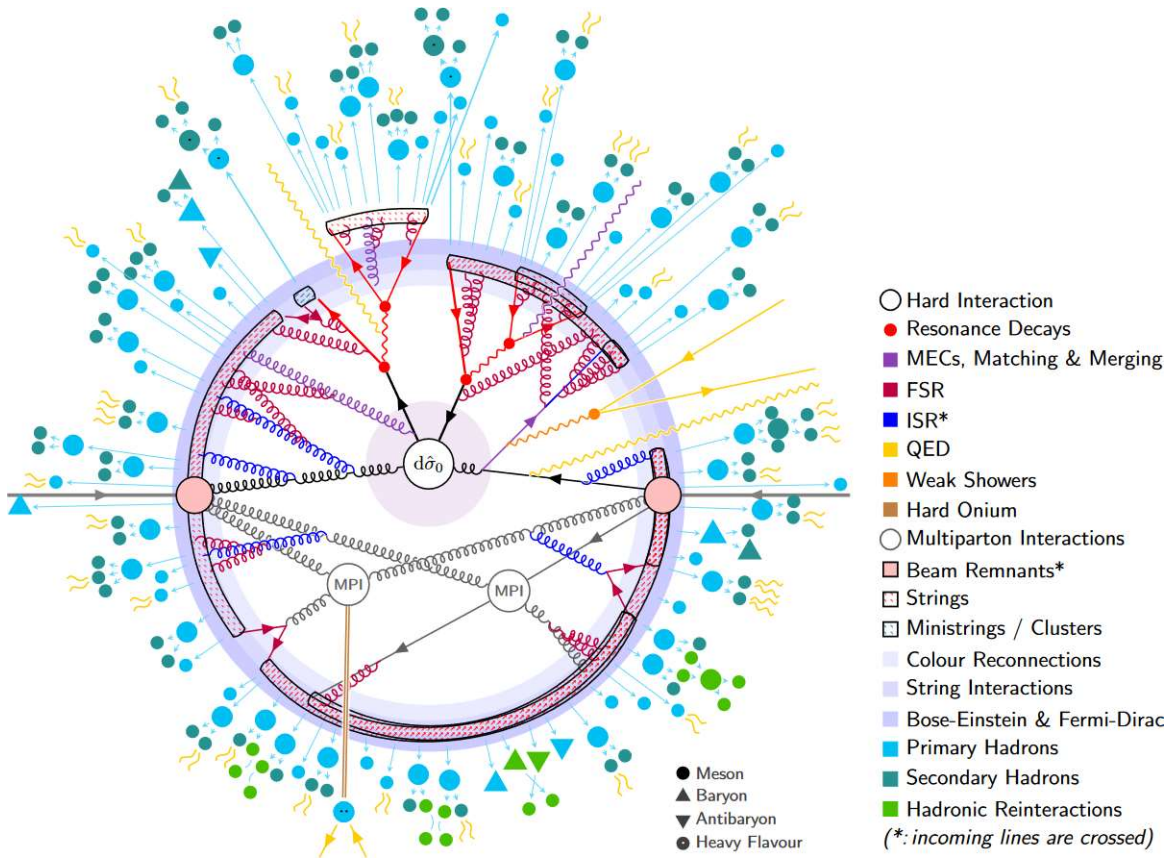


Figure 1.8: Illustration of the structure of a $pp \to t\bar{t}$ event, as modeled by PYTHIA. The phenomena are arranged in decreasing degree of hardness along with the radial coordinate. Reproduced from [7].

1.3.1.1 The hard subprocess in a proton-proton collision

According to the parton model, only two partons of the incoming protons take part in the hard-scatter event. Assuming the matrix element $\mathcal{M}$ can always be computed perturbatively, the cross-section for a proton-proton collision to produce the final state X_n with n the number of final state particles, $P_A P_B \to X_n$, writes:

$$\sigma_{P_A P_B \to X_n} = \sum_{a,b} \int_0^1 \mathrm{d}x_a f_a(x_a, \mu_F) \int_0^1 \mathrm{d}x_b f_b(x_b, \mu_F) \int_\Omega \frac{\mathrm{d}\hat{\sigma}_{ab\to X_n}(\hat{s}, \mu_F)}{\mathrm{d}\Phi_n} \mathrm{d}\Phi_n \tag{1.45}$$

with a and b the possible partons and $f_{a/b}$ the PDFs of the proton, μ_F the factorization scale, $\frac{\mathrm{d}\hat{\sigma}_{ab\to X_n}}{\mathrm{d}\Phi_n} \approx \frac{|\mathcal{M}|^2}{2\hat{s}}$ the differential cross-section at the parton level of the $ab \to X_n$ process (which also has to be averaged over incoming spin and color configurations, and summed over outgoing spin and color configurations), $\hat{s} = x_a x_b s$ and $s = (p_A + p_B)^2$ where p_A and p_B are the 4-momenta of the two incoming protons, $\Omega = \mathbb{R}^{3n}$ is the phase space, and $\mathrm{d}\Phi_n$ is an element of the LORENTZ-invariant phase space:

$$\mathrm{d}\Phi_n(x_A p_A, x_B p_B; p_1, \cdots, p_n) = (2\pi)^4 \delta^{(4)} \left(x_A p_A + x_B p_B + \sum_{i=1}^n p_i \right) \prod_{i=1}^n \frac{\mathrm{d}^3 p_i}{(2\pi)^3 2p_i^0} \tag{1.46}$$

with $p_i = (p_i^0, \vec{p}_i)$ the 4-momentum of the i-th particle, and $p_i^0 = \sqrt{m_i^2 + \vec{p}_i^2}$ with m_i the mass of the i-th particle.

The Monte Carlo method of event generation

To compute this cross-section, the multiple integrals have to be performed, but the phase-space has many dimensions, and standard numerical integration techniques (SIMPSON, GAUSS, *etc*) are not well-suited for a large number of dimensions (larger than 3). The Monte Carlo (MC) method evaluates the integral by uniformly sampling the phase space with N points and by evaluating the integrand at each of these points. For instance, let's say we have a probability density f that is a function of the n-component vector $\vec{x}$, that we want to integrate over the phase space Ω of finite volume V_Ω:

$$I[f] = \int_\Omega f(\vec{x})\, \mathrm{d}^n x \approx \frac{V_\Omega}{N} \sum_{i=1}^N f(\vec{x}_i) \tag{1.47}$$

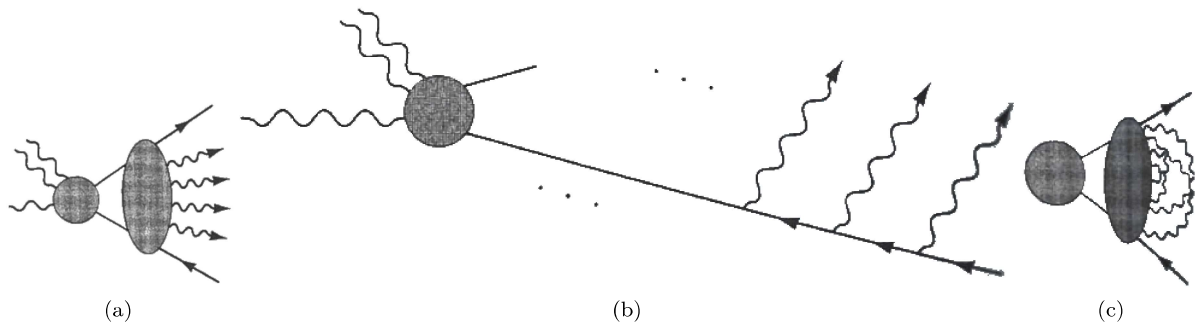


Figure 1.9: **(a)** Illustration of a parton shower (blob on the right and outgoing partons) after a hard process (blob on the left). **(b)** Illustration of the showering of a high-energy parton that radiates multiple partons before reaching the hadronization scale. **(c)** Illustration of higher-order virtual loops, taken into account by the SUDAKOV form factor. Reproduced from [8].

The central limit theorem guarantees that this is an unbiased estimator of the integral in the $N \to \infty$ limit, with an estimator of the error being

$$e[f] = V_\Omega \sqrt{\frac{\mathrm{Var}[f]}{N-1}} \tag{1.48}$$

which shows that the error decreases as $\propto N^{-1/2}$ when N increases.

The points $\{\vec{x}_i\}_i$ generated to sample the phase space can actually be used to represent events, since each of them corresponds to a set of momenta for the particles involved.

This basic Monte Carlo technique presented here is refined by the use of many additional techniques, for instance to improve the error on the estimate (with "importance sampling" and the acceptance/rejection technique), to produce points in the phase space region where the probability density is higher (with MARKOV chains for instance [3]), to reduce or factorize the phase space, *etc.* It is possible to sample more a certain region of the phase space to improve the statistical error in that specific region. In case it is difficult to draw an equal number of events in all phase space regions, events can be weighted such that in a dense phase space region they have a higher weight than events in a sparser region.

1.3.1.2 Parton showers

Angular-ordered shower

When a high-energy parton (quark or gluon) is emitted off-shell from a hard subprocess, it radiates gluons, that can split into $q\bar{q}$ pairs. Each of these radiated quarks or gluons itself carries 1 or 2 color charge(s) and repeats the process of emitting other gluons. This results in what is known as a parton shower, as illustrated in Fig. 1.9a. These can be viewed as higher-order corrections to the hard subprocess, even though it is not feasible to calculate them exactly: an approximation is used, where only the dominant contributions are included. The dominant contributions are collinear parton splittings and soft gluon emissions.

In the collinear approximation, where the initial parton and the new parton from the splitting are almost collinear, it can be shown that the differential cross-section of $i \to j + k$ (for instance $q \to q + g$) factorizes as:

$$\mathrm{d}\sigma_{i\to j+g} \approx \mathrm{d}\sigma_{i\to j} \frac{\alpha_S}{2\pi} \frac{\mathrm{d}t}{t} \, \mathrm{d}z \, \mathrm{d}\phi P_{ji}(z,\phi) \tag{1.49}$$

where P_{ji} is the splitting function already encountered in Section 1.2.5 in the DGLAP evolution of the PDFs (when integrating over ϕ), *i.e.* the distribution of the fraction $z = \frac{E_{\text{daughter } j}}{E_{\text{mother } i}}$ of energy of i carried by j, and t is the virtuality, either chosen as $t = q^2 = (p_k - p_i)^2$ or as $t = \theta^2$ in this section (with θ the angle between k and i), because $\frac{\mathrm{d}t}{t} = \frac{\mathrm{d}q^2}{q^2} = \frac{\mathrm{d}\theta^2}{\theta^2}$.

This splitting equation can be applied repeatedly to partons, using Monte Carlo (MC) methods to generate values of t, z, and ϕ for each splitting, thus producing a parton shower: z and ϕ follow the distribution of $P_{ji}(z,\phi)$, and the choice of t will be explained in the next few lines. This process is repeated from the momentum transfer scale of the hard process Q ($q^2 < Q^2$) until the virtuality reaches the hadronization scale Λ_{QCD} ($q^2 \approx \Lambda^2_{\mathrm{QCD}} \approx 1$ GeV^2). One high-energy parton will therefore radiate multiple partons, as illustrated in Fig. 1.9b, before reaching the hadronization scale.

It might seem like the parton shower only takes into account the real emissions of gluons and not the virtual effects of the same order from the quantum loops (see Fig. 1.9c for an illustration of such a loop). Such effects are taken into account in the probability of *not* splitting during evolution from scale q_1^2 to to scale q_2^2, which is given by the SUDAKOV form factor [2]:

$$\Delta_i(q_1^2, q_2^2) = \exp\left(-\int_{q_2^2}^{q_1^2} \frac{\mathrm{d}t}{t}\frac{\alpha_S}{2\pi}\int_{\Lambda_{\mathrm{QCD}}^2/t}^{1-\Lambda_{\mathrm{QCD}}^2/t} \mathrm{d}z \int_0^{2\pi} \mathrm{d}\phi P_{ji}(z,\phi)\right) \tag{1.50}$$

where the bounds on z come from the fact that an emission outside the range $\left[\frac{\Lambda_{\mathrm{QCD}}^2}{t}, \frac{1-\Lambda_{\mathrm{QCD}}^2}{t}\right]$ is either too soft (below the hadronization scale) or at too small an angle to be detected and is therefore unresolvable.

The SUDAKOV form factor is therefore used to give a virtuality scale for the next splitting. For a parton at a given initial scale Q^2, one draws a random number R_1 in the uniform interval $[0,1]$, then solves the equation $\Delta_i\left(Q^2, q_1^2\right) = R_1$ to obtain the scale q_1^2 of the first splitting. If $q_1^2 < \Lambda_{\mathrm{QCD}}^2$, the splitting is unresolvable, and the showering for that parton is terminated. Otherwise, the splitting is performed, obtaining z and ϕ, providing two partons, and the splitting procedure is repeated for each of them.

The choice of the virtuality variable t comes naturally as θ^2, where θ is the opening angle of splitting, as it treats the coherence of soft radiations correctly. Indeed, a quark-antiquark pair coming from the splitting of a gluon looks resolved at opening angles smaller than θ and unresolved at angles larger than θ: in the latter case, they appear as if they were a gluon, and therefore their radiation pattern looks like that of a gluon, even though the pattern comes from the coherent showering of the $q\bar{q}$ pair.

This is the origin of the angular-ordered showering models.

Dipole shower

An alternative for the showering is the dipole showering, where a gluon is emitted by the dipole radiation of a pair of partons, instead of the splitting of a parton: it is a $2 \to 3$ process instead of a $1 \to 2$ process.

The approximation of a large number of colors is used: each (anti)quark is uniquely connected to one color partner, and each gluon to two partners, forming dipoles. Each dipole splits into two dipoles by emitting a gluon. This gluon's kinematics determine the mass and momenta of the two new dipoles. The splitting procedure is performed until some scale is reached.

This is the basis for dipole showering models, for which many refinements exist.

Initial State Radiations

The showering models described previously are applied to outgoing partons after the hard subprocess, to generate a Final State Radiation (FSR) shower. However, the two incoming partons of the hard subprocess also produce a shower of Initial State Radiations (ISR). The naive treatment, where the 4-momentum of the incoming partons is chosen according to the PDFs followed by a showering and finally by the hard subprocess, is not feasible in simulations of most subprocesses. For instance, if the hard subprocess is $qq \to W$, then the incoming quarks must have an energy close to that of m_W^2 after the showering, which would be very rare and thus the MC efficiency would be very low.

The solution is "backwards evolution": the initial partons taking part in the hard subprocess have their momenta chosen according to the PDFs such that they match the $p^2 = m_W^2$ condition, then they are evolved backwards in time, gaining energy at each emission of the showering, until they reach the low virtuality scale appropriate for partons to be part of the incoming hadrons.

1.3.1.3 Hadronization

When the energy scale becomes close to the infra-red cutoff Λ_{QCD} of perturbative QCD, the QCD coupling constant α_S becomes large due to its running and by definition the perturbative treatment of the QCD effects is impossible. The dynamics of the partons after the showering phase enter a non-perturbative phase called "hadronization", which leads to the formation of final-state hadrons. No non-perturbative mathematical technique is known to treat hadronization from first principles, and therefore event generators have to rely on QCD-inspired phenomenological models.

Lund string model

An important observation of lattice QCD (simulations of the non-perturbative QCD dynamics formulated on a grid of points) is that the potential energy of color sources grows linearly with their separation [3]. The gluonic field collapses in thin tubes that are 1 fm wide (when the separation between the sources becomes larger than this).

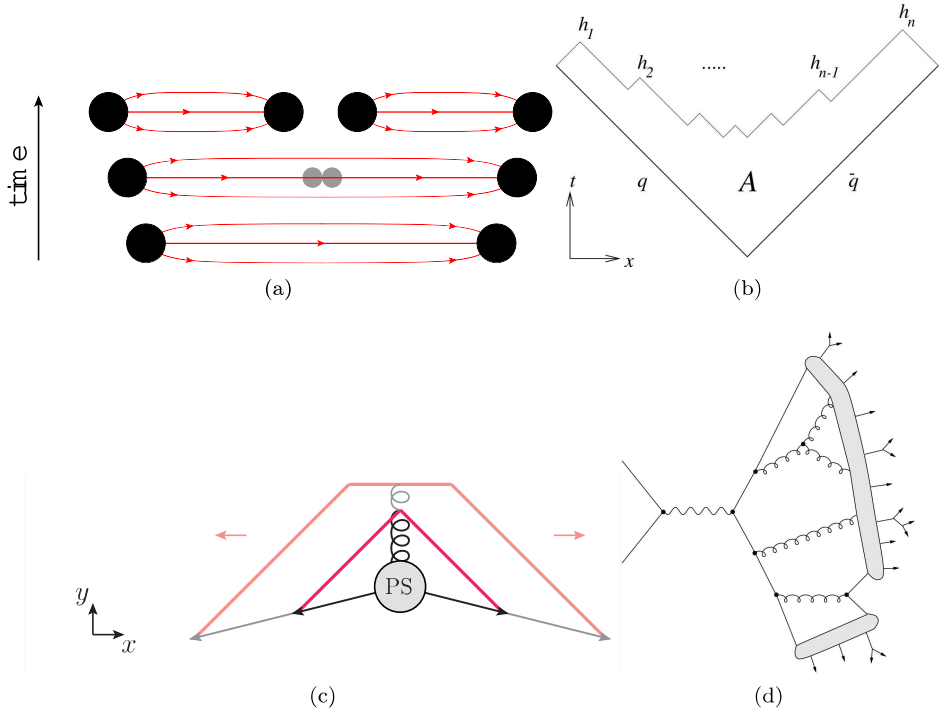


Figure 1.10: **(a)** Illustration of a string breaking from a $q\bar{q}$ pair. **(b)** Space-time picture of a string hadronization. A is the world-sheet of the string, $h_1, \cdots, h_N$ represent the final-state hadrons. **(c)** Illustration of an $e^+e^- \to qg\bar{q}$ configuration emerging from the parton shower (PS). Snapshots of string positions are shown at two different times (full and shaded lines respectively). The gluon forms a transverse kink which grows in the y direction until all the gluon's kinetic energy has been used up. **(d)** String hadronization model. Reproduced from [3, 6].

When a quark-antiquark pair is produced, both quarks are moving rapidly apart in the center-or-mass frame, stretching the string, increasing the potential energy at the expanse of their kinetic energy. When this potential energy reaches the hadron mass, the string breaks and creates a $q\bar{q}$ pair, as illustrated in Fig. 1.10a. The two new strings begin to stretch and break again, and so on until all the kinetic energy has been converted to $q\bar{q}$ pairs, identified with hadrons, as illustrated in Fig. 1.10b. Since the string breaking can be viewed as a vacuum fluctuation, light quark flavors are more likely to be produced. There is a choice for the string breaking to occur at one point or another in the string, but the final fragmentation spectrum should not depend on it, which provides strong constraints on the "fragmentation function" $f(z)$, which governs the probability for a hadron h to be produced at a given step with a fraction $z \in [0, 1]$ of the remaining energy:

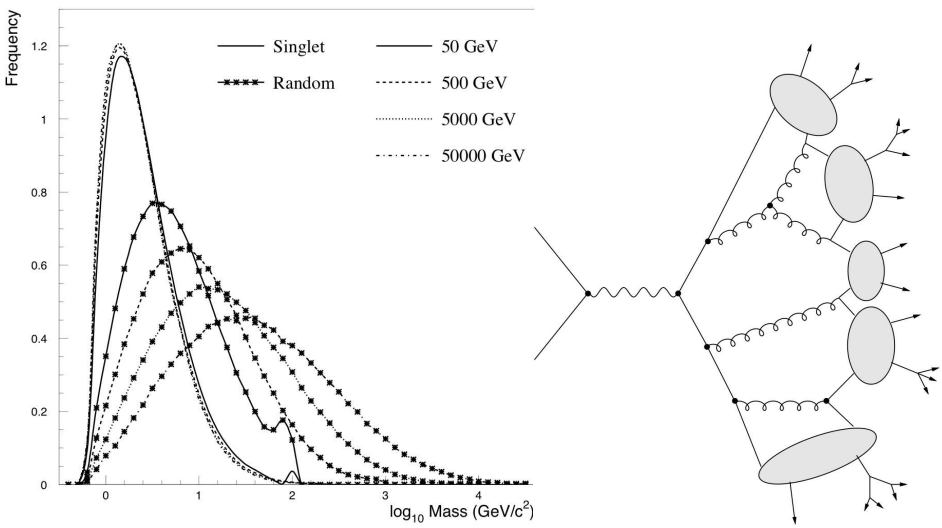
$$f(z) \propto \frac{1}{z}(1-z)^a \exp\left(-\frac{b\left(m_h^2 + p_{T,h}^2\right)}{z}\right) \tag{1.51}$$

with $p_{T,h}$ the hadron transverse momentum and m_h the hadron mass, and where only two free parameters are required: a to control the hard $z \to 1$ tail of the fragmentation function, and b the string tension to control the soft $z \to 0$ tail.

Strings are color-neutral objects, that can involve more than one $q\bar{q}$ pair, for instance it can be a sequence $\bar{q}gggq$. In this case, a gluon introduces a "kink" in the otherwise one-dimensional string object, as illustrated in Fig. 1.10c, making the dynamics of the string evolution more involved mathematically but without the requirement of additional free parameters. A gluon being part of two strings, it looses energy (hence produces hadrons) twice as fast as quarks, which has consequences on the angular distribution of hadrons that are then created preferentially in the direction of the gluon and in the quark-gluon and antiquark-gluon angular regions than in the region between the quark and the antiquark. The hadronization using the Lund string model [3] is illustrated in Fig. 1.10d.

Cluster model

A property of perturbative QCD is the "preconfinement": the partons in a shower, at a certain shower evolution scale q, are grouped in phase space in colorless clusters with an invariant mass distribution that is independent of the nature and scale of the hard subprocess, depending only on the scale q and the fundamental QCD scale Λ_{QCD},



(a) (b)

Figure 1.11: **(a)** Cluster mass distribution. The distribution of color singlet masses tends to a universal curve, whereas that of random quark-antiquark combinations does not. **(b)** Cluster hadronization model. Reproduced from [6].

as shown in Fig. 1.11a. Only the number of clusters depends on the scale of the hard subprocess.

At the hadronization scale, all gluons are forced to decay to $q\bar{q}$ pairs, as illustrated in Fig. 1.11b, then the clusters are identified to "proto-hadrons" that decay into the final-state hadrons: most of the clusters have a low mass and can be decayed using a simple isotropic quasi-two-body phase space model, whereas the few high-mass clusters are decayed using a string-like model of sequential cluster decay.

A difference from the string model is that, in the cluster model [3], the string is always broken at a gluon, rather than having a kink, as illustrated in the comparison between Fig. 1.10d and Fig. 1.11b.

1.3.1.4 Soft QCD: underlying event and collective effects

Multiple other low energy ("soft") effects take place. Only description of these effects is done here for completeness, as they do not play an important role in the physics analyses presented in this thesis: the underlying event is a (minor) source of background noise for the jets stemming from the hard-scatter subprocess, and the non-perturbative effects need to be taken into account by corrections in the cross-section measurements.

The Underlying Event (UE)

The underlying event is an extra hadron production, that cannot be ascribed to the ISR or FSR of the partons involved in the hard subprocess. This activity is believed to come from additional color exchanges between the colliding hadrons, either modeled as additional perturbative exchanges, mainly of gluons, called Multiple Parton-parton Interactions (MPIs) [3, 7], or as non-perturbative effects with "cut pomerons" (roughly equivalent to an exchange of gluons with $p_T \to 0$) [3].

The cross-section of gluon-gluon scattering, and parton-parton interactions in general, is larger than the cross-section of proton-proton collisions, indicating that it is more probable for protons to have multiple parton-parton interactions for each proton-proton interaction than just one.

The extra hadron production requires a model of the spatial structure of hadrons. The partons in each incoming hadron are distributed over a transverse area of 1 fm^2. When the overlap between the areas is small, only a few soft interactions take place and the probability for a hard interaction is low; conversely, when the overlap is large, the probability for a hard-scatter event is larger and there are more MPI interactions.

These other soft interactions from MPI lead to the production of more low-energy hadrons in the final state.

Bose-Einstein and color reconnection effects

Multiple collective effects are responsible for physics correlations between the final-state hadrons, two of which being the color reconnection and the Bose-Einstein (BE) correlations [3, 7].

The color reconnection is the step of color-neutralization between all the partons, after showering and before hadronization, that implies long-range interactions through the gluonic fields between partons with very different

momenta. This color reconnection involves both partons from the hard-scatter subprocess and partons from the MPIs in the underlying event.

The BE correlations are lost at the hadronization stage, due to the string or cluster fragmentations being treated as independent from one another. Models exist to mimic the BE effect, by adding an attractive interaction between pairs of identical particles in the final state.

1.3.1.5 Hadron decay

Only particles that have a long enough lifetime τ reach the detector, with the threshold usually retained being $c\tau > 10$ mm. Hadrons with shorter lifetimes decay into other hadrons. This is important as experimental data shows that a large fraction of the observed final-state particles come from the decays of unstable hadrons. Most of the hadrons listed in [3] have to be taken into account, together with their decay modes. In a simulation, complete flavor multiplets must be included to avoid flavor biases, and the sum of the branching fractions must sum to unity. However, not all particles and excited states have experimentally been observed, especially for heavy flavor hadrons, and not all decay modes have been observed, so models and assumptions are made.

1.3.1.6 Event generators

Multiple General-Purpose Monte Carlo (GPMC) generators implement the different models:

- Pythia8 [7, 9] has the matrix element at leading order, uses a dipole-style showering model that is p_T ordered, and uses the Lund string model. In ATLAS, it uses the A14 tune [10] and the NNPDF2.3 [11] parton distribution function (PDF) set.
- Herwig7 [12–14] generates events either at LO or NLO, has two integrated shower models, either angle-ordered [15] or dipole [16, 17], and uses a cluster hadronization. It is used with the [18] PDF set.
- Sherpa [19] generates LO events, has a dipole showering [20], and has two possible hadronization models: either a cluster one [21] or a Lund string one [7]. It is used with the CT14NNLO [22] PDF set.

Some generators only deal with the matrix element, like Powheg [23–25], that generates NLO events and uses an angle-ordered showering, that can be passed to the hadronization models. It is used with the NNPDF3.0NLO [26] PDF set. For instance, the following combinations are used in ATLAS: Powheg+Pythia8, Powheg+Herwig7.

Having multiple options at our disposal allows to test the different models for the physics predictions, to find the one that best describes the physics of interest, and to evaluate systematic uncertainties by comparing with the other models.

1.3.1.7 Partonic predictions and corrections

Some generators do not serve the purpose of generating events, but use Monte Carlo techniques to integrate over phase-space in order to produce cross-section estimations. For instance, NLOJET++ [27, 28] and NNLOJET++ [29, 30] generate NLO and NNLO events respectively with quarks or gluons in the final state, that are then counted in appropriate bins of kinematic variables and then discarded. The showering, that corresponds to higher order terms, is not simulated so that the distributions correspond to a single order in perturbation theory.

Since non-perturbative effects like the hadronization or the UE are not simulated on an event-by-event basis, corrections have to be made to the parton-level predictions. This can be done by multiplicative factors to the distribution. This also allows to take multiple effects into account, for instance the convolution to a different set of PDFs, the use of a different value of the strong coupling constant $\alpha_S(\mu_R)$, a different tune of the MC generators. The non-perturbative correction factors are shown in Fig. 1.12.

The multiplicative factor $\mathcal{K}^i_{\mathrm{NP}}$ in a given bin i is computed as follows:

$$\mathcal{K}^i_{\mathrm{NP}} = \frac{\mathcal{O}^i\,(\mathrm{ME} + \mathrm{showering} + \mathrm{hadronization} + \mathrm{UE})}{\mathcal{O}^i\,(\mathrm{ME} + \mathrm{showering})} \tag{1.52}$$

with $\mathcal{O}^i\,(\mathrm{ME} + \mathrm{showering})$ the observable $\mathcal{O}$ computed in the bin i using only the matrix element and the showering and $\mathcal{O}^i\,(\mathrm{ME} + \mathrm{showering} + \mathrm{hadronization} + \mathrm{UE})$ using the full event simulation. The non-perturbative correction factor is then applied to the partonic prediction as

$$\mathcal{O}^i\,(\mathrm{partonic} + \mathrm{NP\ effects}) = \mathcal{O}^i\,(\mathrm{partonic}) \times \mathcal{K}^i_{\mathrm{NP}} \tag{1.53}$$

Considering how precise the QCD predictions have become at NLO, Electro-Weak (EW) corrections to the partonic predictions have to be taken into account. The contributions are negligible at LO due to the much larger

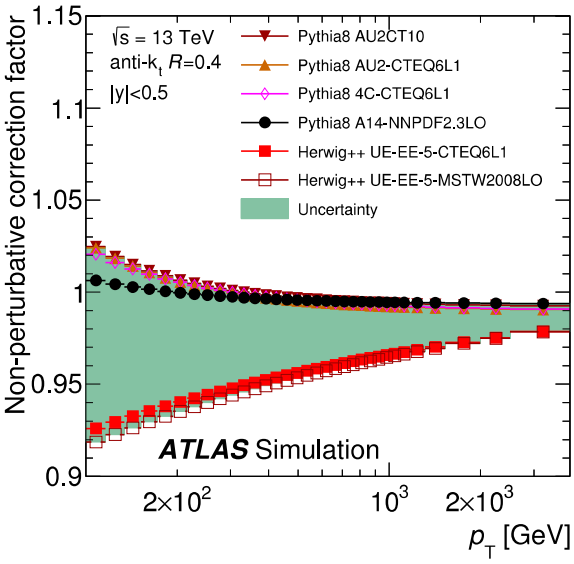


Figure 1.12: Non-perturbative correction factors for the inclusive jet cross-section NLO partonic prediction, in the $|y| < 0.5$ rapidity bin. The correction is obtained for different tunes of the generators and for two generators, providing an uncertainty band. Reproduced from [31].

coupling constant of the strong force, but become large at high p_T, of the order of 10 %, due to an increasing logarithmic SUDAKOV term [32]. Note that in these EW corrections, the weak bosons only intervene in loops, not in real emissions, while photons can be either real or virtual, for IR safety.

The electro-weak correction is computed as the ratio between the QCD+EW partonic predictions to the QCD-only partonic predictions:

$$\mathcal{K}_{\mathrm{EW}} = \frac{\mathcal{O}\left(\alpha_S^n \alpha_{em}^m\right)}{\mathcal{O}\left(\alpha_S^k\right)} \tag{1.54}$$

where $\mathcal{O}\left(\alpha_S^n \alpha_{em}^m\right)$ is the observable computed with specific contributions in the perturbative series: the leading order is given by the terms for which $n + m = 2$, the NLO terms are those for which $n + m = 3$. The contributions to the observable (to the inclusive jet cross-section in the present case) are shown in Fig. 1.13.

1.3.2 Jet definitions

To sum up hadron production in a proton-proton collision event, depicted in Fig. 1.14, one parton of each incoming proton comes into play in the hard scatter event, and in case the event produces quarks or gluons as outgoing particles, they will radiate soft and collinear gluons (as $\int \alpha_S \frac{\mathrm{d}E}{E} \frac{\mathrm{d}\theta}{\theta} \gg 1$) that themselves can split into $q\bar{q}$ pairs (showering) until the non-perturbative scale is reached; then all the partons undergo hadronization: they associate and create new particles to end up with a collection of color-neutral particles, the hadrons. Since the initial parton has a boost, all the final particles share a common boost and therefore are emitted in a cone in the lab frame.

This gives the basis of two natural methods to cluster the hadrons (or the energy deposits in the detector): either by using cones, or by using iterative methods that cluster particles by going through a "hardness" flow. This is called a "jet algorithm", and the final cluster is called a "jet": it is a collimated spray of hadrons.[6]

Then a method should be specified to reconstruct the 4-momentum of the jet, the simplest one being the (weighted) sum of the 4-momenta of the particles. Together with a jet algorithm, this specifies a "jet definition". Both steps represent a choice, that has to be based on the physics studied.

1.3.2.1 Requirements

"Jet [definitions] are legal contracts between theorists and experimentalists."
M. J. TANNENBAUM

For jets to be useful objects, they must give a coherent definition for very different objects, ranging from partons at different orders in perturbation theory, to showers, hadrons and finally energy deposits in detectors, as illustrated in Fig. 1.15. A jet algorithm must satisfy several principles and criteria, defined in the "Snowmass accord" in 1990

[6] This section is mostly based on [33, 34].

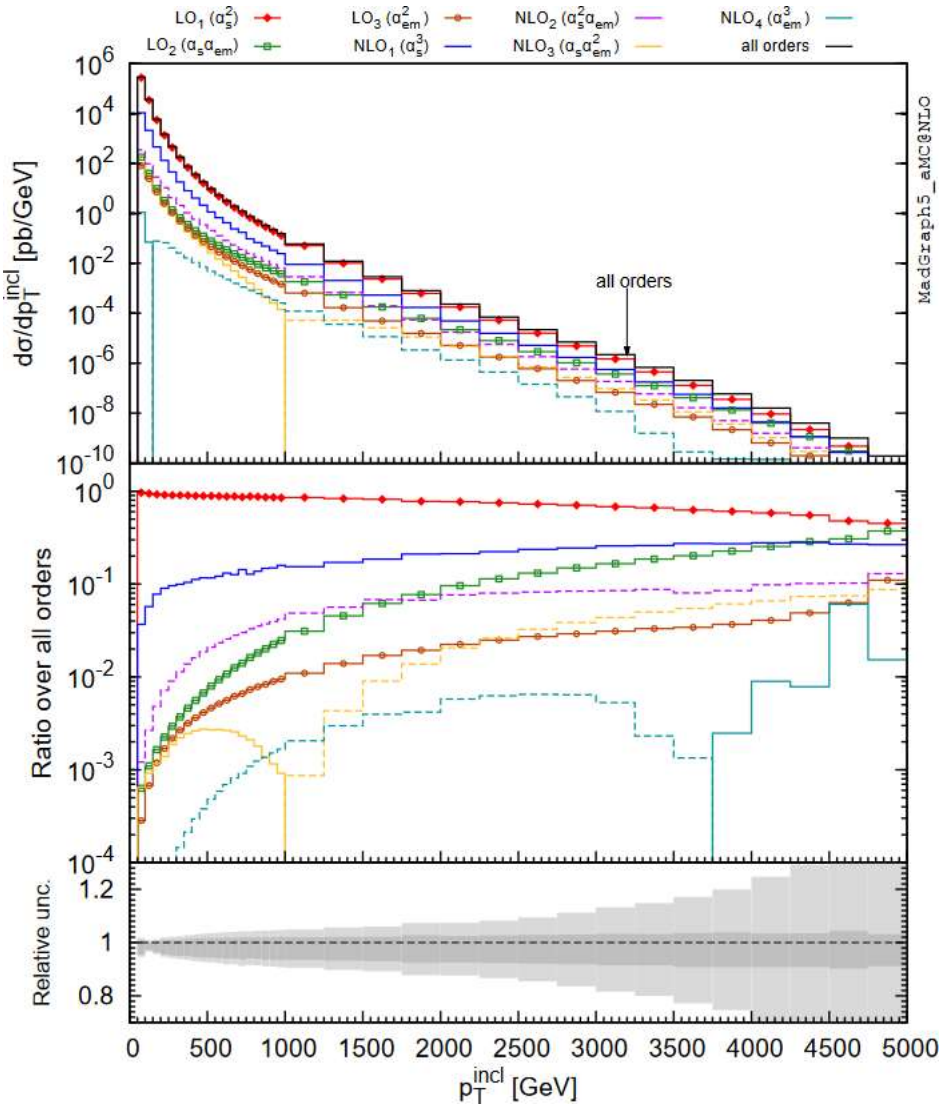


Figure 1.13: Inclusive jet cross-section at parton level at NLO, differential in p_T and inclusive in rapidity, showing the contributions of the different orders in $\alpha_S^n \alpha_{em}^m$ with $n + m = 2$ for LO and $n + m = 3$ for NLO. The solid and dashed lines respectively represent positive and negative contributions. The lower panel shows the uncertainties, where the dark grey band represents renormalization scale uncertainty and the light grey band adds the PDF uncertainties. Reproduced from [32].

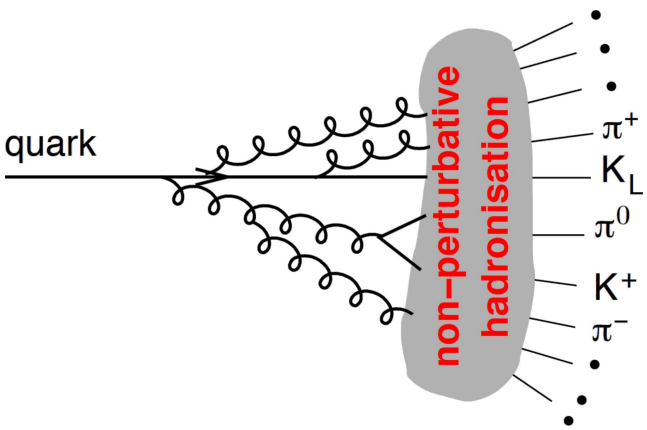


Figure 1.14: Sketch of the different steps of jet production from an outgoing high-energy parton. Reproduced from [33].

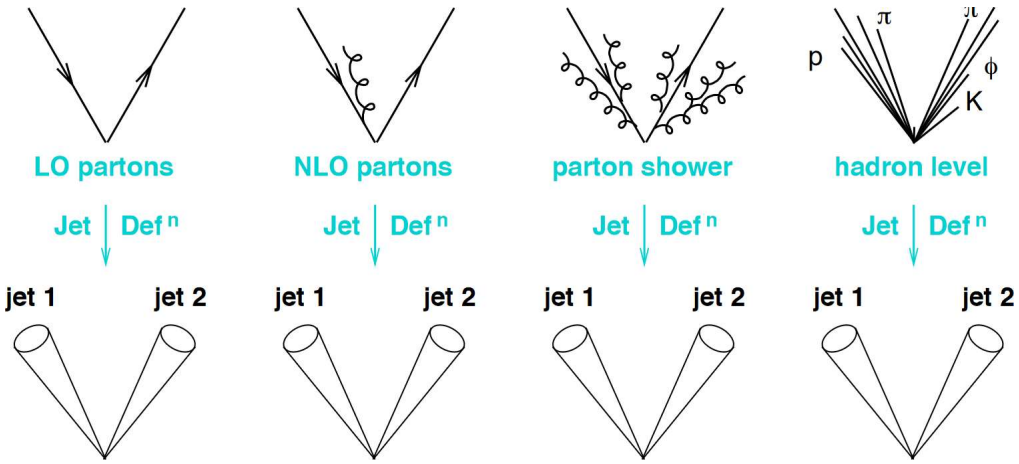


Figure 1.15: Jet finding is a form of projection that should be resilient to QCD effects, like infrared or collinear emissions (IRC safety), showering, hadronization. It has to provide a common definition for different objects: partons at different orders in perturbation theory, showers, hadrons (and energy deposits in detectors, not depicted here). Reproduced from [33].

[35], that are: simplicity to implement both in an experimental analysis and in a theoretical calculation, defined at any order in perturbation theory, yielding a finite cross-section at any order in perturbation theory and yielding a cross-section that is relatively insensitive to hadronization.

In practice, these criteria amount to the following ones:

- Lorentz-invariance. The jet definition should be invariant under the Poincaré group, and notably under longitudinal boosts.
- Infrared safety. The jets obtained should not be significantly altered by the presence or the absence of a low-energy particle (a soft gluon for instance).
- Collinear safety. Infinities in calculations have to cancel, in particular when integrating over the momenta of virtual loops or over the splitting angle of one particle (that splits almost collinearly), both diagrams have infinite weights that cancel. This means that the emission of a collinear radiation should not significantly alter the jets.
- Physical substructure. The particles should be clustered in a way that has physical meaning, *i.e.* such that the history of the clustering steps provides a meaningful hierarchy inside the jet. The showering takes place in decreasing degrees of hardness, therefore natural quantities for the clustering hierarchy are those that qualify as describing the hardness.
- Conical shape. The jet should satisfy basic topographical properties, or else they will not pass the quality criteria. The conical shape is particularly important to simplify the pile-up noise corrections; indeed, area-based pile-up corrections and local pile-up corrections are easier to set up if the jet has a regular shape.
- Weakest sensitivity possible to pile-up (noise coming from the other hard pp collisions than the one of interest) or to the underlying event.
- Computational speed. Algorithms are not equivalent in terms of computational complexity, which is particularly important in the context of the LHC, that involves large amounts of data to be treated for any event.

1.3.2.2 Jet algorithms

Multiple generations of algorithms have been invented, solving different issues along the way, and new algorithms continue being developed, of different kinds: cone algorithms, iterative algorithms are the most prominent types. Among the different algorithms, several ones are infrared-and-collinear safe (IRC safe) and thus very widely used: Cambridge-Aachen (C/A), inclusive k_T, anti-k_T, SISCone [34]. Figure 1.16 shows jets reconstructed for a high jet-multiplicity event from a proton-proton collision in the ATLAS detector with the anti-k_t algorithm; this picture shows that this exercise of clustering particles is an ambiguous and difficult task as jets might be numerous and even overlap in (η, ϕ) space.

The anti-k_T algorithm is a bottom-top algorithm (like all sequential recombination algorithms, unlike cone

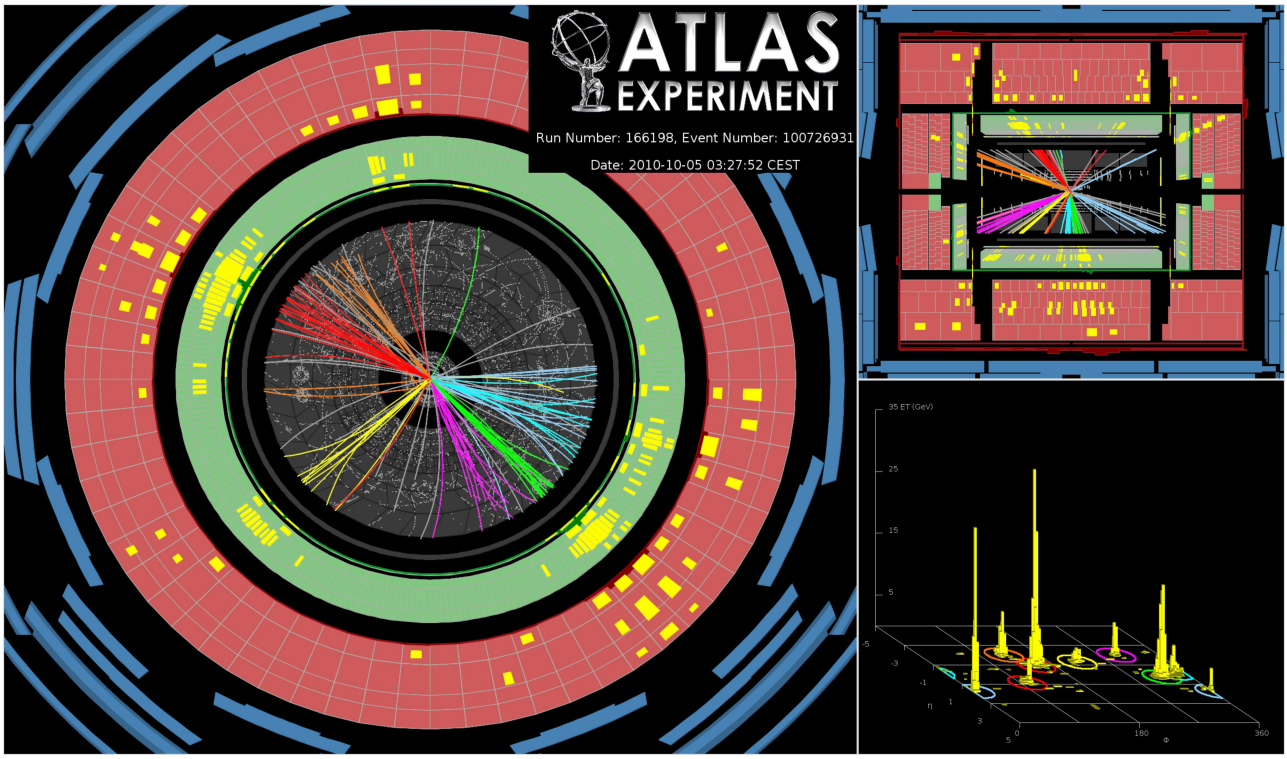


Figure 1.16: The highest jet-multiplicity event collected by the end of October 2010 in ATLAS, counting jets with p_T greater than 60 GeV: this event has eight of them. Left and upper right: sections of ATLAS showing the reconstructed tracks of the jets in different colors, with the energy deposits in the calorimeters in yellow. Lower right: transverse momentum p_T as a function of (η, ϕ), showing the jets isolated by the jet reconstruction algorithm, circled in the same color as the ones used to display the tracks. Reproduced from [36].

algorithms). It clusters sequentially the particles one at a time, using the following "distances":

$$d_{ij} = \frac{1}{\max\left(p_{T,i}^2, p_{T,j}^2\right)} \frac{\Delta R_{i,j}^2}{R^2} \tag{1.55}$$

$$d_{i,B} = \frac{1}{p_{T,i}^2} \tag{1.56}$$

with $\Delta R_{i,j}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ being the angular distance in the (y, ϕ) plane and R being the radius parameter, that acts as a stopping criterion to prevent all particles from being clustered into a single jet. The algorithm is the following one:

1. Work out all the $d_{i,j}$ and $d_{i,B}$.
2. Find the minimum of the $d_{i,j}$ and $d_{i,B}$.
3. If it is a $d_{i,j}$, combine i and j into a single new particle and return to step 1.
4. Otherwise, if it is a $d_{i,B}$, declare i to be a (final-state) jet, and remove it from the list of particles, then return to step 1.
5. Stop when no particles remain.

The k_t algorithm has very similar definition: the distance is defined as $d_{ij} = \min\left(p_{T,i}^2, p_{T,j}^2\right) \frac{\Delta R_{i,j}^2}{R^2}$, giving an algorithm with a meaningful p_T-ordered hierarchy. The Cambridge/Aachen is slightly simpler, and proceeds in two steps: recombine pairs of objects closest in $\Delta R_{i,j}$, and repeat until all $\Delta R_{i,j} > R$: this algorithm has a meaningful angle-ordered hierarchy.

The Seedless IRC-Safe Cone (SISCone) algorithm is different: it is a cone algorithm, meaning that it clusters all particles in cones such that $\Delta R < R$. It takes all subsets of particles and establishes for each one whether it corresponds to a stable cone, *i.e.* whether the addition of a collinear of soft radiation would lead to a new cone or not; to do so, it calculates the total momentum, draws a circle around the resulting axis, and if the points contained in the circle are exactly as those in the initial subset, then one has found a stable cone. A polynomial implementation of this method exists and is called SISCone. More details about these algorithms or other non-IRC-safe algorithms may be found in [34].

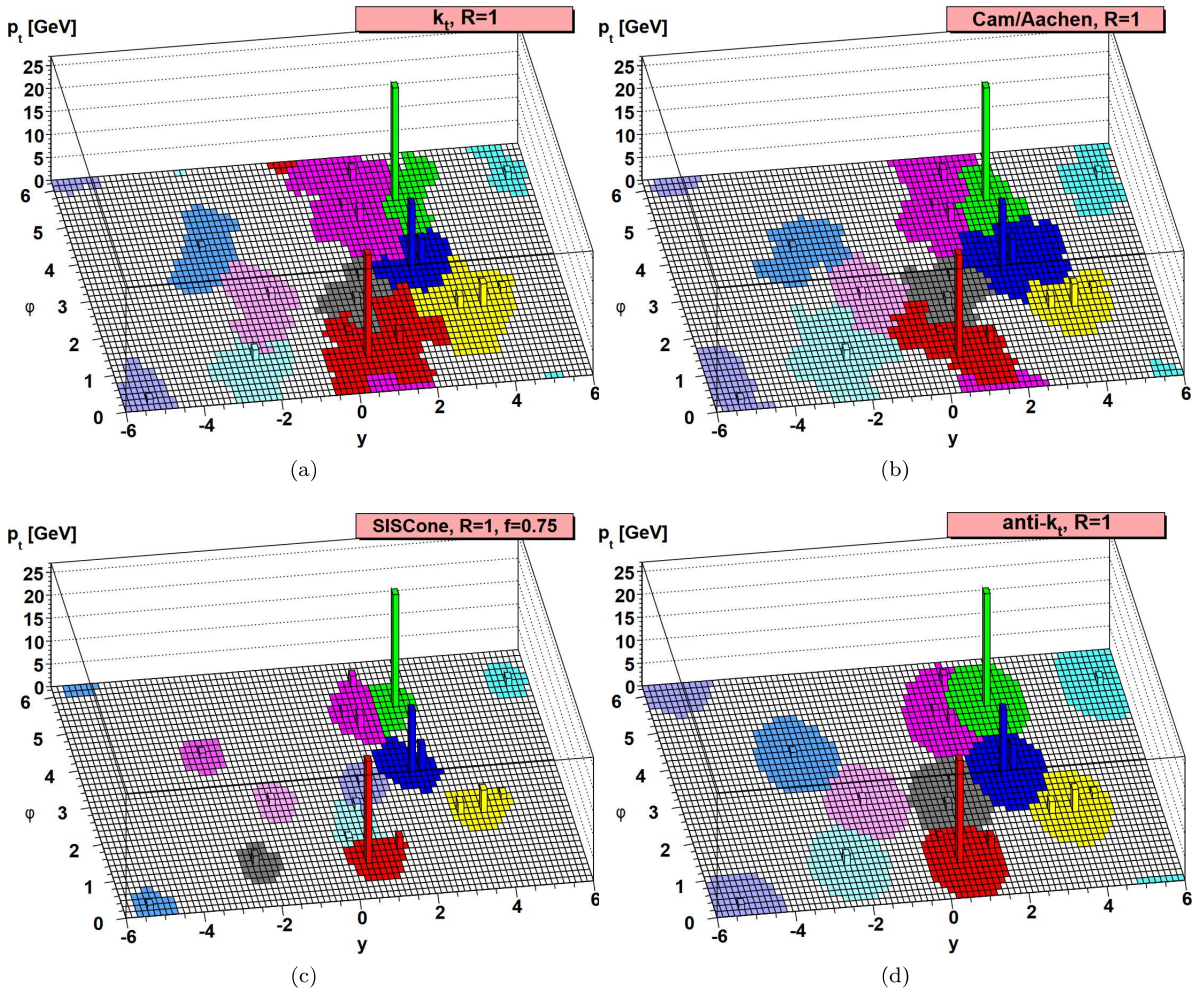


Figure 1.17: Jets clustered from the same inputs (parton-level event together with many soft "ghosts") with different jet algorithms: **(a)** k_t; **(b)** Cambridge/Aachen; **(c)** SISCone; **(d)** anti-k_t. Reproduced from [34].

The anti-k_T algorithm is chosen in ATLAS for multiple reasons:

- It produces circular-shaped cones (as good as SISCone's ones, far better than inclusive k_T or C/A) with the most stable area, as illustrated in Fig. 1.17;
- It is faster than SISCone, and has comparable speed with the other algorithms, as shown in Fig. 1.18;
- It favors clusterings that involve hard particles rather than clusterings that involve soft particles (like the k_T algorithm) or energy-independent clusterings (like C/A), making it almost unaffected by the underlying event and not very sensitive to pile-up.

However, the substructure of the anti-k_t jet does not have a physical meaning (unlike inclusive k_T and C/A), and anti-k_t is more sensitive to pile-up and to the activity of the underlying event than SISCone.[7]

The radius parameter R of the algorithm has to be chosen. It is a compromise between many criteria: resolution of separated jets, capture of perturbative radiations, limitation of the contamination by the underlying event, p_T-dependency, number of jets dependency, computational performance, pile-up resilience, number of particles agglomerated into the jet (flavor dependency for instance), *etc.* Examples of figures of merit are shown in Fig. 1.19, using $Q^w_{f=z}$ as the quality variable (among other possible choices); $Q^w_{f=z}$ is a measure of the quality of the mass peak: it is the width of the smallest mass window around the mass peak that is required for a certain fraction f of the jets to be in the window (it is akin to a mass resolution). The best value of R depends on the physics system under study, and in the example shown on the invariant of the dijet system: it is smaller for quark-initiated jets, that are known to have a larger fraction of their energy in a small core, compared to gluon jets, that radiate more soft gluons at higher angles. In ATLAS, the most widely used radius parameters are $R = 0.4$ and $R = 1.0$, although $R = 0.2$ and $R = 0.6$ are also used for some analyses.

[7]In some cases the k_t algorithm is being used, yielding jets of less regular shape but with a more uniform soft background, see Section 3.3.1.1. The Cambridge-Aachen algorithm is also used since it provides an angle-ordered clustered history, see Section 3.1.2.2.

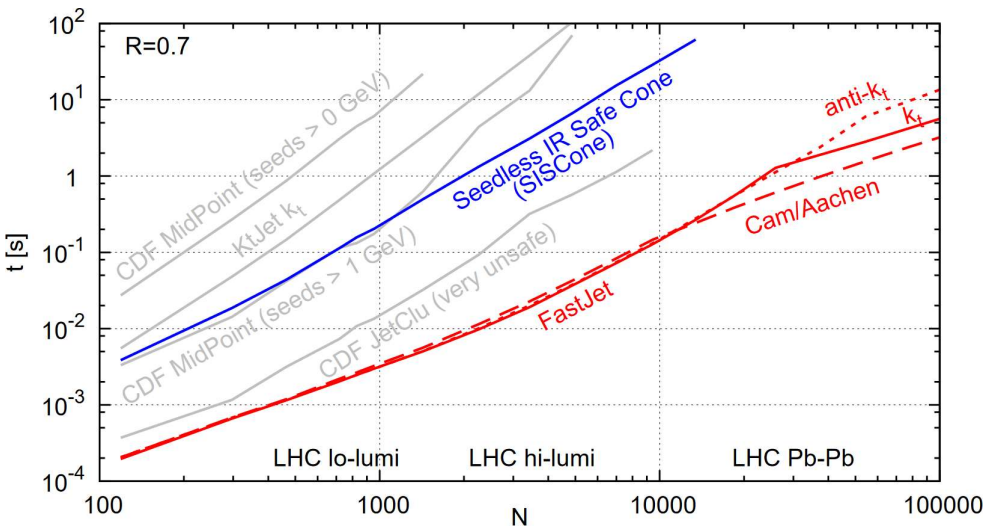


Figure 1.18: Timings for the clustering of a simulated dijet event of approximately 50 GeV, to which increasing numbers of simulated minimum-bias events have been added. One observes k_t, Cambridge/Aachen and anti-k_t in red (implemented in the FastJet package [37]) and SISCone in blue. Other non-IRC-safe algorithms are shown in grey. Reproduced from [34].

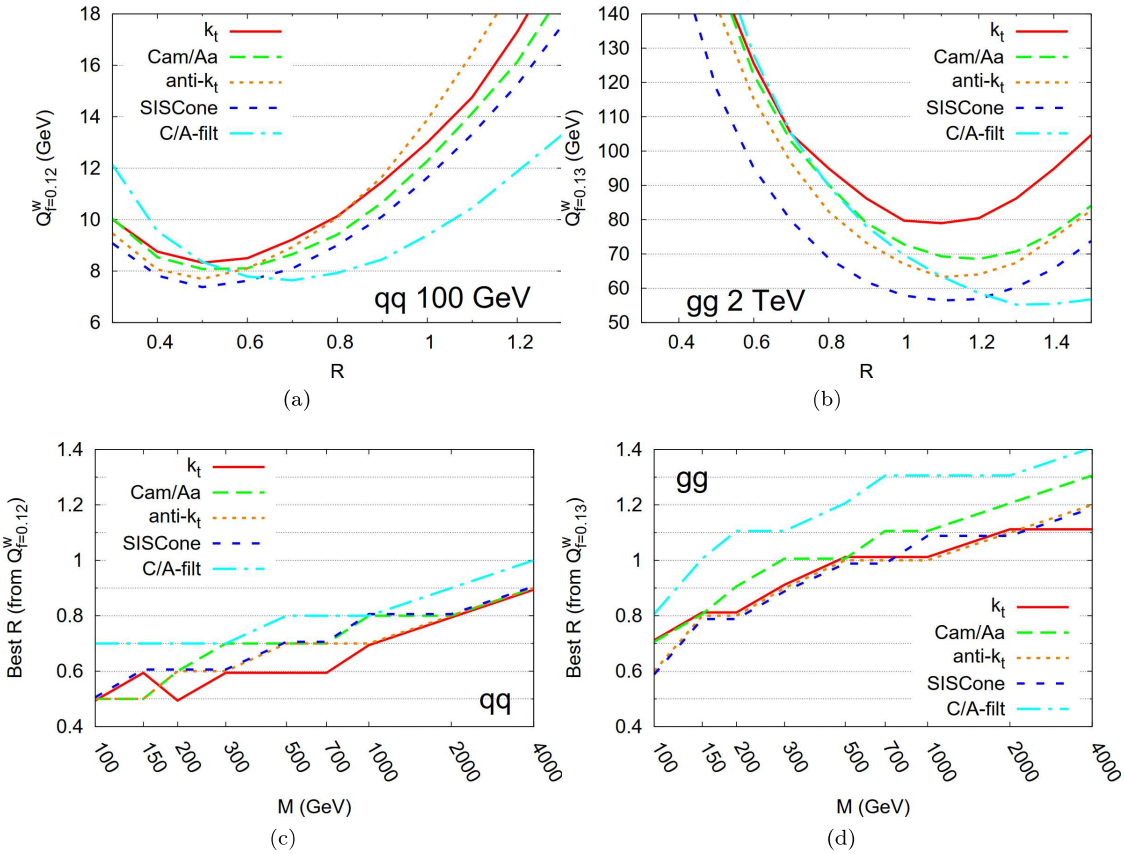


Figure 1.19: Examples of figures of merit for the choice of the jet algorithm to use and the associated radius parameter R as a function of the physics process studied. The quality measure $Q^w_{f=z}$ for **(a)** the 100 GeV $q\bar{q}$ case and **(b)** for the 2 TeV gg case. The optimal value for R, obtained from the minimum of the $Q^w_{f=z}$ *vs.* R distributions, as a function of the invariant mass of the **(c)** $q\bar{q}$ system or **(d)** for the gg system. Note that the optimal value of R changes a bit with the variable used as a quality measure. Reproduced from [38].

1.4 Physics beyond the Standard Model

Despite being incredibly successful at making predictions that are almost all verified by experiments with a high precision [5, 39], the Standard Model (SM) is known to not completely describe nature. This section presents a selected overview of the hints at physics Beyond the Standard Model (BSM), as well as experimental techniques to search for this so-called "new physics".

1.4.1 Hints at physics beyond the Standard Model

Gravity and cosmology

As explained in Section 1.2, the SM describes the properties and dynamics of particles, as well as their interactions through three of the four fundamental forces: the electromagnetic force, the strong nuclear force and the weak nuclear force. However, it does not encompass gravity. The modern framework that describes this force is General Relativity (GR), through geometry of space-time: a massive body curves space-time, thus modifying the geodesics of the motion of other objects, appearing as though these objects were attracted by the massive body. As for the SM, GR revolves around the notion of symmetry, which in this case is a general coordinate invariance. Though quantum field theory in curved space-time is known, a quantum theory of GR is not known. This problem can be thought of in this way: there is no (experimentally verified) theory to deal with the quantum superposition of geometries. To be more precise, GR is perturbatively non-renormalizable, meaning that treating the metric in the Einstein-Hilbert Lagrangian as a perturbation about the Minkowski metric provides interaction operators that are non-renormalizable: an infinite number of counter-terms is needed. This does not make perturbative quantum GR useless, as this theory remains predictive below a certain energy cutoff (below the Planck scale $M_{\mathrm{Pl}} \approx 10^{19}$ GeV), it just indicates that this cannot be the final theory of nature [2], just like the Fermi theory of the weak nuclear force is only predictive below 100 GeV approximately.

Such a quantum theory of GR above the Planck scale is needed in contexts where neither gravitational nor quantum effects can be ignored: for black holes and for the first moments of primordial universe.

On the topic of black holes, there are essentially two open questions [40, 41]. Black holes have a horizon area that cannot decrease, and, if embedded in a quantum field theory, would emit Hawking radiations that satisfy conservation laws, two properties that are very analogous with laws of thermodynamics: entropy and black body radiation [42]. However, in statistical mechanics, entropy counts a number of microscopic configurations. Without a quantum theory of gravity, it is impossible to perform a counting of micro-states to derive the macroscopic entropy. The second question concerns the black hole information paradox: in GR, a black hole is only described by three parameters (its mass, electric charge and angular momentum). Any object falling into the black hole contains a lot of information; however, if Hawking's theory is correct, any black hole completely evaporates through Hawking radiation, that does not carry any information from the matter that initially formed the black hole. This is in contradiction with one essential property of quantum mechanics: unitarity.

The primordial universe question is part of the bigger picture of cosmology, which studies the evolution of space-time. The most widely used model is the Big Bang model, parameterized with the ΛCDM model [43]. Without delving into the details, it describes the expansion of space-time from a single point of high density and temperature. This model is not "standard", because it is in contradiction with quite a few experimental observations (contrarily to the Standard Model of particle physics); still, it explains a lot of properties of the cosmos with reasonably good agreement, and therefore cannot be easily discarded. When going back in time within the ΛCDM model, the universe started as a singularity: a single point of infinite density and temperature. This, however, does not take into account the quantum effects that have taken place at this epoch, thus rendering the model non-predictive at the earliest times (before the Planck time $t_{\mathrm{Pl}} \approx 10^{-43}$ s), where a quantum theory of GR is required. The ΛCDM model comes with a few assumptions and open questions. It assumes the existence of "Cold Dark Matter" (CDM), a currently unknown type of matter that is actively searched for in the field of particle physics, responsible for numerous observations like the "flat" rotation curves of galaxies, the gravitational lensing of light by galaxy clusters or the angular power spectrum of the Cosmic Microwave Background (CMB). This model also assumes the existence of "dark energy" (Λ), an energy of unknown type that fills empty space and is responsible for the accelerated expansion of the universe at contemporary times (that started about 5 billion years ago). When interpreted as a vacuum energy, the cosmological constant Λ can be compared to the predictions of the vacuum energy of the SM, leading to a huge disagreement of 55 orders of magnitude! This is known as the "cosmological constant problem", or "vacuum catastrophe" [43], that can be rephrased in this way: why does the zero-point energy of the vacuum not cause a large cosmological constant?

Another key ingredient of modern cosmology is inflation [43]. Inflation is a short period of exponential expansion of space, taking place from 10^{-36} s to 10^{-33} s or 10^{-32} s after the Big Bang. It accommodates a few experimental observations, like the horizon problem (the universe is very homogeneous and isotropic even if some regions of space have never been in causal contact in the ΛCDM theory), the flatness problem (the density of matter is extremely

close to the critical density required to have a flat universe as the one we live in), or the magnetic monopole problem (magnetic monopoles are hypothetical particles predicted by a lot of Grand Unification Theories of particles physics, that, if they do exist, should have been copiously produced in the primordial universe and would have persisted to present days, being one of the primary constituents of our modern universe, despite not having been observed yet). Inflation commonly uses a quantum scalar field, called the "inflaton", to drive the accelerated expansion of the universe. The inflaton is yet to be observed; the only scalar field known to date, the HIGGS field, could be the inflaton field.

The problems solved by inflation are "naturalness" problems. Naturalness is not a law of nature, but rather an aesthetic principle to guide scientific research, that does not need to be followed but that has proven to be very fruitful. Naturalness states that all parameters in dimensionless theories should be close to 1, otherwise the theory requires "fine-tuning" of the parameters, *i.e.* the values of the parameters have to be extremely precisely "tuned" for our universe to exist as we know it. This fine-tuning calls for an explanation, either by a completion of the model like the addition of the inflation to cosmology or by philosophical considerations like the "anthropic principle" that states that there could be many patches in the "universe", all with different physical properties, and that we live in the one that has the adequate properties for intelligent life to emerge [2].

Another interesting problem is the unknown origin of the equivalence principle [43]. It states that the inertial mass is equal to the gravitational mass. The inertial mass of fundamental particles appears as the mass terms in the SM Lagrangian and results from the interaction with the HIGGS field, while the inertial mass of composite objects has a dominant contribution from the potential energy of interaction between its constituents. The gravitational mass is the one curving space-time in GR, responsible for gravitational attraction. Those two phenomena do not seem to have anything in common, and yet the two masses are measured to be equal to a 10^{-15} precision.

Completing the Standard Model

First, the description of neutrinos is incomplete, on multiple aspects [2]. The adequate representation of the POINCARÉ group for neutrinos is not known: they could be of MAJORANA type, thus being their own anti-particles, or of DIRAC type (like all other fermions), thus requiring the existence of right-handed neutrinos. Right-handed neutrinos (and left-handed anti-neutrinos) have not been observed. The observation of neutrino oscillations (from one flavor to another) implies that they have a mass, that is too small to have been measured and that therefore is currently unknown (those very small masses constitute another naturalness problem). The mass ordering between the three neutrinos is also unknown.

Though the SM explains a lot of natural phenomena, it does still require a relatively large number of parameters, 26 of them (assuming DIRAC neutrino masses), that need to be experimentally measured because their values are not predicted by the theory: three coupling constants g, g' and g_S, six quark masses, three charged lepton masses, three neutrino masses, three mixing angles and one phase among quarks, three mixing angles and one phase among leptons, the HIGGS mass m_h and vacuum expectation value v, and the QCD vacuum angle $\bar{\theta}$ (mentioned hereafter) [2].

Multiple questions are natural to ask: why three families of fermions? Why such different masses? Why such a complicated symmetry group structure and such a complicated representation structure? There are a lot of theories attempting to answer these questions, in different ways, for instance by using a larger symmetry group that would be spontaneously broken, or by providing a UV-completion of the SM like the string theories, to only cite a few approaches.

The matter-antimatter asymmetry is another problem of the SM [2]. This asymmetry is observed in the everyday life: the world is made of matter, not of antimatter. The SM already incorporates CP violating terms in two different ways, the "weak" CP violation by the mixing angles in the fermion and lepton sectors, and the "strong" CP violation in the QCD sector measured by the "strong CP phase" $\bar{\theta}$ (also known as the QCD vacuum angle). But there is not enough CP violation in the SM to explain the observed abundances of matter and antimatter in the universe.

A closely related problem is the "strong CP problem" [2]. The QCD Lagrangian incorporates CP violating terms, that have the important consequence that the neutron picks up an electric dipole proportional to $\bar{\theta}$. This neutron electric dipole has not been observed yet, putting stringent constraints on the value of $\bar{\theta}$, that has to be less than 10^{-10}. This smallness of $\bar{\theta}$ compared to the large amount of CP violation in the weak sector is the strong CP problem, which is another naturalness problem.

The hierarchy problem is another naturalness issue of the SM [2]. The pole mass of the HIGGS boson is measured to be $m_P = 125$ GeV, a value that requires a lot of fine-tuning. Indeed, the mass of scalar particles is UV sensitive: the renormalized mass of a scalar boson depends on the masses of the heavy DIRAC fermions to which it couples. Therefore, the on-shell pole mass (measured experimentally) and the renormalized mass can be significantly different:

assuming a UV-completion of the SM with the HIGGS boson coupling to a fermion around the PLANCK mass, the renormalized mass of the HIGGS boson is approximately $m^2 = \left(1 + 10^{-34}\right) M_{\mathrm{Pl}}^2$. This requires a 34 digit precision.

This fine-tuning problem is related to another one, also referred to as a "hierarchy" problem: why is gravity so much weaker than the three other forces of nature? While there is no explanation for why gravity has the strength it has, the strength of the weak force is closely related to the masses of the W and Z bosons, themselves related to the mass of the HIGGS boson. Therefore, a different HIGGS boson mass could make the weak force a lot weaker, closer to gravity.

Another closely related open question is the potential metastability of the electroweak vacuum [44]. It so happens that the measured values of the HIGGS boson mass and the top quark mass are compatible with the electroweak vacuum being metastable: the HIGGS potential, after renormalization, has a different shape than the usual Mexican-hat potential, with global minima more stable than the local minimum in which our vacuum lies. A brutal transition between both vacua through quantum tunneling could then occur, changing the properties of physics. These metastability criteria could however drastically change in case of the discovery of new physics.

On the topic of force strength, it so happens that the renormalized coupling constants are almost equal around 10^{16} GeV, calling for an explanation. Indeed, the electromagnetic and weak forces are already unified in the electroweak theory (though with two separate coupling constants), but the strong and electroweak forces are still unrelated phenomena. This would explain why the weak hypercharge is quantized (while the interactions based on $SU(3)$ and $SU(2)$ are naturally quantized), thus explaining why the electric charge is quantized, and therefore why the electric charges of the proton and the electron seem to cancel out exactly to extreme precision.

A related open question is: do magnetic monopoles exist? This question has been asked about a century ago by DIRAC, who understood that non-observation of magnetic monopoles is responsible for the MAXWELL equation $\nabla \vec{B} = 0$. The existence of one single magnetic monopole at any point in space-time would explain why electric charges are quantized [45].

Even the mathematical consistency of the SM is questioned. YANG-MILLS theories, *i.e.* non-Abelian quantum field theories, and properties like color confinement, are not proven to exist in a mathematically rigorous way [46]. This is actually a "Millennium Prize Problem".

Also, considering that the SM is based on quantum mechanics, all fundamental open questions related to quantum mechanics apply to the SM as well, like the quantum measurement problem [47]. While the SCHRÖDINGER equation describes the deterministic time evolution of a linear superposition of states, measurement chooses one of these states in a probabilistic way. The paradox appears when considering that the system comprising the experiment, the measurement device and the observer is itself subject to quantum mechanics, *i.e.* to a deterministic time evolution, and therefore no wave-function collapse should occur. Quantum mechanics makes different predictions when considering that the observer is outside (wave-function collapse) or inside (deterministic evolution, no collapse) the system.

Finally, the SM does not explain some fundamental hypotheses of physics that are solely based on observation, like the number of dimensions of space-time (which is predicted by string theories for instance), or the fact that only fields and their first derivatives are required to build Lagrangians that describe nature (*i.e.* second derivatives are uniquely determined by the knowledge of fields and their first derivatives, *i.e.* Lagrangians do not depend on the second derivatives of fields) [48].

Non-confirmed anomalies

Some experimental results are not conclusive yet, as they need more studies to reach the high significance thresholds applied for discovery claims in High Energy Physics (HEP), but are interesting glimpses at physics Beyond the Standard Model (BSM) [49]. In particle physics, the SM has been very extensively tested and its predictions have withstood the test of time until now, so any deviation from it would need a very solid justification, as goes with the saying attributed to Carl SAGAN: "extraordinary claims require extraordinary evidence". Furthermore, experimental results require a lot of resources, and strongly orient future research, so this high standard is a necessity.

The standard in HEP is usually "5 σ", which quantifies the threshold of statistical significance above which an experimental result is claimed to be a discovery; it means that the probability for the observed phenomenon to come from a statistical fluctuation is less than 0.00006 % (a percentage that depends a bit on the context and on the technical details of the search). However, this threshold is somewhat arbitrary, and it is argued that discoveries that contradict the most well established theories should exhibit a significance of more than 5 σ [50]; for instance, the impact of systematic uncertainties, that are usually less well defined from a statistical point of view, or the presence of an important look-elsewhere effect, increase the requirements in number of σ.

The anomalous magnetic dipole moment of a fermion $a = \frac{g-2}{2}$ is a measurement of the quantum corrections to the tree-level prediction of the magnetic dipole moment g (which quantifies the strength of the fermion as a

magnetic source). In the case of the electron, the QED prediction agrees extremely well with the experiment to more than 10 digits. In the case of the muon, a 4.2 σ tension is observed [51].

The Cabibbo-Kobayashi-Maskawa (CKM) matrix describes the mixing of quark flavors in the mass eigenstates. Two coefficients of this matrix, $|V_{ub}|$ and $|V_{cb}|$, are measured through semi-leptonic decays of B mesons, in which the following process takes place: $\bar{b} \to \bar{c}(\bar{u})l^+\nu_l$ (with l a lepton). Those two coefficients are in tension with the SM by about 3 σ [52].

A few "small" excesses in neutral scalar channels have been reported by multiple experiments, one of which is the 95 GeV excess in the diphoton channel reported by CMS, in tension with the SM at the level of 2.9 σ [53]. This excess is consistent with other small excesses in other channels or by other experiments.

There are also some experimental results that are in tension with the SM or with other experimental results by more than 5 σ, but that are not as widely accepted as the BSM hints mentioned in the previous paragraph, due to suspected systematic effects that require more careful understanding. One such results is the W mass anomaly, that has recently been measured by the CDF experiment to be in 7 σ tension with the SM expectation [54], bringing the world average of the W boson mass to a 3.7 σ tension with the SM. This result is currently scrutinized in detail, as it is very sensitive to systematic effects and in strong tension with all other measurements, including a more recent one by ATLAS [55].

The non-perturbative frontier

Though not being Beyond the Standard Model *stricto sensu*, as it relies on the SM Lagrangian and its QFT framework, little is known about the non-perturbative aspects of the SM, making it a frontier of current knowledge. A fluid mechanics analogy is the following: only studying perturbative QFT is like restricting the study of the sea to waves, "neglecting" the ocean currents or turbulence for instance. To draw another parallel, non-perturbative effects also play a very important role in solid state physics in the description of superconductors. Not all aspects of nature can be described as small perturbations, and hadronization in QCD is one such examples in particle physics: it is encompassed within the SM, but analytical predictions are currently impossible, so "semi-classical" QCD-inspired models are used (see Section 1.3.1.3). This frontier is worth exploring, as the SM probably encompasses many fascinating phenomena that we are currently unaware of, just like high-temperature superconductors were not theory predictions but experimental discoveries (and that are still not understood theoretically).

Some non-perturbative phenomena are already known within the SM with an analytical description. For instance, the Schwinger effect is the spontaneous decay of a (very high) electric field into electron-positron pairs [2]. Another example is instantons, that are perturbations around local minima of the action (obeying the equations of motion) which are not global minima; hence instantons are non-perturbative phenomena [44, 45].

Of course, the reason why the perturbative approach of QFT is still dominant is simply due to the lack of general mathematical tools for non-perturbative physics. A very successful approach is a numerical one, lattice QCD: it consists in discretizing space-time into a lattice of spacing a, which acts as a regularization criterion of the theory, and taking the limit $a \to 0$ [3]. It is quite successful in predicting various hadronic states, hadron masses, evaluating form factors for semi-leptonic meson decays, evaluating α_S, evaluating the hadronic vacuum polarization contribution to the muon $g-2$ and to the running of α_{QED}, *etc.*

New mathematical tools are also being developed, like the correspondence between a strongly coupled Conformal Field Theory in flat space-time (that contains non-perturbative effects) to a weakly coupled string theory in curved anti-de Sitter space-time (with more space dimensions), a tool known as "AdS/CFT correspondence". Though fairly new, the applications are flourishing; for instance, it provides a good agreement for the quenching of jets in quark-gluon plasma [56].

1.4.2 Search strategies for physics Beyond the Standard Model

With the numerous hints enumerated above, the search for BSM physics appears justified, both theoretically and experimentally.

There is a plethora of BSM models, each attempting to solve one of the aforementioned issues but rarely all of them, of different categories:

- One can add new fields, providing them with quantum numbers and a representation structure of the SM symmetry group, like right-handed neutrinos or a fourth family of fermions.
- One can add couplings between SM particles, hypothesizing that those (not necessarily renormalizable) couplings are effective manifestations of a more fundamental microscopic theory that is yet to be discovered, just like the Fermi theory of weak interaction is an effective low-energy theory of the SM weak interaction mediated by W and Z bosons. This is the approach of Effective Field Theories (EFTs).
- One can extend the symmetry group of the SM and add a spontaneous symmetry breaking of the newly added symmetry like the Higgs mechanism. More specifically, the symmetry can be extended in two ways.

One way is to add supersymmetry, that predicts that to every SM particle corresponds another particle (like the CP symmetry that relates matter and antimatter).

The other possibility is to extend the gauge group. For instance, one can add other gauge groups like the PECCEI-QUINN $U(1)$ group to solve the strong CP problem and predict a scalar boson called the "axion", or one can predict compositness of some SM particles by adding other symmetry groups corresponding to a new fundamental force (like the compositness of hadrons). Or one can embed the SM gauge group into a larger one like $SU(5)$, $SO(10)$, E_6, *etc*, which is one way to obtain a "Grand Unification" of the strong and electroweak forces.

- One can provide a full UV completion of the SM. The models that try to do so are called "Theories of Everything", and are usually based on mathematical grounds that are quite significantly different from "standard" QFT as introduced in this thesis. For instance, string theories change one fundamental assumption of QFT, which is that particles are described by one-dimensional objects called "strings" instead of by point-like particles, having numerous implications and a very rich phenomenology.

Of course, this short list only scratches the surface of all the possibilities, that nowadays are more limited by the imagination of theoretical physicists than by experimental observations. This is due to the excellent agreement between the SM and experimental observations, that does not lead to any very clear breach where to focus the efforts of the theoretical community, making new experimental results that would contradict the SM very awaited.

Due to the huge diversity of predictions of all these possible models, ranging from subtle corrections applied to current observables to countless new particles at every possible range in energy and coupling, the experiments searching for BSM physics are also extremely diverse. The ATLAS experiment at the LHC, presented in Chapter 2, is a general purpose detector, meaning that a single experiment is versatile enough to make measurements of many different observables. The approaches carried out in ATLAS can essentially be classified into two categories, differing on the energy at which the additional particles appear:

- "Direct" search: if the new particles are in a phase space region that is accessible by experiment, they will appear as resonances in a spectrum for some final state observable. Direct searches typically look for bumps in mass spectra. They can be model-agnostic, but are usually guided by existing models for the definition of their observables.
- "Indirect" search: if the new particles are out-of-reach, they can still manifest themselves through quantum corrections to the observables, *i.e.* loops in the FEYNMAN diagrams. Indirect searches typically look for smooth deviations from the SM in precision measurements.

The physics analysis presented in this thesis in Chapter 5 focuses on this second type of search. The inclusive and leading jet production cross-sections are precision measurements of an SM prediction, done double-differentially to allow for the potential observation of smooth deviations in spectra, that could then be interpreted within different BSM models.

Chapter 2

The ATLAS Experiment

One ring to rule them all, one ring to find them,
One ring to bring them all and in the darkness bind them.

J. R. R. Tolkien, *The Lord of the Rings*, volume 1, 1954

Introduction: CERN

CERN is the European organization for nuclear research, located at the border between France and Switzerland near Geneva [57]. The acronym stands for "Conseil Européen pour la Recherche Nucléaire", a temporary provisional council for building the laboratory. CERN was founded in 1954 with the support of UNESCO, following the initiative of Louis de Broglie at the 1949 European conference dedicated to culture, for Europe to restart working on nuclear physics after the Second World War. Nowadays, more than 17500 people contribute to CERN, comprising over 12200 scientists representing more than 500 institutes from more than 110 countries, as shown in Fig. 2.1.

Over the years, a large network of interconnected accelerators and experiments grew: a dozen accelerators are currently in use, and another dozen have been decommissioned, allowing the study of many aspects of nuclear and particle physics. The website of CERN [58] currently lists 1 256 experiments and projects, of which about 300 are in progress, taking or analyzing data. CERN has a large record of fundamental discoveries, including the discovery of neutral currents in 1973, the discovery of the W and Z bosons in 1983, the determination of the number of light neutrino families in 1989, the creation of antihydrogen in 1995 and its isolation in 2010, the discovery of direct CP violation in 1999, the discovery of the Quark-Gluon Plasma (QGP) state of matter in 2000, the discovery of the Higgs boson in 2012.

Besides its scientific missions, CERN also has societal missions, the first of which being explicitly "bringing nations together", but also transferring technologies to industry. Most notably, CERN is at the origin of the World Wide Web, invented in 1989. Technologies developed at CERN for particle acceleration and detection are now widespread. To only give a few examples, Charpak's multi-wire proportional chamber, invented in 1968, greatly improved X-ray imaging; crystal-based detectors developed for experiments in the 1980s are now ubiquitous in PET scanners; accelerator technologies are used in many domains of science like chemistry, solid state physics, biology, art restoration; finally, the high demand in state-of-the-art technologies and materials pushes industry, for instance to develop and produce better superconducting magnets.

The ability that CERN has to reuse its already existing infrastructure is a strength, used for modern-day accelerators like the Large Hadron Collider (LHC), as we will see in Section 2.1. Particle collisions at high energy are produced inside huge detectors, one of which being ATLAS, described in Section 2.2.

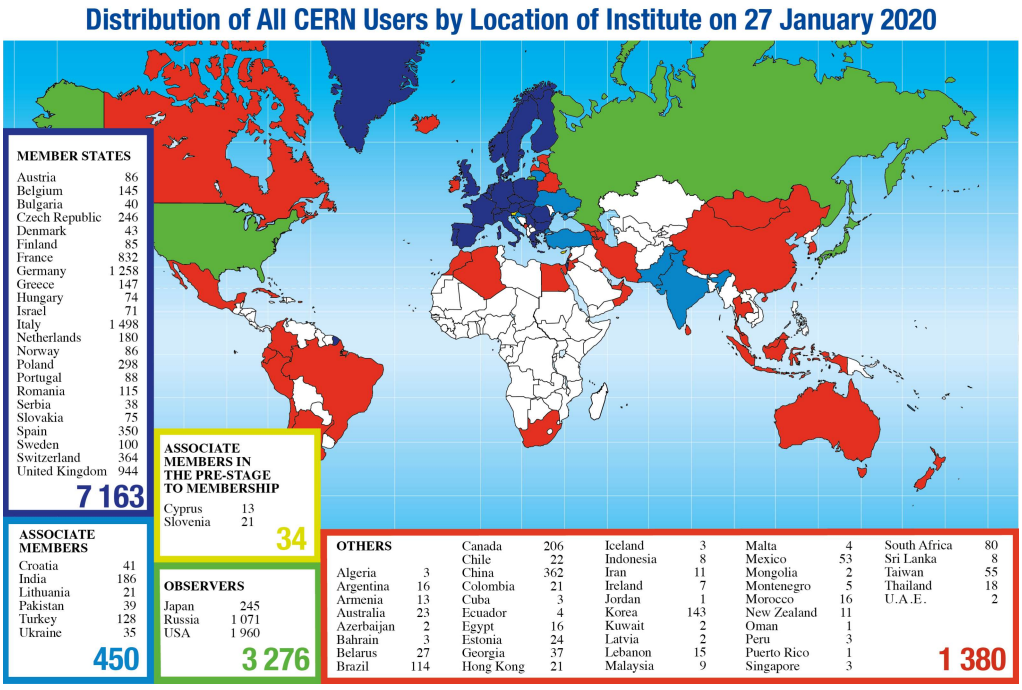


Figure 2.1: CERN users from the participating countries over the world on 27th January, 2020. Reproduced from [57].

2.1 The Large Hadron Collider

The Large Hadron Collider (LHC) accelerates hadrons to nearly the speed of light in two counter-rotating beams, reaching the highest kinematic energies ever attained in an accelerator, to produce collisions in specific instrumented areas.[1] The LHC project was initiated in 1984, as a proposal of double-bore proton accelerator to be placed in the Large Electron-Positron collider (LEP) tunnel. It got approved by the CERN Council in 1994, and after the decommissioning of the LEP in 2000, the construction of the LHC began. The last dipole magnet was delivered in 2006, and the LHC got completed in 2008. The first beam was obtained in 2008, and the first proton-proton collisions in 2009, starting the "Run 1" data-taking period of the LHC. The design energy of LHC is $\sqrt{s} = 14$ TeV; this energy has not been achieved yet but operators are getting closer and closer with every upgrade.

2.1.1 Operation periods

The energies achieved during the multiple runs are:

- During Run 1, proton-proton collisions were at $\sqrt{s} = 7$ TeV during 2010 and 2011, and 8 TeV during 2012. Proton-lead nucleus collisions were at $\sqrt{s_{pN}} = 5.02$ TeV and lead-lead collisions at $\sqrt{s_{NN}} = 2.76$ TeV (with N a nucleon).
- During Run 2, proton-proton collisions were at $\sqrt{s} = 13$ TeV from 2015 to 2018. Proton-lead nucleus collisions were at $\sqrt{s_{pN}} = 5.02$ TeV and 8.16 TeV, lead-lead collisions were at $\sqrt{s_{NN}} = 5.02$ TeV and xenon-xenon collisions were at $\sqrt{s_{NN}} = 5.44$ TeV. To give an order of magnitude, lead has 208 nucleons, and therefore $\sqrt{s_{NN}} = 5.02$ TeV collisions are equivalent to $\sqrt{s_{\mathrm{PbPb}}} = 1.04$ PeV.
- During the ongoing Run 3, proton-proton collisions are at $\sqrt{s} = 13.6$ TeV, and lead-lead collisions have been performed at $\sqrt{s_{NN}} = 5.36$ TeV.

During these periods, there have also been series of "special" runs, like proton-proton collisions at 900 GeV, 2.76 GeV, 5.02 GeV, *etc.* This thesis focuses on proton-proton collisions produced during the Run 2 data-taking period, therefore at 13 TeV.

The energy of collisions is an important parameter, but the amount of data collected is also crucial. Indeed, collision events that generate particles with a 1 GeV energy are common, but events generating 1 TeV particles are a lot rarer. In general, rare events are sought for, not necessarily the highest-energy ones. Finally, precision measurements require good statistical uncertainties, and therefore a lot of events, as well as a good calibration of the reconstructed objects that also highly depends on the number of events. The ability for the LHC to produce a

[1] This section is mostly based on the lectures of Nicolas Delerue, given for the doctoral school PHENIICS in 2021 [59], and on [60].

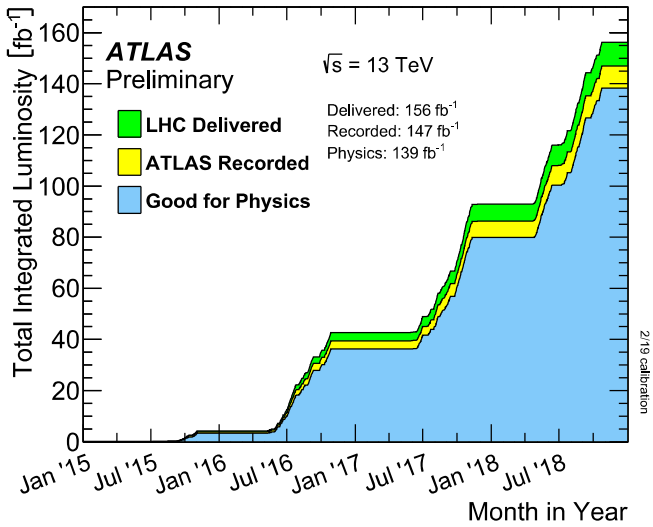


Figure 2.2: Integrated luminosity during the Run 2 data-taking period of LHC. Reproduced from [63].

certain number of events with a given cross-section σ per unit of time is quantified by the instantaneous luminosity, defined as

$$\mathcal{L} = \frac{1}{\sigma}\frac{\mathrm{d}N}{\mathrm{d}t} \tag{2.1}$$

The LHC reached its design instantaneous luminosity value of $\mathcal{L} = 1.0 \cdot 10^{34}\ \mathrm{cm}^{-2}\mathrm{s}^{-1}$ in 2017, and exceeded it by 40 % with subsequent upgrades, becoming the world's highest luminosity hadron-collider [61]. The instantaneous luminosity varies over a run, for instance because of the number of protons per bunch that decreases with time, and also from one run to another one, so the useful quantity is the integrated luminosity

$$\mathcal{L}_{\mathrm{int}} = \int \mathcal{L}\ \mathrm{d}t \tag{2.2}$$

Indeed, when multiplied with a cross-section, it gives the expected number of events. The integrated luminosity delivered by the LHC over the Run 2 period is of 156 fb^{-1}. The barn b is the unit of cross-sections, and is defined as $1\ \mathrm{b} = 10^{-28}\ \mathrm{m}^2 = 100\ \mathrm{fm}^2$. Not all of these collisions were recorded by all the experiments, and not all collisions are of a high enough quality to be useful for physics purposes. For instance, ATLAS recorded 147 fb^{-1}, of which 139 fb^{-1} are useful for physics, as can be seen in Fig. 2.2. This last value has recently been re-evaluated as 140.1 fb^{-1} with a relative uncertainty of 0.83 % [62]. This is currently the most precise luminosity determination ever achieved at a hadron collider.

In between the runs, there were "long shutdowns" for LHC and detector upgrades, that allowed to increase security, collision energy and luminosity. New upgrades are planned after the end of the current Run 3 to increase the instantaneous luminosity by a factor 5 approximately, a phase called the High Luminosity LHC (HL-LHC), with the aim of recording 3 ab^{-1} of 14 TeV proton-proton collisions.

2.1.2 The acceleration infrastructure

Architecture

Before entering into the LHC to be accelerated to their final energy of 6.5 TeV, protons are accelerated by multiple other accelerators. The injection chain is shown in Fig. 2.3. Since the strength of the magnets must be increased when the energy of the particles increases, making use of the previously built accelerators is very useful because it allows the different elements to have a smaller dynamical range, compared to having less accelerating elements. This reduces the R&D complexity and the costs of the magnets and power supplies. Also, as the energy of the beam increases, its emittance decreases, and therefore its size decreases. Early accelerators must have a wide aperture whereas the LHC has a small aperture.

The injection chain comprises the following elements:

1. The proton source: particles are extracted by ionization of hydrogen in a Duoplasmotron Proton Ion Source. In 2020, this has been replaced by a negative hydrogen ions H^- source.
2. The protons first enter a linear accelerator, the Linac 2, to reach an energy of 50 MeV. In 2020, this has been replaced by the Linac 4 to accelerate the H^- to an energy of 160 MeV, before stripping both electrons to obtain a proton.

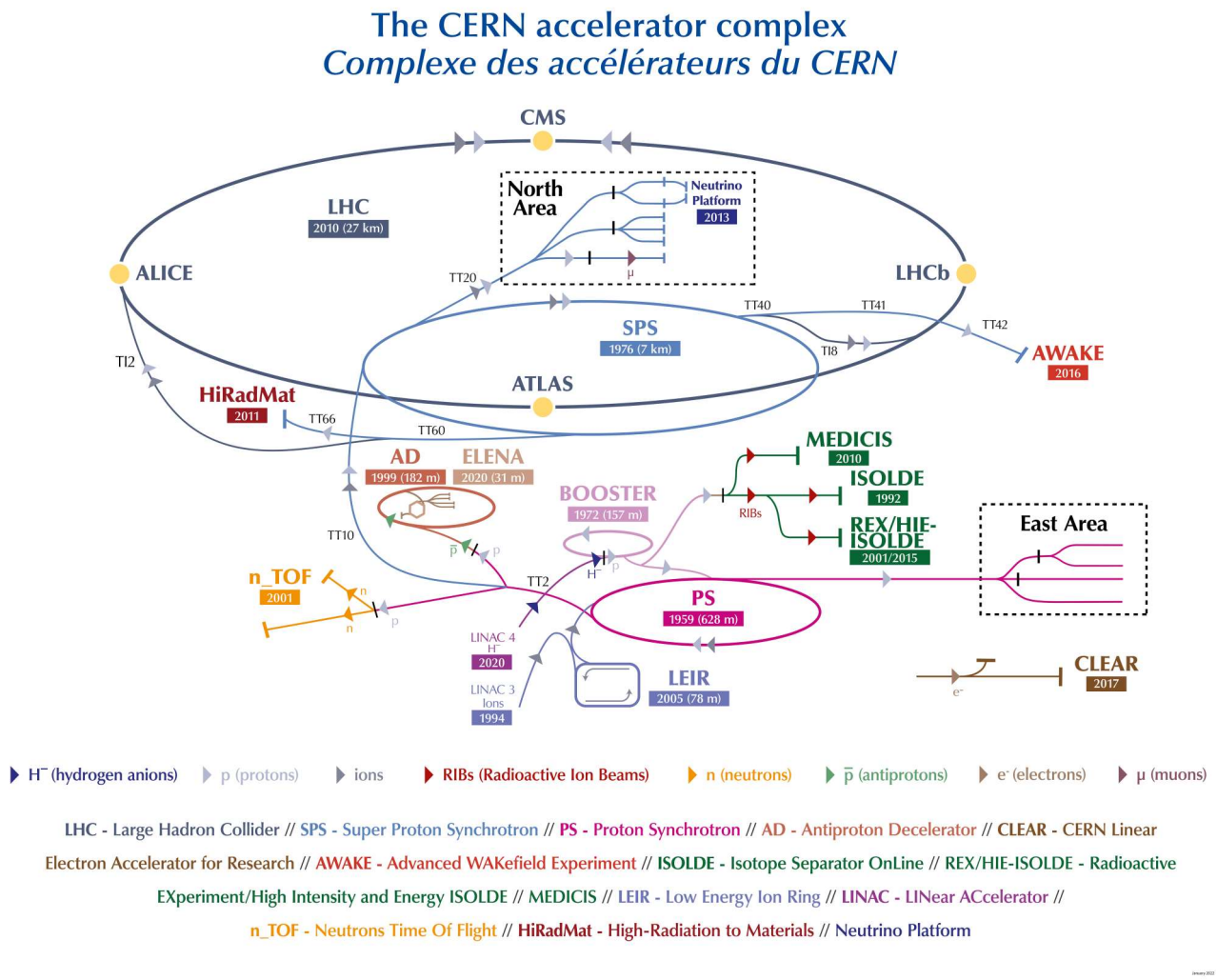


Figure 2.3: CERN accelerator complex, showing the LHC injection chain. Reproduced from [57].

3. The Proton Synchrotron Booster (PBS) accelerates the protons to 2 GeV.
4. The Proton Synchrotron (PS) accelerates them to 26 GeV. It is composed of 4 rings stacked onto each other to save space. The rings use pulsed magnets which allow to quickly change the beam configuration: 8 proton bunches are split into 84 bunches.
5. The Super Proton Synchrotron (SPS) accelerates the protons to 450 GeV.
6. Proton bunches are finally injected into the Large Hadron Collider (LHC), and accumulated until 2808 of them compose the beam, an operation that takes several minutes. Then the protons are accelerated to their final energy of 6.5 TeV, taking about 20 minutes. This energy corresponds to 99.9999991 % of the speed of light in the vacuum.

The LHC roughly has a circular shape, of length 27 km approximately, and is divided into octants with straight sections, as can be seen in Fig. 2.4:

- The Radio-Frequency (RF) cavity accelerates protons.
- Two cleaning sections, with diverse collimation systems, to correct the shape of the bunches.
- Four interaction points where the main particle detectors are located:
  - ATLAS and CMS, dedicated to the study of the Standard Model, to the discovery and study of the Higgs boson, to the teraelectronvolt scale, and to the search for new particles.
  - LHCb, dedicated to the study of B mesons, flavor physics, CP violation and rare decays.
  - ALICE, dedicated to the study of the quark-gluon plasma from collisions of heavy ions.
- The beam dump, which is the only element of the LHC capable of absorbing the energy of the proton bunches without damage.

In between those straight sections, there are arc sections to bend the trajectory of the protons.

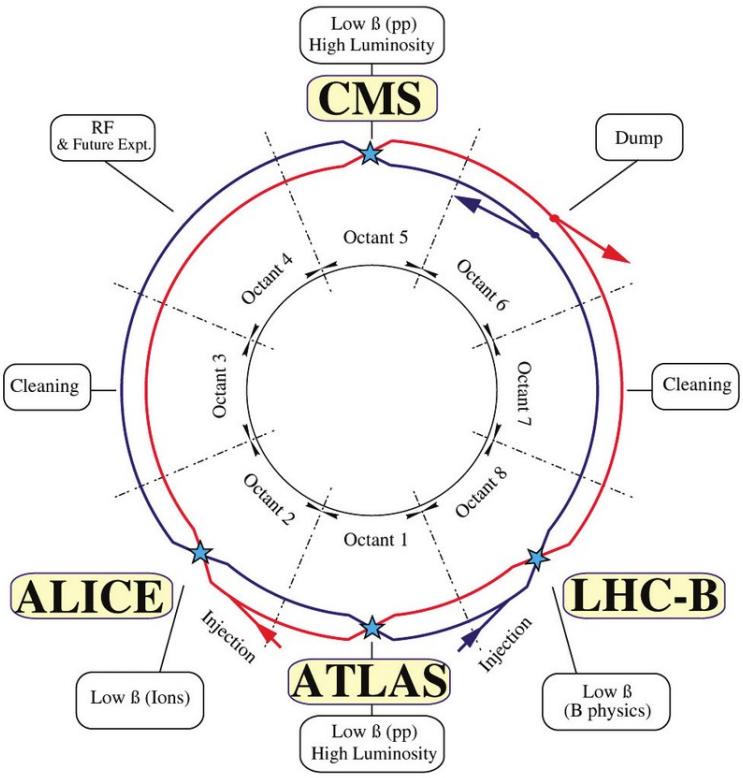


Figure 2.4: LHC layout with its instrumentation and the main four experiments in the different octants. Reproduced from [60].

Elements of the LHC

There is a total of 9593 superconducting magnets in the LHC, including 1232 main dipoles used to bend the trajectory of the protons, shown in Fig. 2.5. Another 392 quadrupole magnets keep the beam focused. More quadrupoles increase the chances of interactions by focusing the beam, and then defocusing it to re-enter the LHC after the interaction point; the beam size is shown in Fig. 2.6 for the case of the ATLAS interaction point. Other quadrupoles or higher multipole orders correct for various types of imperfections. The main dipole magnets generate a field of 8.3 T. They are based on NiTi RUTHERFORD cables, cooled to a temperature of 1.9 K with superfluid helium, and the electric current in the magnets is of 12 kA approximately.

The injected beam is accelerated with a 400 MHz superconducting cavity system, shown in Fig. 2.7. The high voltage of the oscillating electric field of 8 MV in coast at injection and 16 MV at 7 TeV is generated by a klystron, one per cavity. Complex feedback loops around each cavity allow precise control of the field in each cavity. Additional systems allow to dampen the injection errors.

The beam vacuum has to be lower than 10^{-11} mbar, to have as little beam-gas collisions as possible. This is important to avoid the decay of the beam intensity, the energy lost in cryomagnets, and the noise to the experiments. Another vacuum is necessary, to insulate the cryomagnets and the helium distribution, but the requirements are less stringent: 10^{-6} mbar. During the LHC operation, the beam pipe vacuum has to be actively maintained, due to the synchrotron radiations that extract electrons from the beam pipe, creating an electron cloud that accumulates and shields the beam from the magnetic elements. This effect is mitigated by multiple elements: a beam-screen, along with surface conditioning with scrubbing runs; the geometry of the beam pipe; and the bunch repetition patterns.

The beam dumping is ensured by a dedicated system, shown in Fig. 2.8, able to extract the beam from the beam pipe extremely fast. It does so with fast kicker magnets, that deflect the beam to the outside, and septum magnets that deflect them even more. Then kicker magnets dilute the beam, that is disposed of in beam dump blocks, that are 8 m long graphite blocks shielded by concrete. The beam is swept over the dump block surface in an 'e' shape with the dilution kicker magnets, to lower the power density, otherwise the beam could drill a hole a few meters deep into the block.

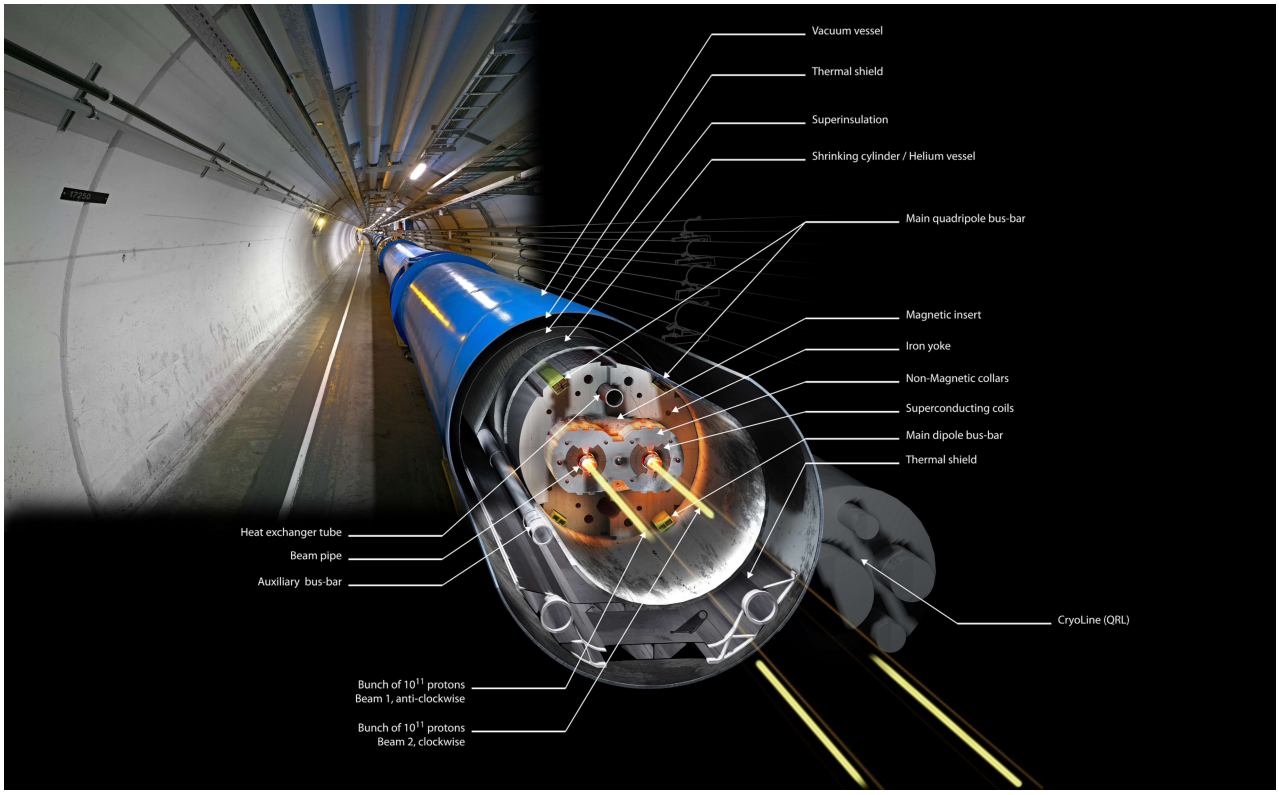


Figure 2.5: One of the 1232 LHC dipole magnets used to bend the trajectory of the protons around the LHC ring. Reproduced from [57].

2.1.3 The beam of the Large Hadron Collider

Beam characteristics

Assuming two Gaussian bunches colliding head-on, the luminosity writes:

$$\mathcal{L} = \frac{N_1 N_2 f_{\text{rev}} n_b}{4\pi\sigma_x^* \sigma_y^*}\mathcal{F} \tag{2.3}$$

with N_1 and N_2 the number of protons in each of the two colliding bunches, f_{rev} the revolution frequency of the two beams, n_b the number of bunches per beam, σ_x^* and σ_y^* the (root mean square) transverse beam sizes in the horizontal and vertical directions at the interaction point, and $\mathcal{F}$ is a factor of order 1 that takes into account multiple effects: inefficient geometric overlapping of the beam due to the crossing angle and finite bunch length, and dynamic effects.

In standard runs, the beam has the following characteristics: $n_b = 2808$ bunches per beam, A number of protons per bunch of$N_1 \approx N_2 \approx 1.15 \cdot 10^{11}$, $f_{\text{rev}} \approx 11$ Hz, $\sigma_x^* \approx \sigma_y^* \approx 8.5$ µm. The time span between collisions is 25 ns. The bunch length is $\sigma_z \approx 8$ cm.

Another parametrization of the luminosity function is often used. Due to LIOUVILLE's theorem, the volume occupied in phase space by a system of particles (here, by the two colliding bunches) is constant, and is called the emittance ε. In trace space (*i.e.* looking only at one Cartesian coordinate and its derivative, like x and x'), the beam size is related to the emittance by the β function: $\beta = \frac{\sigma^2}{\varepsilon}$. Magnets do not affect the emittance, they can only affect the bunch shape. The emittance only changes with particle acceleration, due to the change in longitudinal momentum, so the emittance at 7 TeV is actually a function of the emittance when the bunches first enter the acceleration chain (which is the reason for the new source of H^- ions: they have a higher emittance). When quadrupoles are used to focus the beam for the collision, the bunches are non-Gaussian, they have an hourglass shape, and the β value has to be taken at the waist because this is the interaction point, in which case it is written β^*. Due to the acceleration of the protons near the speed of light, relativistic effects are to be taken into account when defining the phase space volume: the normalized emittance is $\varepsilon_n = \gamma\varepsilon$ with γ the LORENTZ factor and ε the geometric emittance. With this parametrization, and assuming that the beam has the same size in the x and y directions, the luminosity is expressed:

$$\mathcal{L} = \frac{N_1 N_2 f_{\text{rev}} n_b \gamma}{4\pi\varepsilon_n \beta^*}\mathcal{F} \tag{2.4}$$

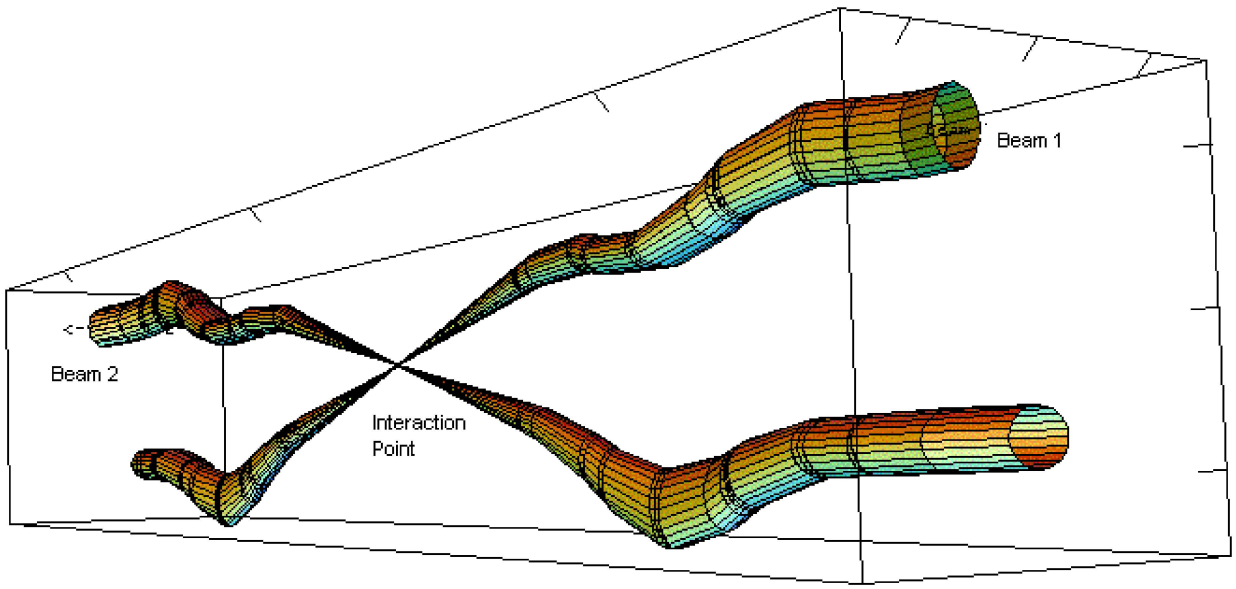


Figure 2.6: Relative beam sizes around the interaction point IP1, where the ATLAS experiment is located. Reproduced from [57].

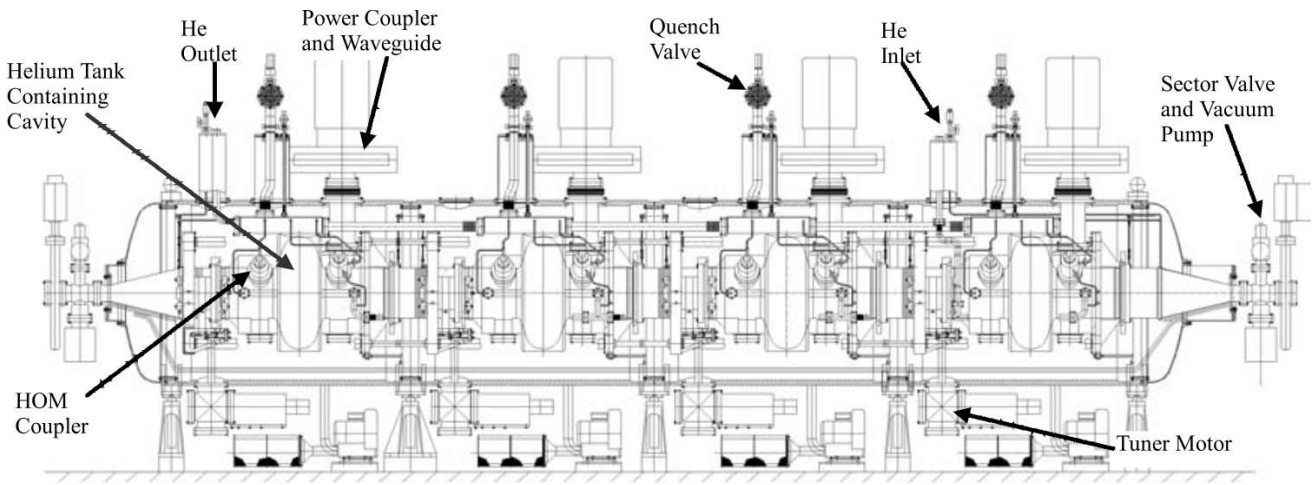


Figure 2.7: Four-cavity Radio Frequency (RF) cryomodule of the LHC. The klystrons are not shown, they are connected to the RF cavities by the power couplers. Reproduced from [60].

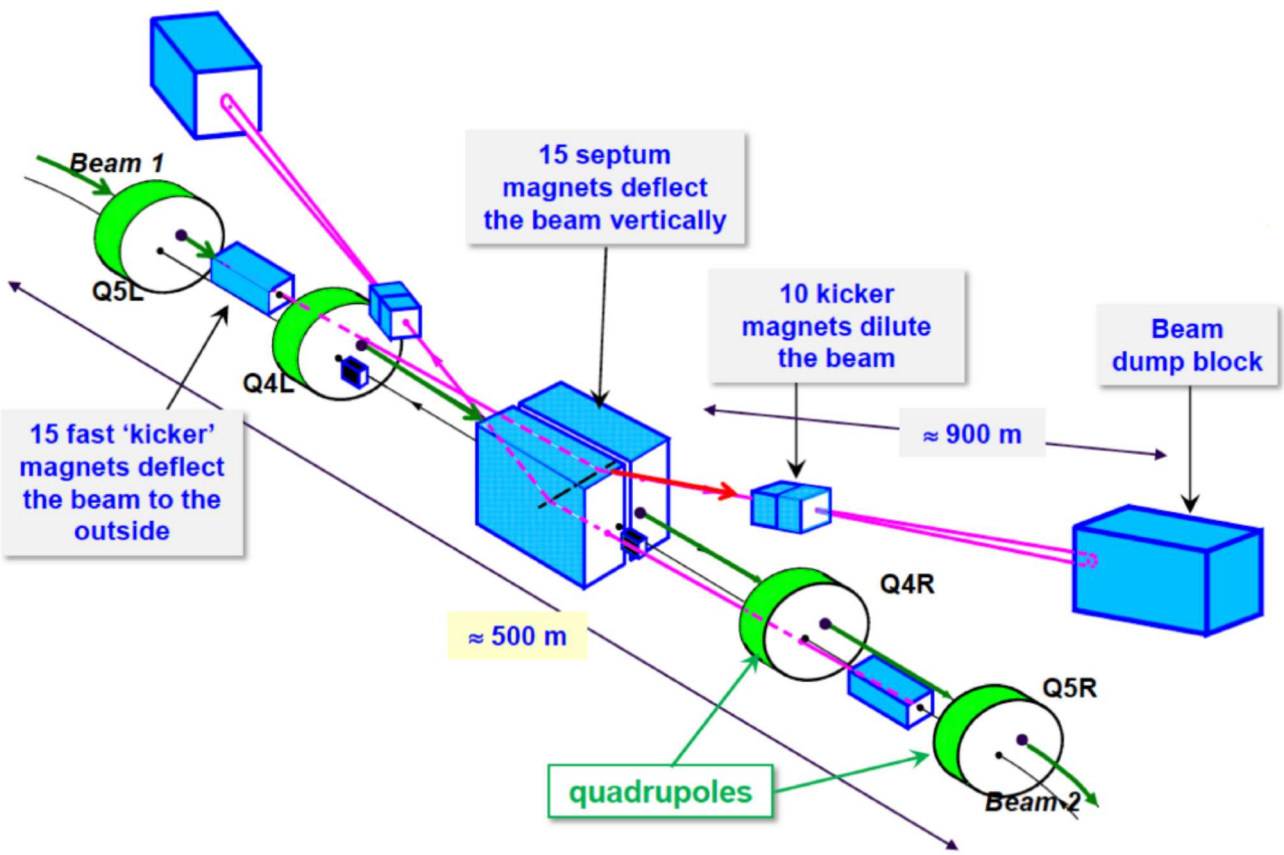


Figure 2.8: Beam dump system of the LHC. Reproduced from [59].

A typical value of β^* is between 0.15 m and 0.3 m, and the nominal normalized transverse emittance for the LHC is $\varepsilon_n = 3.75$ µm.

Of course, many more dynamic effects are taken into account to control the beam, like the electromagnetic wake field, the betatron oscillations (the beam focusing is done alternatively in the x and y directions, hence the bunch transverse sizes oscillate), the crossing angle (bunches do not collide head-on), intra-beam COULOMB scattering, the TOUSCHEK effect (intra-beam hard scattering), *etc.*

Bunch filling scheme

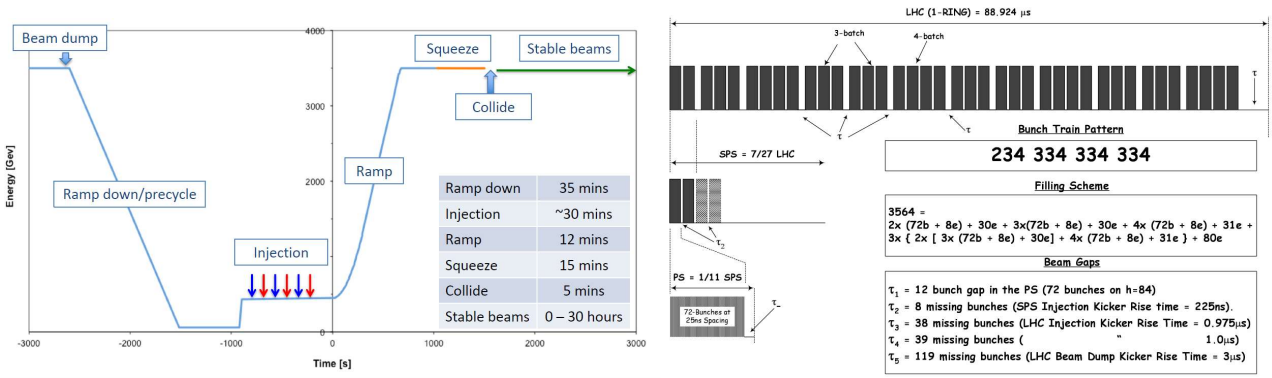
The operation cycle is shown in Fig. 2.9a. The beam preparation takes about an hour, including beam filling, ramp up and squeezing. Then the collisions for physics experiments can start, and last as long as the beams are stable, without any risk for the detector or for the experiments, and while there are enough protons in the bunches to generate enough collisions.

The number of collisions per bunch crossing depends on the instantaneous luminosity, and can be adjusted. The β^* parameter can be tuned during the run, to adjust the desired luminosity and therefore the expected number of collisions per bunch crossing. A similar effect is reached with the adjustment of the crossing angle, that modifies the $\mathcal{F}$ parameter: the crossing angle decreases from 320 µrad to 260 µrad during a run, for luminosity leveling. This will be extensively used during the HL-LHC phase: the peak luminosity is designed to be $2 \cdot 10^{35}$ cm^{-2}s^{-1}, but the number of collisions per bunch crossing would be very high, and the luminosity would not be constant, making it harder for physics analyses. Therefore, luminosity will be leveled to $5 \cdot 10^{34}$ cm^{-2}s^{-1}, to limit the number of collisions per bunch crossing around 140.

There are actually 3564 possible bunch positions in the circumference of an LHC ring [64], 2808 of which are filled with the "25 ns filling scheme" shown in Fig. 2.9b. This leaves empty spaces to satisfy a few constraints, mainly the time needed to allow the injection kicker magnet and dump kicker magnet to rise, or the four-fold symmetry to reduce the beam-beam interactions. The bunches are separated by a 25 ns spacing from the SPS, and grouped in packets of 72 bunches, themselves arranged in the 234 334 334 334 train pattern.

Beam monitoring

Finally, one of the most crucial aspects of the LHC operation is the beam monitoring. Indeed, each of the two



(a) (b)

Figure 2.9: **(a)** LHC operation cycle. Reproduced from [59]. **(b)** Bunch filling scheme of the LHC beam for the 25 ns scheme. *b* indicates a position with beam and *e* an empty position. Reproduced from [64].

beam carries 360 MJ, which is roughly the energy released by the explosion of 100 kg of TNT. Only one element of the LHC is capable of absorbing that much energy: the beam dump. Losing control of the beam might lead to a complete destruction of the LHC. At these energies, small beam losses can lead to tremendous damage, for instance by heating a superconducting magnet above its critical temperature, thus releasing all its energy causing a cascade reaction and other subsequent damage. A similar incident occurred in 2008, though not involving any beam: a faulty junction between two magnets dissipated too much heat, leading to the quench of a magnet that released 400 MJ of its total 600 MJ energy, creating electric arcs that punctured the insulating vacuum, releasing 6 t of helium into the insulating vacuum. The pressure wave propagated too fast, ten times faster than what the pressure valves were designed to handle, thus propagating from one dipole to another one, until finally reaching a stronger pressure valve capable of handling the pressure. This sudden increase in pressure displaced some magnets by 50 cm, broke the beam vacuum over 2 km. This event led to the replacement of an entire LHC section, and to the overall increase in monitoring and security.

There are numerous beam monitors along the LHC, and a sophisticated control system to take the decision to dump the beam as fast as possible in case of any danger signal. A non-exhaustive list of monitoring devices is:

- Beam profile monitoring, using wire-scanners. A thin (33 μm in diameter) carbon wire is inserted into the beam during 1 ms approximately, colliding with the beam and thus generating secondary particles that are collected and analyzed.
- The synchrotron light is collected near the RF zone. It is amplified and collected by a camera, allowing to monitor the beam size (integrated over 4 turns of the beam in the LHC).
- Beam loss monitors are ionization chambers with a very fast reaction time (40 μs, *i.e.* half a turn) and a very large dynamic range. About 3600 of them are distributed over the LHC.

2.2 The ATLAS detector

ATLAS is one of the four main experiments that exploit particle collisions produced by the LHC.[2] Its acronym stands for "A Toroidal LHC ApparatuS", due to its huge toroidal magnetic field for the muon chambers, as described hereafter. It is a general-purpose detector, able to detect almost all types of long-lived Standard Model particles: photons, electrons, muons, hadrons, with the exception of neutrinos that do not interact strongly enough with matter to be detectable with those technologies. It has a wide angular coverage, 2π in longitude around the beam axis and from -4.9 to 4.9 in pseudo-rapidity (see definition hereafter).

The main two goals for the design of this detector were the detection of the Higgs boson, that was known to lie in the mass range between a few GeV to about 1 TeV, and supersymmetric particles. To that extent, having estimates for the energy and luminosity at which the LHC would operate, a few detector concepts were proposed in the late 1980s, and two projects involving large toroidal magnet configurations were merged in 1992 to become ATLAS. The letter of intent was submitted in October 1992 to the CERN LHC Experiments Committee (LHCC), followed in 1994 by a technical proposal for the various detector components, followed in the next ten years by further optimizations and cost reductions. Approved in 1996, the construction began, with pieces constructed in laboratories from all over the world.

[2] This section is mostly based on [1, 65–68].

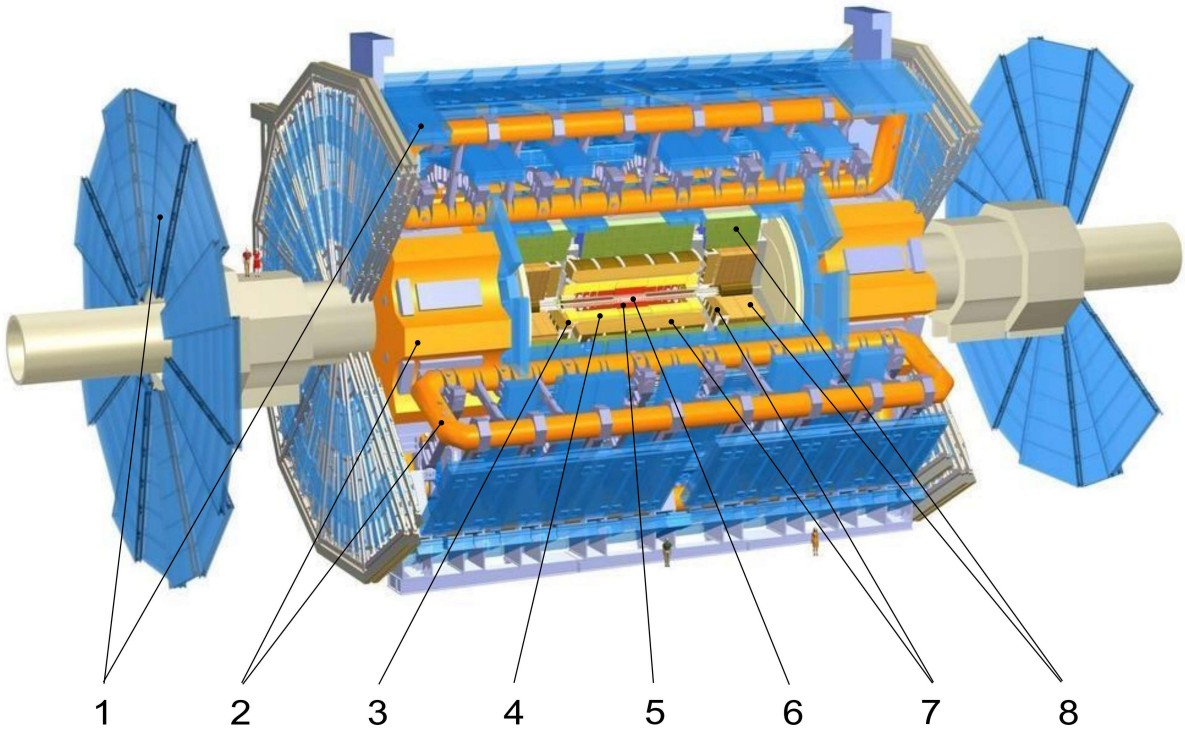


Figure 2.10: Cut-away view of the ATLAS detector showing its various components. (1) Muon spectrometer, including its barrel and end-caps. (2) Toroid magnets. (3) Solenoid magnet. (4) Transition Radiation Tracker. (5) Semi-Conductor Tracker. (6) Pixel Detector. (7) Liquid argon electromagnetic calorimeter. (8) Hadronic calorimeters: tile calorimeters in the barrel, liquid argon end-cap and forward calorimeters. Reproduced from [57].

The ATLAS detector is exploited by the ATLAS Collaboration, which nowadays consists of about 5500 people from more than 180 institutions and 38 countries, of whom 44 % are physicists, 22 % are Ph.D. students, 19 % are engineers and technical staff (in January 2018) [66]. The number of people who author ATLAS Collaboration publications total approximately 2900 people.

The ATLAS detector, shown in Fig. 2.10, is the largest particle detector at CERN: it is 44 m long, 25 m in diameter, and weighs about 7000 t. Considering all the different particles that can reach the detector, ATLAS is composed of multiple sub-detectors, each specialized in one type of particle, one type of interaction, or one property of those particles. ATLAS is organized in nested sub-detectors, with (nearly) cylindrical and forward-backward symmetries. The sub-detectors are shown in Fig. 2.11 with the type of particle that they are designed to detect; they are detailed in the following sections and listed from the closest to the interaction point to the furthest:

- The Inner Detector (ID) is the closest detector to the beam pipe. It is immersed in a 2 T magnetic field, that bends the trajectory of electrically charged particles (electrons, positrons, protons, charged pions, *etc*). Those particles make multiple small energy deposits ("hits") along their trajectory, that allow to reconstruct their electric charge to momentum ratio, as well as their impact parameters. It is also crucial for vertex reconstruction, which is needed to discriminate between the collision of interest and pile-up, or for flavor-tagging algorithms for instance.
- The Electromagnetic CALorimeter (ECAL) and Hadronic CALorimeter (HCAL) aim at measuring the energy of particles, respectively electromagnetic particles (photons, electrons, pions, *etc*) and hadrons (protons, neutrons, pions, *etc*.) They have a coarser granularity than the ID, but a good energy resolution. The shower development inside the detector also allows for particle identification.
- The Muon Spectrometer (MS) allows for the identification of muons and the measurement of their momenta. Indeed, those particles only deposit a small fraction of their energy in the previous detector due to their high mass, making them low-ionizing particles. Complete energy measurement is not possible, as they are not fully absorbed in the MS.

The main active detector components of the ATLAS detector are given in Table 2.1, with the total number of channels and the pseudorapidity coverage.

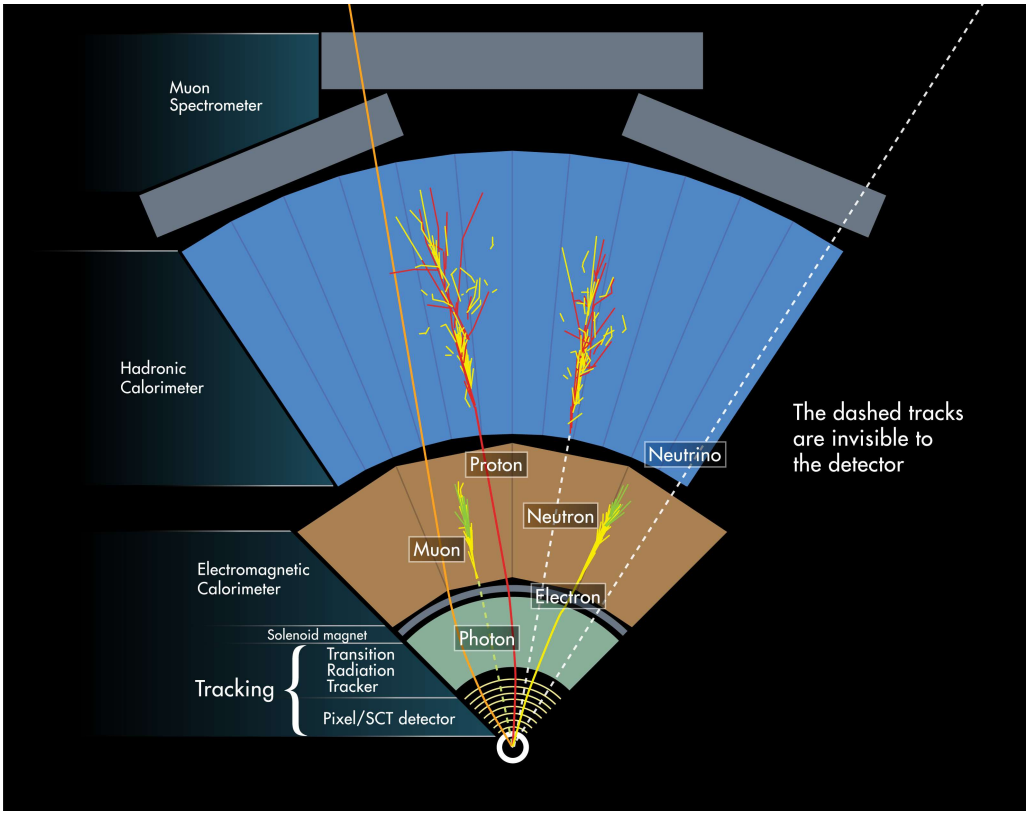


Figure 2.11: Paths of different particles entering the ATLAS detector. Electrically charged particles have a bent trajectory in the tracking systems due to the 2 T magnetic field provided by the central solenoid magnet. The charged particles produce small energy deposits along their trajectories in the tracking systems, allowing to measure their electric charge to momentum ratio. Photons, electrons, positrons and electrically charged hadrons produce energy deposits in the electromagnetic calorimeter, allowing for a measurement of their energies. Hadrons, like protons or neutrons, produce energy deposits in the hadronic calorimeter, allowing for a measurement of their energies. Muons produce small energy deposits in the muon spectrometer, along a curved trajectory due to the 4 T magnetic field provided by the barrel and end-cap toroids, allowing for a measurement of their momenta. Neutrinos do not interact with the ATLAS detector. Reproduced from [57].

2.2.1 System of coordinates

ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point (IP) in the center of the detector, as shown in Fig. 2.12. The z-axis is along the beam pipe, the x-axis points from the IP to the center of the LHC ring, and the y-axis points upwards. Spherical coordinates (r, θ, ϕ) are used: $\theta \in [0, \pi]$ is the polar angle with respect to the transverse plane (x, y), $\phi \in [0, 2\pi[$ is the azimuthal angle around the z-axis.

Let $(E, \vec{p})$ be the four-momentum of a particle with E its energy and $\vec{p} = (p_x, p_y, p_z)$ the momentum of the particle. The rapidity of a particle is a measure of its relativistic velocity, which is natural as Lorentz boosts are hyperbolic rotations of the space-time coordinates:

$$w = \operatorname{atanh} \beta = \frac{1}{2} \ln \frac{E + |\vec{p}|}{E - |\vec{p}|} \tag{2.5}$$

with $\beta = \frac{v}{c}$ the velocity of the particle (in natural units). In particle physics, another definition is more commonly used, where the rapidity is defined relative to the beam axis:

$$y = \frac{1}{2} \ln \frac{E + p_z}{E - p_z} \tag{2.6}$$

The rapidity y is closely related to the pseudo-rapidity η, that is defined as follows with respect to the polar angle θ as:

$$\eta = -\ln \left(\tan \frac{\theta}{2} \right) \tag{2.7}$$

Detector component	Position	Channels (total)	η coverage
Tracking			
Pixel B-layer (IBL, added for Run 2)	1 cylindrical barrel layer	6 million	$\lvert\eta\rvert < 2.5$
	Average radius 33 mm		
Pixel	3 cylindrical barrel layers	80.4 million	$\lvert\eta\rvert < 2.5$
	3 end-cap disks on each side		
	Radial envelope 45.5 to 242 mm		
SCT strips	4 cylindrical barrel layers	6.3 million	$\lvert\eta\rvert < 2.5$
	9 end-cap disks on each side		
	Radial envelope 251 to 610 mm		
TRT	73 barrel straw planes	351'000	$\lvert\eta\rvert < 2.5$
	80 end-cap straw planes		
	Radial envelope 554 to 1106 mm		
Calorimetry			
EM presampler	Barrel	7'808	$\lvert\eta\rvert < 2.52$
	End-caps	1'536	$1.5 < \lvert\eta\rvert < 1.8$
EM calorimeter	3 depth samples barrel	101'760	$\lvert\eta\rvert < 1.48$
	3 depth layers end-caps	62'208	$1.375 < \lvert\eta\rvert < 3.2$
Hadronic tile calorimeter	3 depth samples barrel	5'760	$\lvert\eta\rvert < 1.0$
	3 depth samples extended barrel	4'092	$0.8 < \lvert\eta\rvert < 1.7$
LAr hadronic end-caps	4 depth layers	5'632	$1.5 < \lvert\eta\rvert < 3.2$
LAr forward hadronic calorimeter	3 depth layers	3'524	$3.1 < \lvert\eta\rvert < 4.9$
Muon spectrometer			
MDT precision tracking	3 multi-layer stations	354'000	$\lvert\eta\rvert < 2.7$
CSC precision tracking	1 innermost station end-caps	31'000	$2.0 < \lvert\eta\rvert < 2.7$
RPC trigger chambers	2 multi-layer stations barrel	373'000	$\lvert\eta\rvert < 1.05$
TGC trigger chambers	2 multi-layer stations end-cap	318'000	$1.05 < \lvert\eta\rvert < 2.4$

Table 2.1: The main active detector components of the ATLAS detector, from the beam line towards the outside. The total readout channels for each component is given, as well as its pseudorapidity coverage. Reproduced from [69].

Some pseudorapidity values are shown in Fig. 2.13 with their correspondence to the polar angle. Most importantly, $\eta = 0$ in the transverse plane and $\eta = \pm\infty$ along the beam axis. For an ultra-relativistic particle, *i.e.* in the massless particle limit ($E \gg m$), the pseudorapidity is equal to the rapidity. More precisely, the relationship between rapidity and pseudo-rapidity for a massive particle is:

$$y = \ln\left(\frac{\sqrt{m^2 + p_T^2 \cosh^2\eta} + p_T \sinh\eta}{\sqrt{m^2 + p_T^2}}\right) \tag{2.8}$$

The transverse momentum p_T is the component of the momentum in the transverse plane. The momentum coordinates are expressed as follows:

$$p_T = \sqrt{p_x^2 + p_y^2} = |\vec{p}|\sin(\theta) = \frac{|\vec{p}|}{\cosh(\eta)} \tag{2.9}$$

$$p_x = p_T \cos(\phi) \tag{2.10}$$

$$p_y = p_T \sin(\phi) \tag{2.11}$$

$$p_z = \frac{p_T}{\tan(\theta)} = p_T \sinh(\eta) \tag{2.12}$$

Angular distances between two particles are defined in a Lorentz-invariant way as

$$\Delta R = \sqrt{(\Delta y)^2 + (\Delta\phi)^2} \underset{\substack{\text{massless}\\ \text{particle}}}{\approx} \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} \tag{2.13}$$

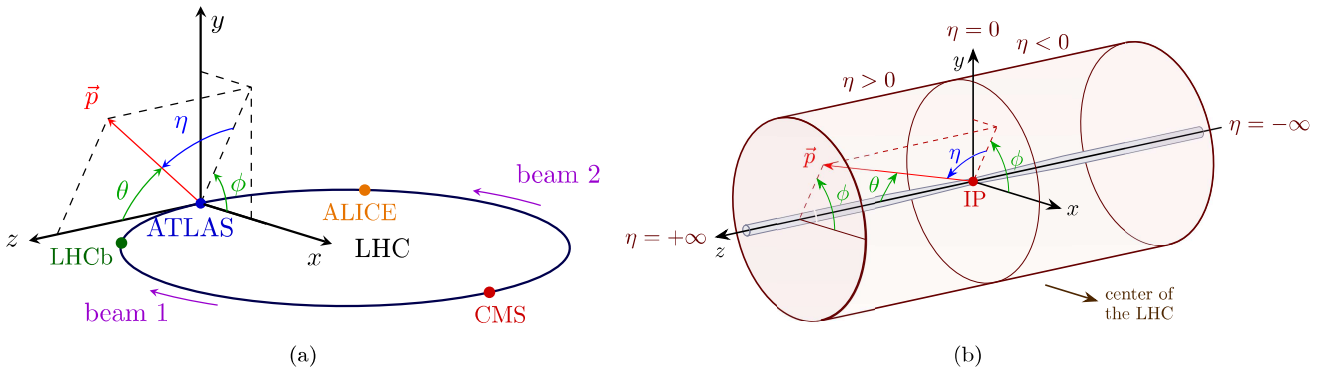


Figure 2.12: LHC scheme and coordinate system for the ATLAS detector. Reproduced from [70] and slightly adapted.

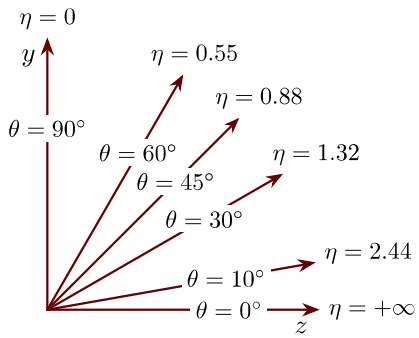


Figure 2.13: Correspondence between the pseudorapidity η and the polar angle θ. Reproduced from [70].

2.2.2 Particle energy loss and shower development

In ATLAS, particle detectors are essentially an absorbing material that converts energy deposited by a particle to an electric signal. The signal is read and amplified before being converted to numerical values and analyzed. Characterizing the absorption of particles by materials is essential to design detectors and to understand the properties of the detected particles. When a particle passes through matter, the interaction depends on the type of particle, on its energy and on the type of material. At low energy, a particle will mostly see matter as atoms and ionize them via electromagnetic interaction, while at high energies it will interact through the strong nuclear force with smaller constituents like the nucleons or even with quarks and gluons. Other detection techniques exist, like the transition radiation, used in ATLAS in the Inner Detector, or the Cherenkov radiation.

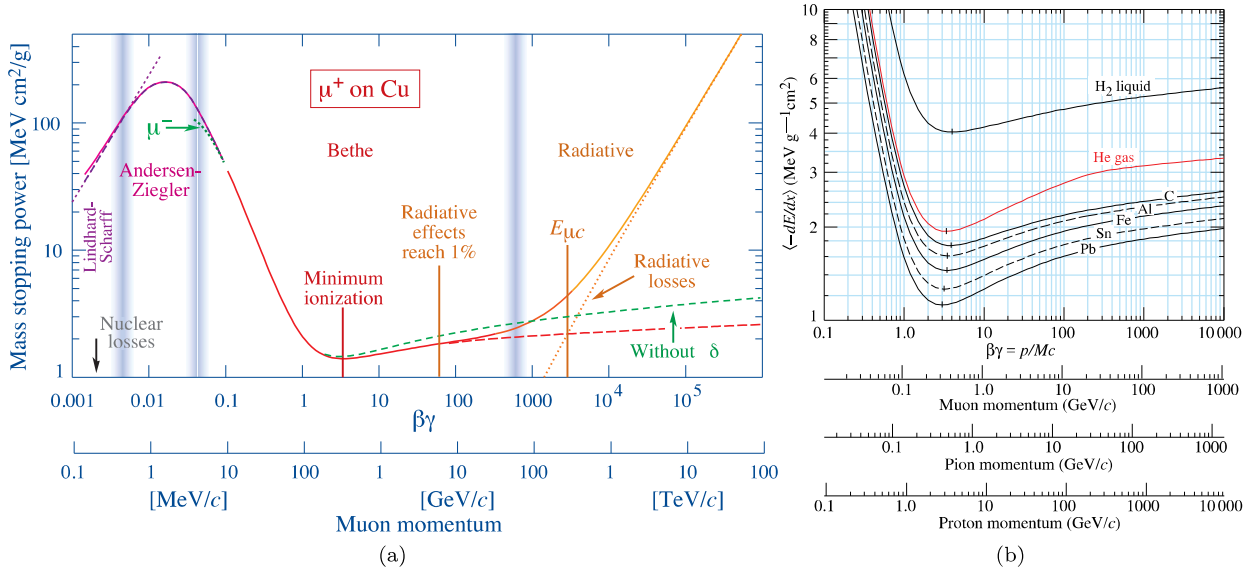
The radiation length

Matter is essentially characterized by two lengths: its radiation length X_0 and its nuclear interaction length λ [1]. X_0 is the mean distance covered by an electron before its first electromagnetic interaction, alternatively defined as the distance over which an electron loses 63 % of its energy [3]. λ is the mean distance traveled by a hadronic particle before undergoing an inelastic nuclear interaction. For orders of magnitude, lead is characterized by $X_0 = 5.6$ mm and $\lambda = 17$ cm, and aluminum is characterized by $X_0 = 8.9$ mm and $\lambda = 39$ cm. For a pure material, with atomic number $Z > 4$ and atomic mass A (assigned unit g $\cdot$ mol^{-1}), the radiation length can be estimated with Dahl's formula which is accurate to few percent [3]:

$$X_0 = \frac{716.4\mathrm{A}}{\mathrm{Z}(\mathrm{Z}+1)\ln\left(\frac{287}{\sqrt{\mathrm{Z}}}\right)}\ \mathrm{g}\cdot\mathrm{cm}^{-2} \tag{2.14}$$

Electrically charged particles

The main types of electromagnetic interactions of charged particles are ionization, scattering and brehmsstrahlung. Ionization is a Coulomb interaction with the electrons of the atom, where the particle transfers enough energy to extract an electron from its orbital. For scattering, the Coulomb interaction occurs with the nucleus of the atom, that mostly deflects the electron from its trajectory. This occurs multiple times, and the deflection angle is



(a) (b)

Figure 2.14: Mass stopping power $\left\langle -\frac{\mathrm{d}E}{\mathrm{d}x} \right\rangle$ for **(a)** anti-muons in copper, with the vertical lines indicating the boundaries between the different regimes; and **(b)** for muons, pions and protons in various materials. Reproduced from [3].

roughly $\theta \approx \frac{14\ \mathrm{MeV}}{p}\sqrt{\frac{d}{X_0}}$, that decreases with increasing momentum p (and d is the distance traveled in matter). Brehmsstrahlung is a radiative process by which an electron emits a photon and loses energy. Brehmsstrahlung dominates at high energy, for instance above 10 MeV for electrons and positrons, and above 100 GeV for muons and anti-muons.

The energy losses are characterized by the mean rate of energy loss $\left\langle -\frac{\mathrm{d}E}{\mathrm{d}x} \right\rangle$, also called the "mass stopping power", shown for different particles in different materials in Fig. 2.14. It can be expressed analytically for moderately relativistic particles ($0.1 \lesssim \beta\gamma \lesssim 1000$) with the Bethe equation [3]. To first order, the mass stopping power only has a weak material dependency, as can be observed in Fig. 2.14b, and it mostly depends on the charge of the particle and not on its mass. In the low-energy regime ($\beta\gamma < 1$), the energy loss increases at β^{-2}, meaning that slow particles are more ionizing than fast particles. In the ultra-relativistic regime ($\beta\gamma > 1000$), brehmsstrahlung dominates. In between, ionization increases logarithmically above $\beta\gamma \approx 1$ due to the increase in electric field. Around $\beta\gamma \approx 3$, *i.e.* at energies typical of those generated at the LHC, ionization is minimal, and particles at that energy are referred to as Minimum Ionizing Particles (MIP).

Photons

Photons interact with matter in a few different manners shown in Figs. 2.15a and 2.15b, mostly depending on their energy [3, 71]:

- Atomic photoelectric effect (electron ejection, photon absorption), with $\sigma_{\mathrm{p.e.}} \propto Z^{4-5}$.
- Coherent Rayleigh scattering, where the atom is neither ionized nor excited, with $\sigma_{\mathrm{Rayleigh}} \propto Z^2$.
- Incoherent Compton scattering off an electron, with $\sigma_{\mathrm{Compton}} \propto Z$.
- Pair production in a nuclear field, that requires an incident energy greater than 1.022 MeV, with $\kappa_{\mathrm{nuc}} \propto Z^2$.
- Pair production in an electron field, that requires an incident energy greater than 2.044 MeV, with $\kappa_{\mathrm{e}} \propto Z^2$.
- Photonuclear interactions, like the giant dipole resonance, where the target nucleus is usually broken up, $\sigma_{\mathrm{g.d.r.}}$

As shown in Figs. 2.15a and 2.15b, the photon interaction with matter is dominated by the photoelectric effect at low energy, by electron-positron pair production at high energy, and by Compton scattering in between. Above a few dozen MeV, pair production dominates. Typical energies for photons produced at the LHC are above this threshold, which is also the energy at which brehmsstrahlung dominates for electrons. This leads to a cascade phenomenon called an electromagnetic shower: a photon creates an electron-positron pair, that themselves radiate photons, and the process repeats until all particles reach an energy below the threshold of pair production or brehmsstrahlung. Electromagnetic showers occur not only for photons but also for electrons, and have similar shapes, therefore identification is done thanks to the tracking systems: if the shower is matched to a track, the absorbed particle is an electron (or a positron), otherwise it is a photon. In the case of a converted photon, *i.e.* a photon that created an electron-positron pair upstream of the calorimeter, in which case there are tracks,

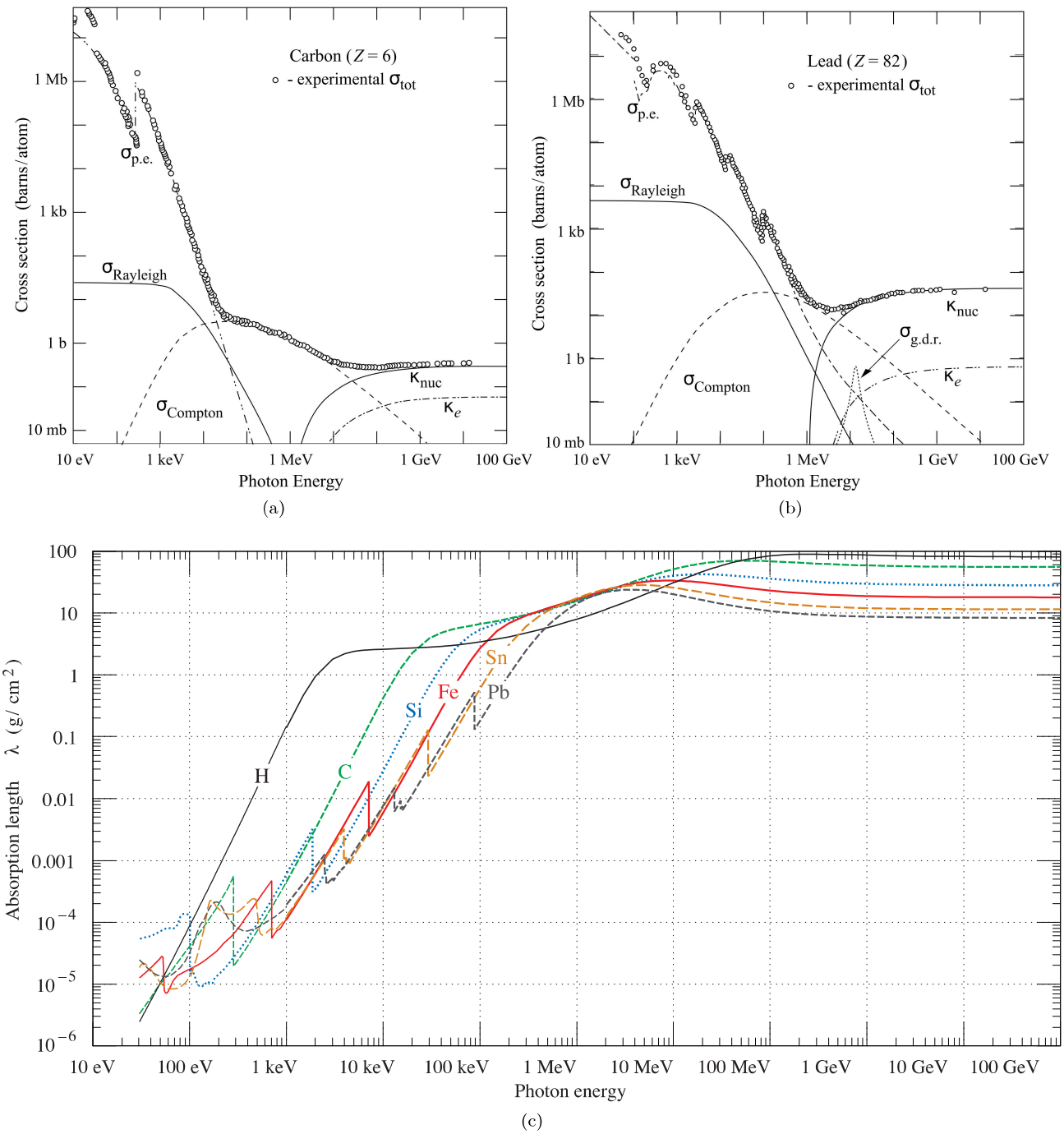


Figure 2.15: **(a,b)** Photon total cross sections as a function of energy in **(a)** carbon and **(b)** in lead, showing the contributions of different processes. **(c)** The photon absorption length $\lambda\rho$ (not λ as shown in the graph) for various elemental absorbers, as a function of photon energy. Reproduced from [3].

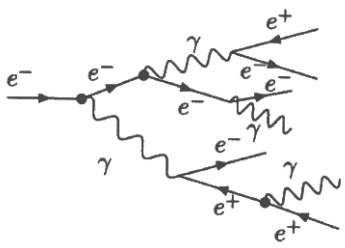


Figure 2.16: Representation of an electromagnetic shower. Reproduced from [1].

another identification mechanism is required; based on the shower shape and energy deposition characteristics. An electromagnetic shower starting from an electron is represented in Fig. 2.16.

When a high energy photon enters the electromagnetic calorimeter, it showers, and its energy is absorbed in a characteristic distance λ (not to be confused with the interaction length that characterizes hadronic interactions) that depends on the material, as shown in Fig. 2.15c (where the figure shows $\lambda\rho$ and not λ, with ρ the material density). This is obtained with $\lambda\rho = \frac{M}{\mathcal{N}_A \sigma_{\text{tot}}}$, where M is the molar mass, $\mathcal{N}_A$ is AVOGADRO's constant, and $\sigma_{\text{tot}} = \sum_i \sigma_i$ is the total cross-section from all processes. For instance, at 1 MeV, $\lambda\rho \approx 15$ g $\cdot$ cm^{-2}, and therefore $\lambda = 1.2$ cm in lead and $\lambda = 9.2$ cm in graphite.

Hadrons

There are also multiple processes for hadronic interaction with matter, highlighted in Fig. 2.17: evaporation, fragmentation, fission, deep spallation, spallation, quasi-elastic scattering. In those processes, some energy can be deposited in the nucleus, leaving an excited state, or energetic particles can leave the nucleus, potentially colliding with other nuclei, thus leading to a cascade as depicted in Fig. 2.18.

At small time scales (of the order of 10^{-22} s), spallation occurs: the hadron interacts with the nucleons, an intra-nuclear cascade of p, n and π occurs, and particle production is possible. At longer time scales (of the order of 10^{-16} s), multiple processes compete in making the isotope stable: fission, where the nucleus splits into other nuclei with high A; fragmentation, where the nucleus emits a small nucleus; evaporation, where light fragments (like p, n, α) are emitted.

Shower characterization

The transverse and longitudinal profiles of the showers can help characterize the particle that initiated it [3].

For instance, for electromagnetic cascades, the transverse radius of the shower containing 90 % of the shower energy, called the MOLIÈRE radius R_M, can be expressed as:

$$R_M = E_s \frac{X_0}{E_c} \tag{2.15}$$

where $E_s \approx 21$ MeV and E_c is the critical energy for which the particle radiative energy loss is equal to its ionization energy loss, thus depending on the type of particle.

Moreover, the longitudinal profile of the energy deposition in an electromagnetic shower, at energies between 1 GeV and 100 GeV, can be characterized. An electromagnetic shower reaches its maximal energy deposition at a normalized depth:

$$t_{\max} = \frac{x_{\max}}{X_0} = \ln\left(\frac{E_0}{E_c}\right) + C_{e,\gamma} \tag{2.16}$$

where $C_\gamma = +0.5$ for photon-induced showers or $C_e = -0.5$ for electron-induced (or positron-induced) showers. Finally the depth containing 95 % of the shower energy can be approximated by:

$$L = (t_{\max} + 0.08Z + 9.6)\, X_0 \tag{2.17}$$

2.2.3 Magnet systems

Due to LORENTZ's force, the trajectory of a particle with momentum p and charge ze in a constant magnetic field $\vec{B}$ is a helix, with radius of curvature R and pitch angle λ. The momentum component perpendicular to $\vec{B}$ is expressed [3]:

$$p \cos\lambda = 0.3zBR \tag{2.18}$$

Knowing the magnetic field and estimating the trajectory of the particle through multiple energy deposits (known as a "track"), one can deduce the $\frac{z}{p}$ ratio.

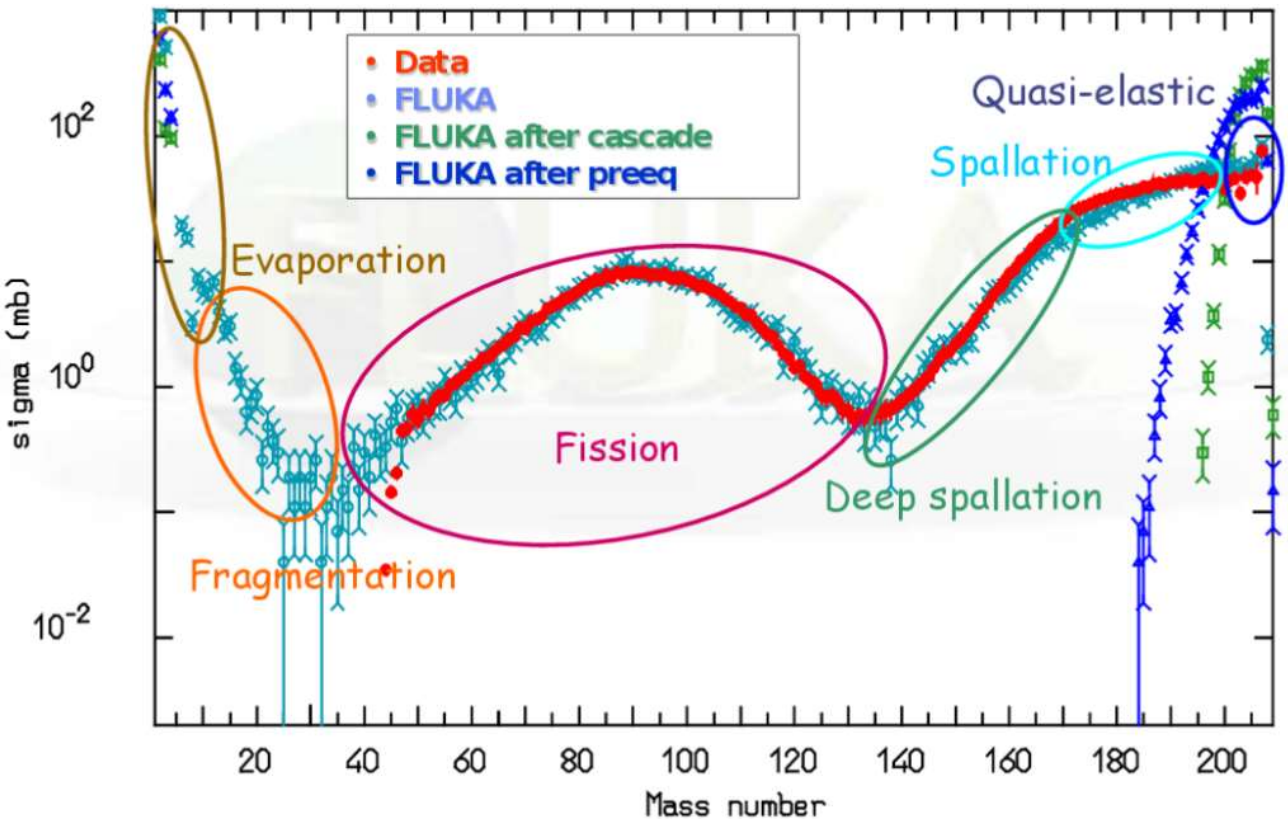


Figure 2.17: Interaction cross-section for the Pb + p reaction at the energy 1 A GeV, with A the mass number. The different regimes of hadronic interactions are highlighted. Reproduced from [71].

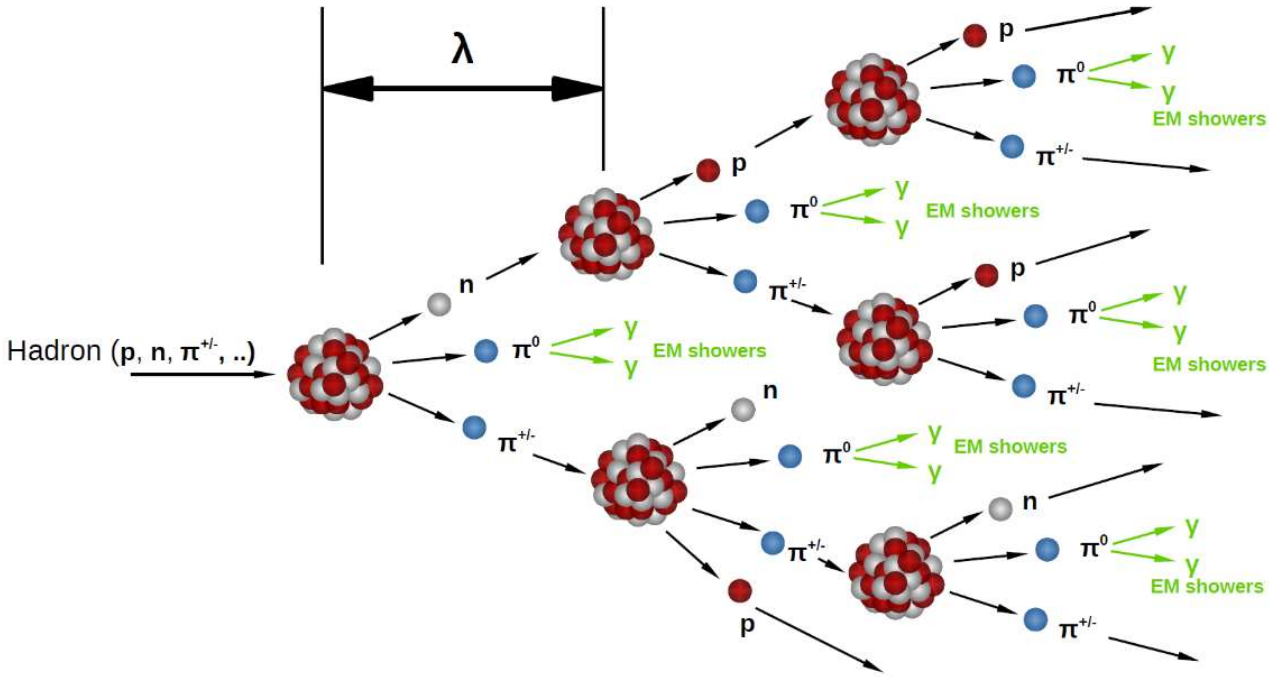
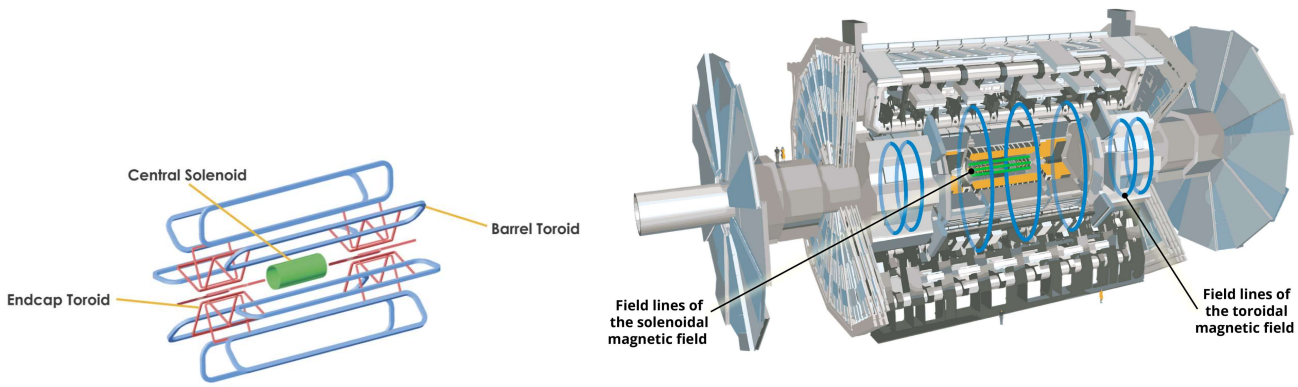


Figure 2.18: Representation of a hadronic shower. Reproduced from [71].



(a) (b)

Figure 2.19: **(a)** Magnets of the ATLAS detector: solenoid, barrel and endcap toroid magnets. **(b)** Magnetic field lines that they produce. Reproduced from [57].

The ATLAS detector has four large superconducting magnets [65, 66], shown in Fig. 2.19. The Inner Detector is immersed in a magnetic field created by the Central Solenoid (CS). The Muon Spectrometer is immersed in a toroidal magnetic field created by three air-core magnets: the Barrel Toroid (BT) and the EndCap Toroids (ECT), giving ATLAS its name.

The magnetic field strength varies along the trajectory of the particle. The CS provides a central field of 2 T with a peak magnetic field of 2.6 T. The peak magnetic fields in the BT and ECT are 3.9 T and 4.1 T respectively. The bending power of a magnet is characterized by the field integral $\int B \, \mathrm{d}l$, where B is the azimuthal field component integrated over a straight-line trajectory between the inner and outer radius of the toroids. The BT provides 2 to 6 T $\cdot$ m and the ECT contributes with 4 to 8 T $\cdot$ m in the $[0.0, 1.3]$ and $[1.6, 2.7]$ pseudorapidity ranges respectively.

2.2.4 Inner detector and tracking system

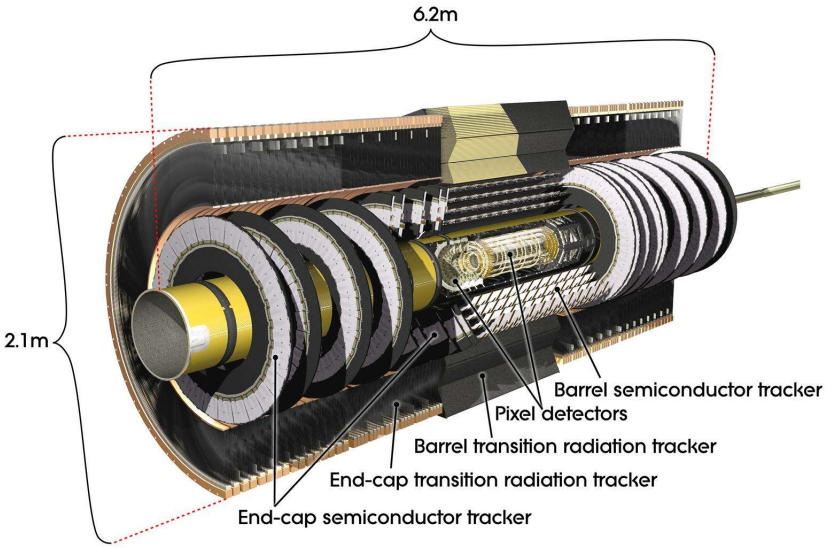
For bunch crossings with 25 proton-proton collisions, there are typically 1000 particles reaching the detector every 25 ns. The Inner Detector (ID) is designed to precisely measure the point of origin and trajectory of charged particles in the 2 T magnetic field of the solenoid magnet [65]. This is achieved by combining very precise measurements from silicon detectors and continuous tracking from gaseous detectors, resulting in excellent pattern recognition and precise momentum measurements. The ID has been designed to have a momentum resolution of:

$$\frac{\sigma_{p_T}}{p_T} = 0.05\ \% \times p_T \oplus 1\ \% \tag{2.19}$$

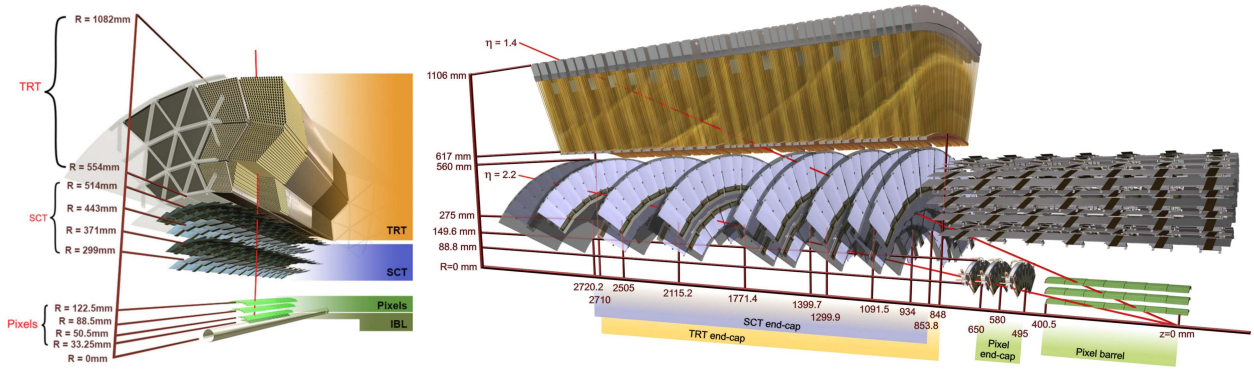
The ID covers the region $|\eta| < 2.5$. It is composed of three sub-detectors: the Pixel Detector (PD), the Semi-Conductor Tracker (SCT) and the Transition Radiation Tracker (TRT). It is divided into three regions: the barrel, concentric around the beam axis, measures particles in $|\eta| \lesssim 1$, and two endcaps, parallel to the beam axis, cover $1 \lesssim |\eta| < 2.5$. The layout of the ID is shown in Fig. 2.20. Considering its position in front of the calorimeters, the ID is designed to be as thin as possible to minimize particle interaction, given the required power, mechanical stability and cooling requirements; it is characterized in terms of radiation length X_0 in Fig. 2.21.

The Pixel Detector is the most granular detector with 92 million pixels, and the closest to the interaction point: up to 15 cm away. It uses hybrid silicon pixel sensors, that create electron-hole pairs along the particle path, that drift through the sensors and are read out to provide "hits". The pixels are divided into four barrel layers and three endcap layers. The first barrel layer, the Insertable B-Layer (IBL), was added during the long shutdown period of 2013-2014; its pixels are 50×250 µm^2. The IBL achieves intrinsic accuracy of 10 µm in the (R, ϕ) plane and 66.5 µm along the longitudinal z axis. The next three barrel layers and the three endcaps have 50×400 µm^2 pixels with 250 µm thickness. The intrinsic accuracies of those three barrel layers are 10 µm in the (R, ϕ) plane and 115 µm along the longitudinal z axis. For the endcaps, the intrinsic accuracies are 10 µm in the (R, ϕ) plane and 115 µm in the radial R direction.

The Semi-Conductor Tracker (SCT) has approximately 6.3 million read-out channels. It uses silicon micro-strip with single-sided p-in-n technology. The SCT is composed of four barrel layers and eight planar endcap discs. The intrinsic accuracy of the barrel channels is 17 µm in the (R, ϕ) plane and 580 µm in the z direction. The intrinsic accuracy for the endcaps is 17 µm in the (R, ϕ) plane and 580 µm in the R direction.



(a)



(b) (c)

Figure 2.20: Inner Detector. **(a)** Three dimensional view. **(b)** Transverse view. **(c)** Longitudinal views. The Insertable B-Layer (IBL) is only shown in **(b)**. Reproduced from [65].

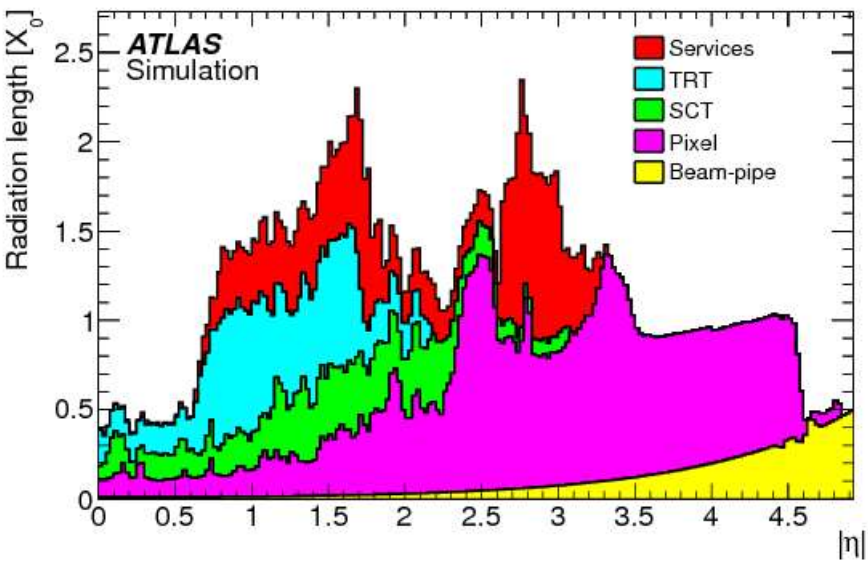


Figure 2.21: Radiation length X_0 as a function of the absolute pseudorapidity $|\eta|$ for the different components of the Inner Detector. Reproduced from [65].

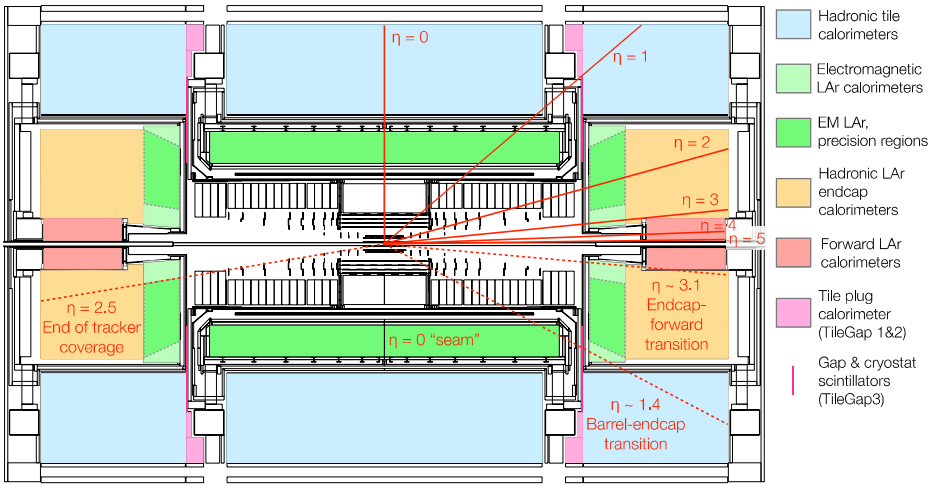


Figure 2.22: Layout of the ATLAS calorimeters with pseudorapitidy η values marked for reference. The inner detector systems can be seen in black-and-white in the center of the diagram; tracking is provided up to $\eta = 2.5$. The electromagnetic (EM) barrel and endcap calorimeters are shown in green. The EM barrel has consistent performance throughout, but has a seam in the construction at $\eta = 0$ which can impact jet energy resolution. The EM endcap has a precision region marked in darker green and an extended region in light green, and the transition from one to the other around $\eta \approx 2.5$ involves a dramatic change in the material layers. The hadronic tile calorimeter is shown in light blue while the hadronic endcap calorimeters based on liquid argon are illustrated in orange. The forward calorimeters are shown in red. Pink filled regions represent the tile plug calorimeter, often referred to as TileGap1 and TileGap2. The thin hot pink line marks the location of the very narrow gap and cryostat scintillators (TileGap3). The regions corresponding to the transition from barrel to endcap ($\eta \approx 1.4$) and from endcap to forward calorimeter ($\eta \approx 3.1$) are given for reference. Reproduced from [72].

The Transition Radiation Tracker (TRT) is a straw-tube tracker. Each straw is a drift tube, 4 mm in diameter, filled with a gas mixture of Xe (70 %), CO_2 (27 %) and O_2 (3 %), and with a thin tungsten wire at its center. A charged particle ionizes the gas, and the free electrons are collected by the central wire. The central barrel contains 52'544 straw tubes, 1.44 m in length, that are parallel to the beam axis. Each endcap contains 122'880 straws tubes, 0.37 m in length, that are arranged radially from the beam axis. The intrinsic accuracy of the barrel straws is 130 µm in the (R, ϕ) plane, and the straws do not provide information in the z direction. The maximum electron collection time is approximately 48 ns.

2.2.5 Electromagnetic and hadronic calorimeters

A calorimeter aims at measuring the energy carried by a particle. The calorimeter system of ATLAS is divided into multiple sub-systems [65–68], with different technologies to face the unprecedented challenges raised by the physics at LHC and the operation conditions: physics of interest was known to have low cross-sections and to be complex events, obscured by the high number of collisions per bunch crossing coupled to the high frequency of bunch crossings. The calorimeters of ATLAS provide high-precision measurements and high background rejection (in particular it is possible to isolate pile-up background from electromagnetic objects such as electrons and photons), with relatively fast response, excellent time stability and high radiation resistance.

The structure of the calorimeter system is shown in Fig. 2.22. The ElectroMagnetic (EM) calorimeter covers the region $|\eta| < 3.2$, a hadronic barrel calorimeter covers $|\eta| < 1.7$ (Tile barrel and extended barrel), Hadronic EndCap (HEC) calorimeters cover $1.5 < |\eta| < 3.2$, and Forward Calorimeters (FCal) cover $3.1 < |\eta| < 4.9$. The large angular coverage $|\eta| < 4.9$ is necessary for good missing transverse momentum (p_T^{miss}) measurement. The fine granularity of the EM calorimeter is suited for high background rejection and precision measurements of electrons and photons.

The inactive material in front of the EM calorimeter worsens the energy resolution for electrons and photons, mostly due to the solenoid, and therefore a liquid argon presampler in the barrel and endcaps is placed inside the solenoid, surrounding the ID, covering $|\eta| < 1.8$; this is shown in Fig. 2.23a. The calorimeters are thick enough for the showers (electromagnetic or hadronic) to be almost fully contained, and therefore not to leak in the muon spectrometer: the total thickness of the calorimeters is $> 22X_0$ for the EM calorimeter and $> 9.7\lambda$ for the total

calorimeter. This is shown in Figs. 2.23c to 2.23e. The complex transition region between the barrel and the endcap, around $|\eta| \approx 1.4$, is detailed in Fig. 2.23b; this region is of particular importance for the calibration in η of the jet energy scale, which is the subject of Chapter 4.

An important difference between electromagnetic and hadronic calorimeters is that the response to hadronic showers is usually lower than the response to electromagnetic showers. Indeed, part of the energy of hadronic showers is used to release protons and neutrons from calorimeter nuclei, and in the kinetic energy carried by recoil nuclei, and therefore does not lead to a calorimeter signal. Some detectors use specific combinations of detector and read-out materials to obtain a similar response for the two showers [73], but this is not the case of the calorimeter system of ATLAS, that is said to be "non-compensating".

For calorimeters, the energy resolution $\sigma(E)$ is parameterized as:

$$\frac{\sigma(E)}{E} = \frac{N}{E} \oplus \frac{S}{\sqrt{E}} \oplus C \tag{2.20}$$

with N the noise term, S the stochastic term and C the constant term. The three parameters are η-dependent. This will be explained in more detail in Section 3.3.3.

Electromagnetic calorimeter (ECal)

The Electromagnetic Calorimeter (ECAL or EM calorimeter) is a lead-liquid argon (LAr) sampling calorimeter with accordion geometry as shown in Fig. 2.24a. Liquid argon has a very linear behavior, is stable over time and is radiation-hard. The absorbers are made of lead plates. The accordion geometry provides a full coverage around ϕ, and enables a fast extraction of the signal. The linear behavior is improved by changes in thickness of the lead plates with η, and by changes with the radius of the folding angle of the accordions.

The calorimeter is divided in two parts: the central ($|\eta| < 1.475$) barrel and the two endcaps (EMEC, $1.375 < |\eta| < 3.2$). The transition region ($1.37 < |\eta| < 1.52$) between the barrel and the endcaps calorimeter has a large amount of material upstream of the first calorimeter layer as illustrated in Fig. 2.23b. This section is instrumented with scintillators located between the barrel and endcap cryostats, and extending up to $|\eta| = 1.63$. The barrel is segmented in three layers in depth (two layers for the EMEC), as shown in Fig. 2.24a. An accurate position measurement is obtained with the fine segmentation of the first layer in the η direction, optimized to differentiate pions ($\pi^0 \to \gamma\gamma$) from photons: $\Delta\eta \times \Delta\phi = 0.0031 \times 0.0245$. Its depth is $4.3X_0$. The second layer collects most of the electromagnetic shower, with a depth of $16X_0$, and its angular resolution is $\Delta\eta \times \Delta\phi = 0.025 \times 0.0245$. The third layer is thinner, with $2X_0$; it measures the tails of the showers, useful to estimate the amount of energy leaking beyond the ECAL. Its angular resolution is $\Delta\eta \times \Delta\phi = 0.05 \times 0.0245$.

In the barrel, the size of the drift gap on each side of the electrode is 2.1 mm, which corresponds to a total drift time of about 450 ns for its nominal operating voltage (2 kV). This is large compared to the duration between two proton bunch crossings, which is of 25 ns: the signal of the calorimeter integrates over multiple bunch crossings, leading to a dependency of the signal to multiple bunch crossings, which is called "out-of-time pile-up". An example of signal for the barrel is shown in Fig. 2.24b. Due to a non-constant gap on each side of the electrodes in the EMEC, the voltage is varied with η to obtain a uniform η-independent detector response.

The design resolution of the ECAL is:

$$\frac{\sigma(E)}{E} = \frac{10\ \%}{\sqrt{E}} \oplus 0.7\ \% \tag{2.21}$$

Hadronic calorimeters (TileCal and HEC)

The Hadronic Calorimeter (HCAL) measures the energy of hadrons that did not deposit all of their energy in the previous detector layers, either because they are neutral, or because they are very energetic. The HCAL is composed of two regions: the barrel (called the Tile calorimeter, or TileCal) covering $|\eta| < 1.7$, and the Hadronic Endcap Calorimeters (HEC) covering $1.5 < |\eta| < 3.2$.

The TileCal uses steel (absorber) and plastic scintillator (active medium) tiles, with a volume ratio 4.7 : 1, as shown in Fig. 2.25. The scintillator produces ultraviolet photons that are collected by fibres that also shift their wavelength to visible light, and that are converted to an electric current by linear photomultipliers. The tiles are grouped in modules by the fibres, so as to form three radial sampling layers shown in Fig. 2.25a, of granularity $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ for the first two layers and $\Delta\eta \times \Delta\phi = 0.2 \times 0.1$ for the third one. The TileCal is divided into a barrel that covers $|\eta| < 1$ and two extended barrels that cover $0.8 < |\eta| < 1.7$. In total, the TileCal has a depth of more than 9.7λ.

The HEC is a copper-LAr calorimeter, which layout is shown in Fig. 2.26a. It consists of two independent wheels per endcap, each of them being built from 32 identical wedge-shaped modules and being divided into two segments in depth, for a total of four samplings per endcap. The wheels closest to the interaction point are built from 25 mm parallel copper plates, while those further away use 50 mm plates. The plates are interleaved with 8.5 mm liquid

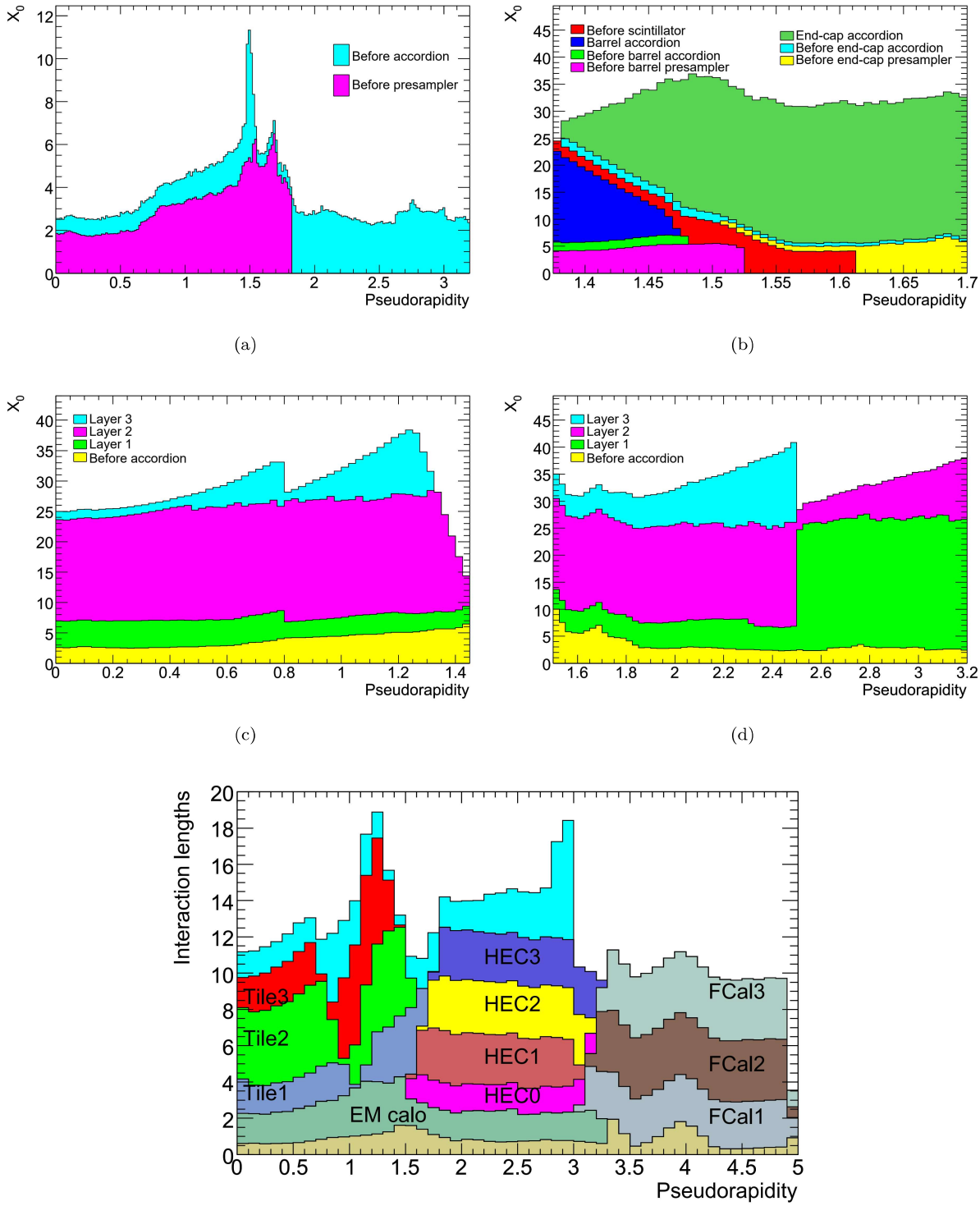
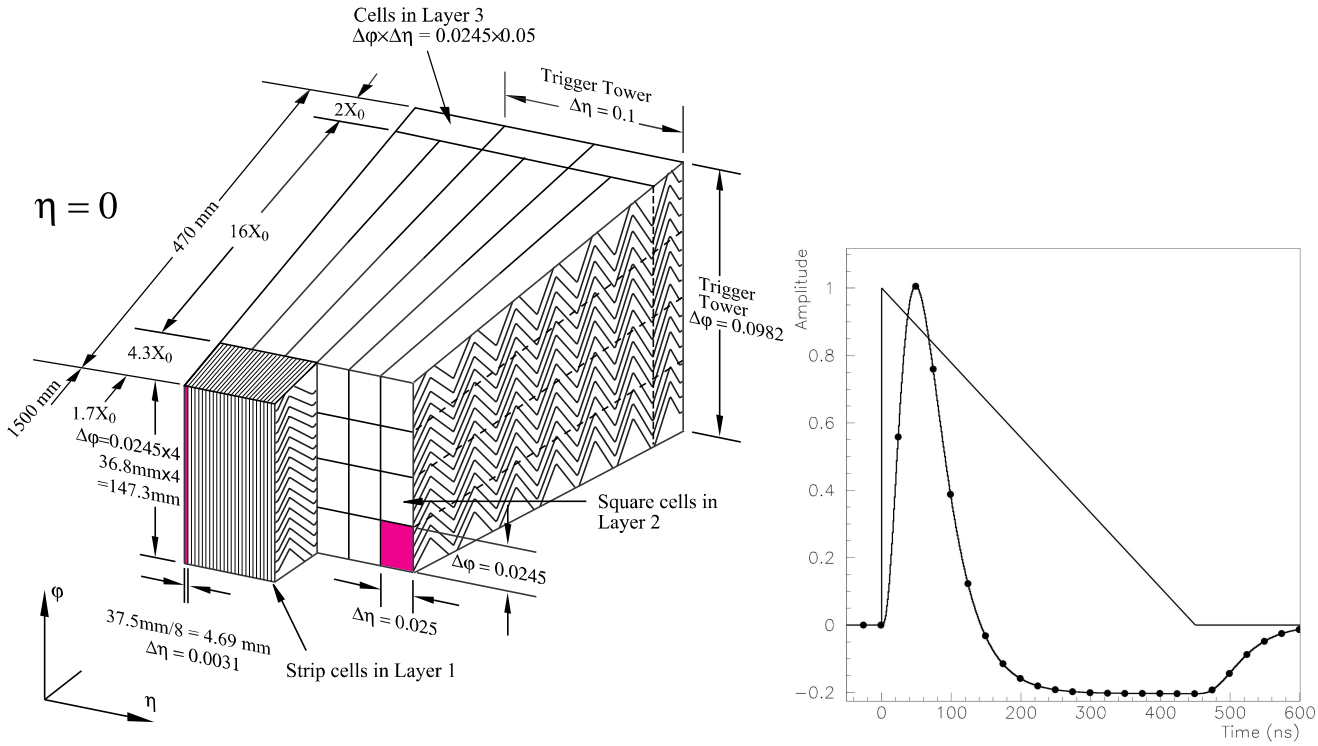
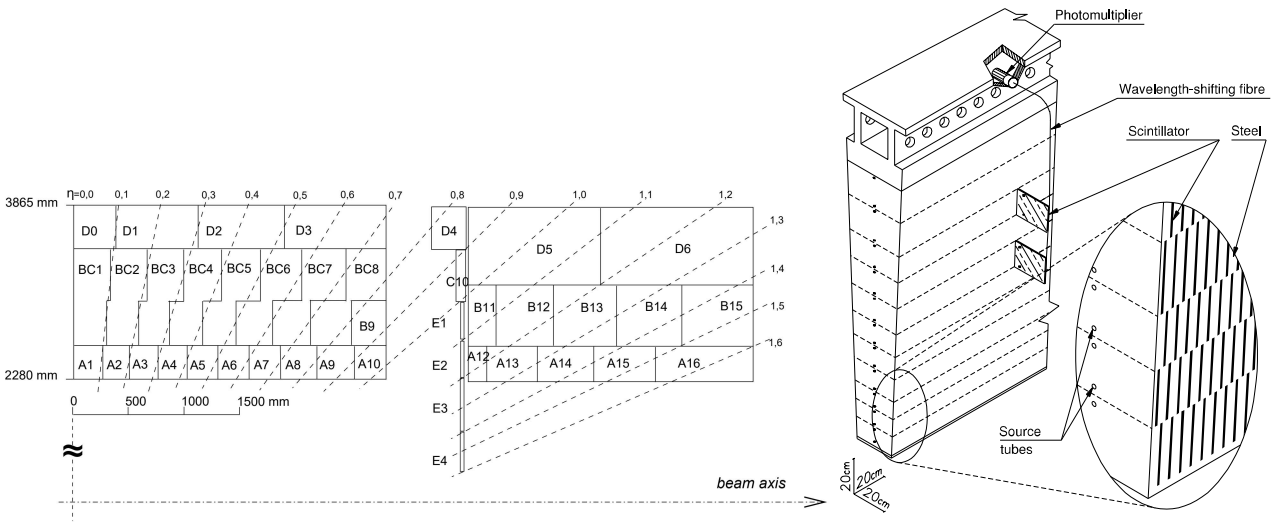


Figure 2.23: **(a-d)** Cumulative amount of material in the EM calorimeters (and in front of them) as a function of $|\eta|$ in terms of radiation length X_0. **(a)** Total amount of material in front of the presampler layer and in front of the accordion. **(b)** Details of the crack region between the barrel and endcap cryostats, both in terms of material in front of the active layers (including the crack scintillator) and of the total thickness of the active calorimeter. **(c,d)** Thicknesses of each accordion layer as well as the amount of material in front of the accordion, **(c)** for the barrel and **(d)** for the endcap. **(e)** Cumulative amount of material in the EM and hadronic calorimeters as a function of $|\eta|$ in terms of interaction length λ. The total amount of material in front of the EM calorimeter is also shown in light brown and the total amount of material in front of the muon spectrometer is shown in light blue (for $|\eta| < 3.0$). Reproduced from [65].



(a) (b)

Figure 2.24: ECal. **(a)** Diagram of a barrel module where the different layers are visible with the ganging of electrodes in ϕ. The granularity in η and ϕ of the cells of each of the three layers and of the trigger towers is also shown. **(b)** Amplitude versus time for triangular pulse of the current in a LAr barrel electromagnetic cell and of the FEB output signal after bi-polar shaping. Also indicated are the sampling points every 25 ns. Reproduced from [65].



(a) (b)

Figure 2.25: TileCal. **(a)** Segmentation in depth R and pseudorapidity η of the tile-calorimeter modules in the central (left) and extended (right) barrels. The bottom of the picture corresponds to the inner radius of the tile calorimeter. The tile calorimeter is symmetric about the interaction point at the origin. **(b)** Diagram showing how the mechanical assembly and the optical readout of the tile calorimeter are integrated together. The various components of the optical readout, namely the tiles, the fibres and the photomultipliers, are shown. Reproduced from [65].

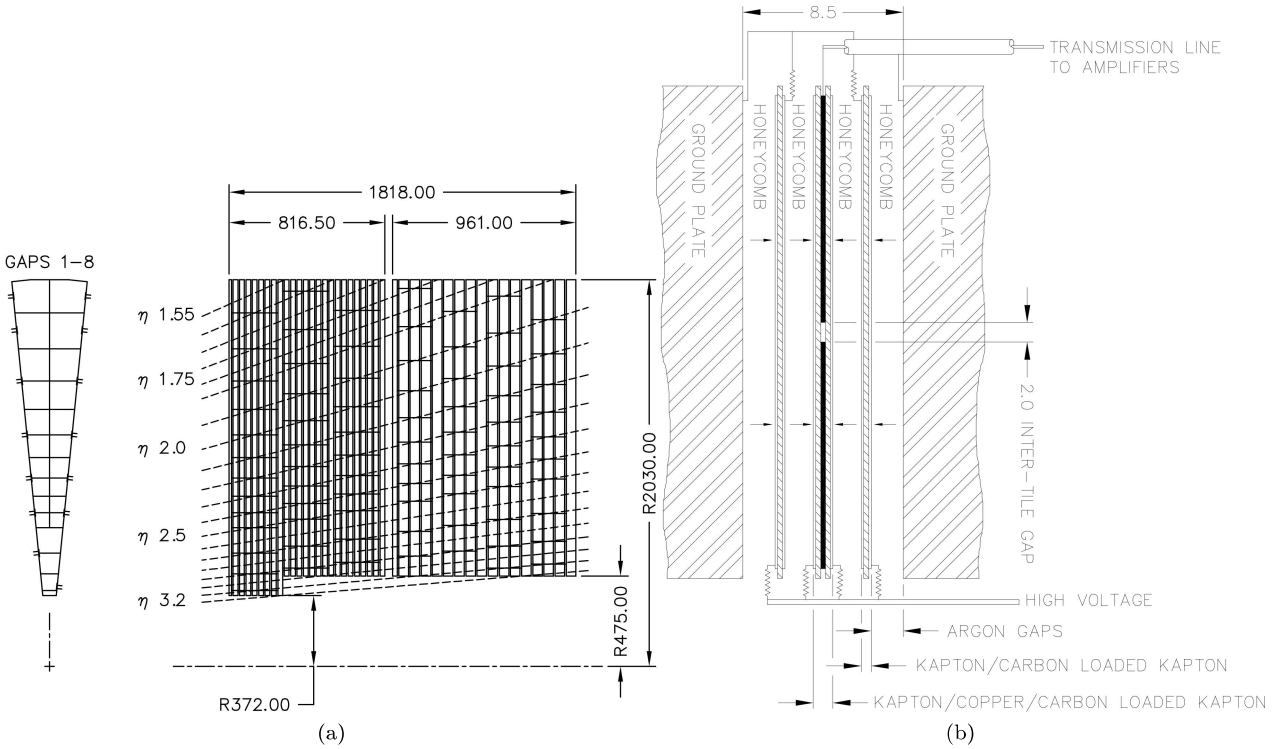


Figure 2.26: HEC. **(a)** Schematic (R, ϕ) (left) and (R, z) (right) views of the HEC. The semi-pointing layout of the readout electrodes is indicated by the dashed lines. Dimensions are in mm. **(b)** Diagram of the arrangement of the HEC readout structure in the 8.5 mm interplate gap. All dimensions are in mm. Reproduced from [65].

argon (LAr) gaps, in which three read-out planes define four gaps connected as an electrostatic transformer, as shown in Fig. 2.26b. The angular granularity is $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ for $1.5 < |\eta| < 2.5$ and $\Delta\eta \times \Delta\phi = 0.2 \times 0.2$ for $2.5 < |\eta| < 3.2$. The typical drift time for electrons is 430 ns.

The design resolution of the TileCal and HEC is:

$$\frac{\sigma(E)}{E} = \frac{50\ \%}{\sqrt{E}} \oplus 3\ \% \tag{2.22}$$

Forward calorimetry (FCal)

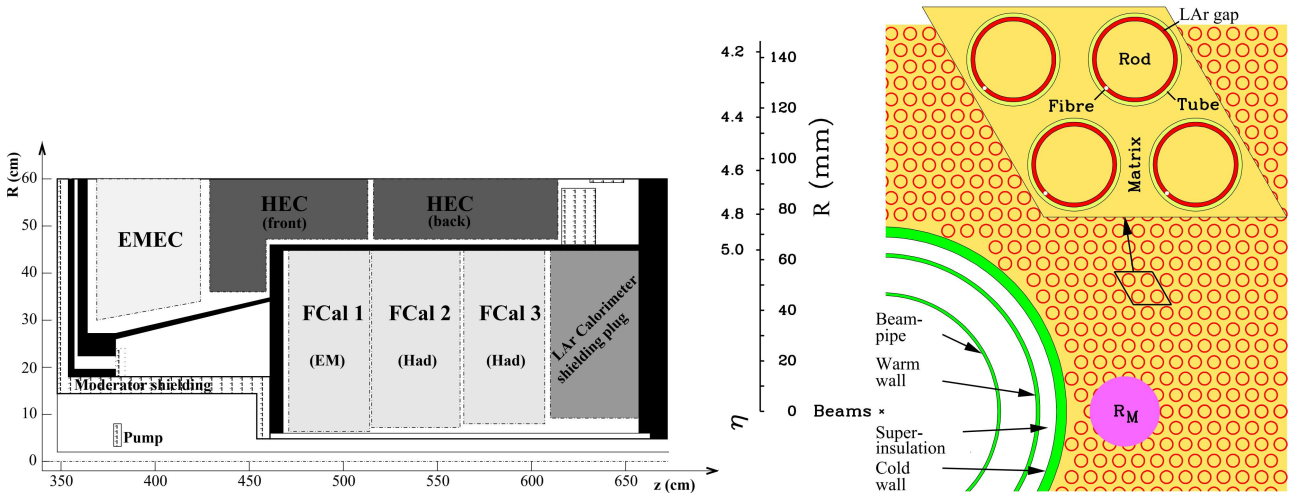
The Forward Calorimeter (FCal) covers the $3.1 < |\eta| < 4.9$ that the ID and the other calorimeters do not cover (except for a small overlap at $|\eta| \approx 3.1$). It consists of three 45 cm long modules, as shown in Fig. 2.27a: the first (FCal 1), made of copper, is suited for EM measurements, while the other two (FCal 2 and FCal 3), made of tungsten, measure predominantly the hadronic component of jets. Copper optimizes the resolution and the heat removal, while tungsten minimizes the lateral shower spread of hadronic showers. Each module consists of a metal matrix, with regularly spaced longitudinal channels filled with the electrode structure consisting of concentric rods and tubes parallel to the beam axis, shown in Fig. 2.27b. The LAr in the gap between the rod and the tube is the sensitive medium. In FCal 1, the full drift time is 60 ns, and in FCal 2 and FCal 3 it scales with the gap size.

The design resolution of the FCAL is:

$$\frac{\sigma(E)}{E} = \frac{100\ \%}{\sqrt{E}} \oplus 10\ \% \tag{2.23}$$

2.2.6 Muon spectrometer

Muons have a heavier mass than electrons, and therefore their brehmsstrahlung radiations are reduced: they pass through the ID and the calorimeters with minimal energy loss. The Muon Spectrometer (MS) is designed such that muons produce "hits", to reconstruct tracks and therefore to measure their momentum, in the range $|\eta| < 2.7$ [65, 66]. For muons with an energy greater than 3 GeV, the MS can measure their properties on its own (direction, electric charge, momentum); for energies below 3 GeV though, the energy loss in the calorimeter is significant and the measurement of the properties of the muon requires the information from the tracks in the inner detector.



(a) (b)

Figure 2.27: FCal. **(a)** Diagram showing the three FCal modules located in the end-cap cryostat. The material in front of the FCal and the shielding plug behind it are also shown. The black regions are structural parts of the cryostat. The diagram has a larger vertical scale for clarity. **(b)** Electrode structure of FCal1 with the matrix of copper plates and the copper tubes and rods with the LAr gap for the electrodes. The MOLIÈRE radius, R_M, is represented by the solid disk. Reproduced from [65].

The MS is also used to detect punch-through signals, *i.e.* hadronic showers that are not fully contained inside the hadronic calorimeter; this allows to correct the energy measured in the calorimeters.

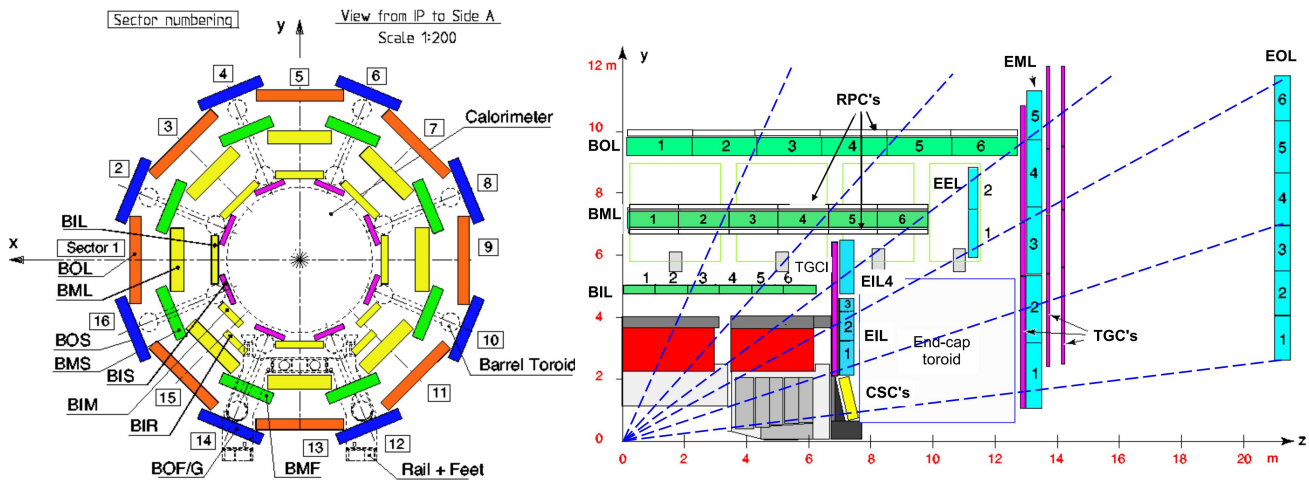
To detect muons, one has to be able to trigger on them and identify the specific bunch crossing in which the particle was produced, every 25 ns. This requires very good timing resolution of a few ns. This criterion is met by two technologies used in ATLAS: Resistive Plate Chambers (RPC), used in the barrel region ($|\eta| < 1.2$), and Thin Gap Chambers (TGC), used in the endcap region ($1 < |\eta| < 2.7$).

Physics analyses, for instance for the measurement of the mass of the HIGGS boson, require very accurate measurement of the momentum of the muon: $2-3$ % below 100 GeV; at 1 TeV, the goal was 10 % for the search for new massive objects. Since the momentum resolution is $\sigma(p_T) \propto \left(BL^2\right)^{-1}$, with B the magnetic field and L the length of the measurement device, the large toroid complex provides the right lever arm to maximize L. The position of the hits is known accurately thanks to Monitor Drift Tubes (MDTs), complemented by multi-wire proportional chambers in the forward region, in this case high-precision Cathode Strips Chambers (CSCs). To achieve the desired precision, the position of the MDTs and CSCs must be known better than 0.05 mm inside a volume of $45 \times 25 \times 25$ m^3. The MDTs allow the monitoring of the deformations of the structure to this precision. This is however not enough: the relative positions of the detector layers must be precisely known; this is achieved with a complex projective optical alignment system. Finally, magnetic sensors are mounted onto the detectors to provide a few thousand measurements of the complex magnetic field of the toroids and their local interference. This last system provides the position of the very large coils (25×5 m^2) with a precision of 1 mm.

The different detectors that compose the MS are shown in Fig. 2.28. There are three layers of MDTs in the barrel and the endcap, providing the three hits that allow the estimation of the momentum.

The different technologies, shown in Fig. 2.29, are:

- The Monitored Drift Tubes (MDTs) measure the curvature of tracks (precision measurement of the position of hits) in the $|\eta| < 2.7$ region. Each chamber is composed of three to eight layers of drift tubes. A drift tube is 30 mm in diameter, with a wall made of aluminum, and contains a Ar/CO_2 gas mixture (in 93 : 7 proportion) at 3 bar. The electrons resulting from the ionization are collected at the central gold-plated tungsten wire. The maximum drift tube time is about 700 ns. The average resolution achieved is of 80 µm per tube, or about 35 µm per chamber.
- The Cathode Strip Chambers (CSC) measure the curvature of tracks (precision measurement of the position of hits) in the $2 < |\eta| < 2.7$ region. CSCs are multi-wire proportional chambers with cathode planes segmented into strips in orthogonal directions. The resolution of a chamber is 40 µm in the bending (R, η) plane and about 5 mm in the transverse (R, ϕ) plane.
- The Resistive Plate Chambers (RPC) are used for triggering thanks to a very good time resolution, and provide a second coordinate measurement in the central region ($|\eta| < 1.05$). An RPC is a gaseous parallel



(a) (b)

Figure 2.28: **(a)** Section of the barrel muon system perpendicular to the beam axis (non-bending (R, ϕ) plane), showing the three concentric cylindrical layers composed of eight large and eight small chambers symmetric in ϕ. **(b)** Section of the muon system in a plane containing the beam axis (bending (R, η) plane) showing the layers in the barrel and endcap regions. MDTs are shown in green in the barrel and cyan in the endcap, while CSCs are shown in yellow. Reproduced from [65].

electrode-plate detector (no wire). Two resistive plates are kept parallel to each other at a distance of 2 mm and an electric field of about 4.9 kV/mm between them is applied. When a muon enters the chamber it ionizes the gas, creating an avalanche of charges along its track that are collected at the anode. The signal width is of about 5 ns.

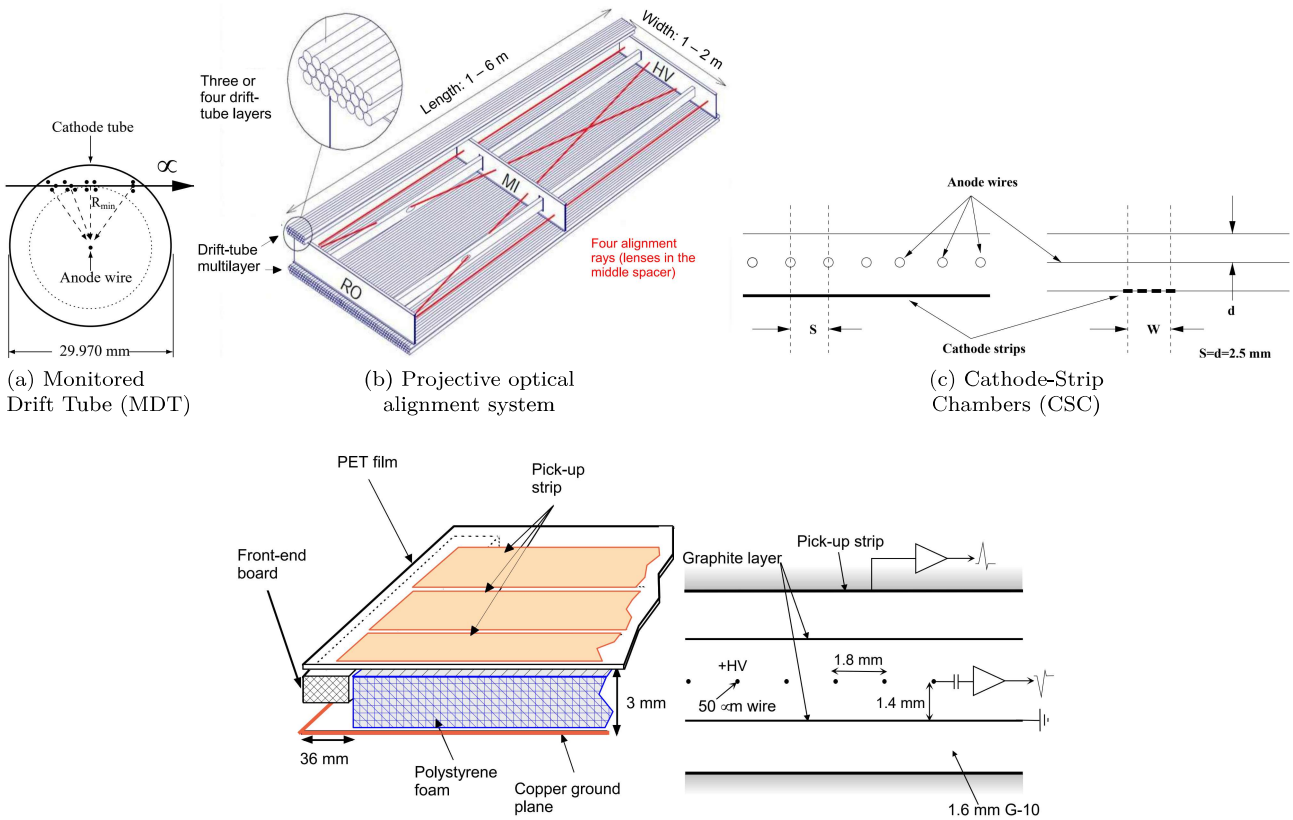
- The Thin Gap Chambers (TGC), like RPCs, are used for triggering thanks to a very good time resolution, and provide a second coordinate measurement in the forward region ($1.05 < |\eta| < 2.4$). They are multi-wire proportional chambers, using a highly quenched mixture of CO_2 and n-pentane, with a wire-to-cathode distance of 1.4 mm that is smaller than the wire-to-wire distance of 1.8 mm. The signal arrives in a time window of 25 ns, more than 99 % of the time.

2.2.7 Forward detectors

The main part of the ATLAS detector is designed to reconstruct the products from head-on collisions, in which the particles are created with large angles relative to the incoming protons. There are also collisions that lead to products with small angles relative to the beam line, *i.e.* in the forward direction. Those interactions are interesting by themselves, and also provide an estimation of the total cross-section of proton-proton collision and of the luminosity delivered by the LHC. The latter information allows to make absolute production rate measurements, *i.e.* cross-section measurements, like the one presented in Chapter 5. The integrated luminosity is shown in Fig. 2.2, and the luminosity-weighted number of interactions per bunch crossing is shown Fig. 2.35a.

There are four detectors in the forward regions of ATLAS [65, 66], shown relatively to ATLAS in Fig. 2.30:

- LUCID stands for LUminosity measurement using Cherenkov Integrating Detector. It is located 17 m away from the interaction point of ATLAS, and as its name suggests it uses a Cherenkov detector. It is used for luminosity measurement and monitoring. It recently released the highest-precision luminosity measurement ever achieved for a particle collider, evaluated as 140.1 fb^{-1} with a relative uncertainty of 0.83 % for the Run 2 data-taking period [62]. The number of interactions in a bunch-crossing is proportional to the number of particles detected in this detector. The Cherenkov light is emitted directly in the quartz window of photo-multiplicators, that are located 12 cm from the beam line.
- ALFA stands for Absolute Luminosity For ATLAS. It is located about 240 m from the ATLAS interaction point. ALFA is a scintillating fibre tracker, housed in so-called Roman Pots, with detectors that can move in the vertical plane and can approach the beam line to distances of the order of a few millimetres. ALFA measures protons that interact after the collision, and that originate mainly from elastic or diffractive scattering. This provides a measurement of the total interaction cross-section and of the luminosity.
- AFP, like ALFA, is designed using Roman Pots to measure protons that emerge intact from the interaction region. It consists of a silicon tracker located 220 m away from the interaction point of ATLAS.



(a) Monitored Drift Tube (MDT)

(b) Projective optical alignment system

(c) Cathode-Strip Chambers (CSC)

(d) Resistive Plate Chambers (RPC)

(e) Thin Gap Chambers (TGC)

Figure 2.29: Diagrams of the different technologies used in the muon spectrometer. Reproduced from [65].

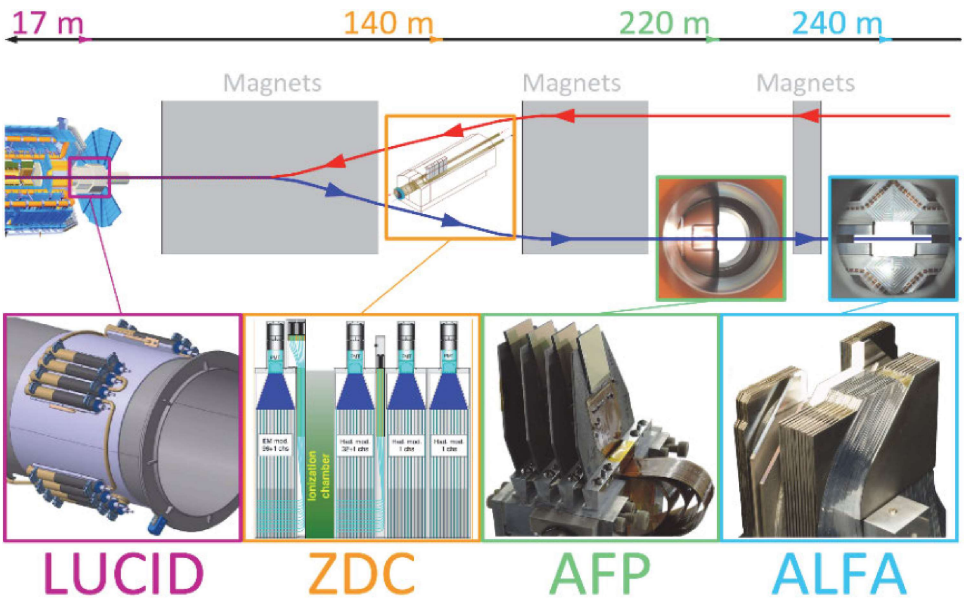


Figure 2.30: Diagrams of the different technologies used in the muon spectrometer. Reproduced from [66].

- The Zero-Degree Calorimeter (ZDC) is mainly used for heavy ion collisions. It is a tungsten quartz sampling calorimeter installed between the two LHC beam pipes, located 140 m away from the interaction point of ATLAS, in a configuration such that only neutral particles produced at the interaction point can reach the detector. It is primarily used as a trigger for ultra-peripheral heavy ion collisions.

2.2.8 Triggers and data acquisition

In ATLAS, during Run 2, there were about 1.3×10^9 collisions per second (considering a mean number of interactions per bunch crossing of $\langle\mu\rangle = 33.7$). At each bunch crossing, 40×10^6 times per second, all the $\approx 10^8$ channels of the ATLAS detector are read, generating enormous quantities of data: of the order of 60 TB/s, as shown in Fig. 2.31a. Recording all those events is impossible. This is however not a problem. Indeed, most of the events display physics that is already quite well-known, and we are now mostly interested either in rare events, or by subtle effects that can be put forth thanks to precision measurement and to the huge statistics at our disposal. Even in the latter case, recording all the events would not be useful because the limiting factor in this case is the systematic uncertainties (like in the measurements of jet cross-sections).

A two-level triggering system is employed to select the events that will be stored for physics analyses, shown in Fig. 2.31b.

1. A hardware-based set of triggers, called the "Level 1 (L1) triggers", do a partial and fast reconstruction of the event to determine which events could potentially be interesting based on physics criteria, and keep all of them or in some cases only a fraction of them. The selection of the events kept, among the events that pass the physics requirement of the L1 trigger, is random with a given probability. The L1 performs the raw selection on a limited amount of information from the detector, and only using a partial event reconstruction for the selection to be extremely fast: a decision is made in approximately 2.5 µs.
2. Then a software-based set of triggers, called the "High Level Triggers" (HLT), are seeded by the L1 triggers (meaning that only events selected by the L1 triggers are considered) and reconstruct the event a lot more thoroughly. Again, the HLT triggers only keep a fraction of the events that are potentially interesting, randomly with a given probability. The HLT uses about 40'000 CPU cores and takes a decision in about 200 µs.

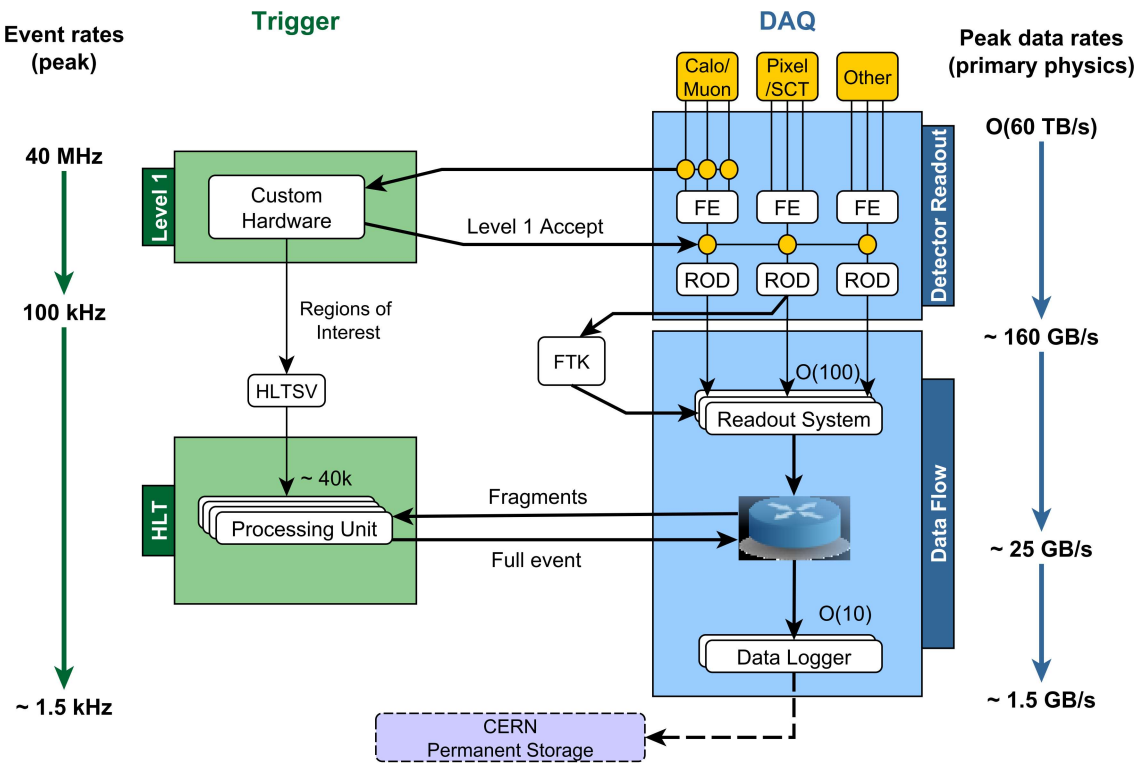
At the L1 stage, only one event in 400 is kept; at the HLT stage, only one in 100 is kept. The final rate is still big but is a lot more manageable: "only" 1500 bunch crossings are saved per second, that amounts to 1.5 GB/s that are stored permanently.

The L1 algorithm searches for high transverse momentum electrons, muons, hadronically-decaying taus, photons, jets and missing transverse momentum (the offline reconstruction of those objects is described in Section 2.2.9), to know whether the event is worth keeping or not. It faces a few difficulties. It has to identify the precise bunch-crossing of interest, which is not an easy feat considering the physical size of the muon spectrometer (implying times-of-flight exceeding the bunch-crossing interval) and considering the large integration time of the pulse shape of the calorimeters (typically extending over four bunch crossings). For the muon spectrometer, the L1 trigger only uses the information from the trigger chambers (RPCs and TGCs). For the calorimeters, the spatial granularity is reduced.

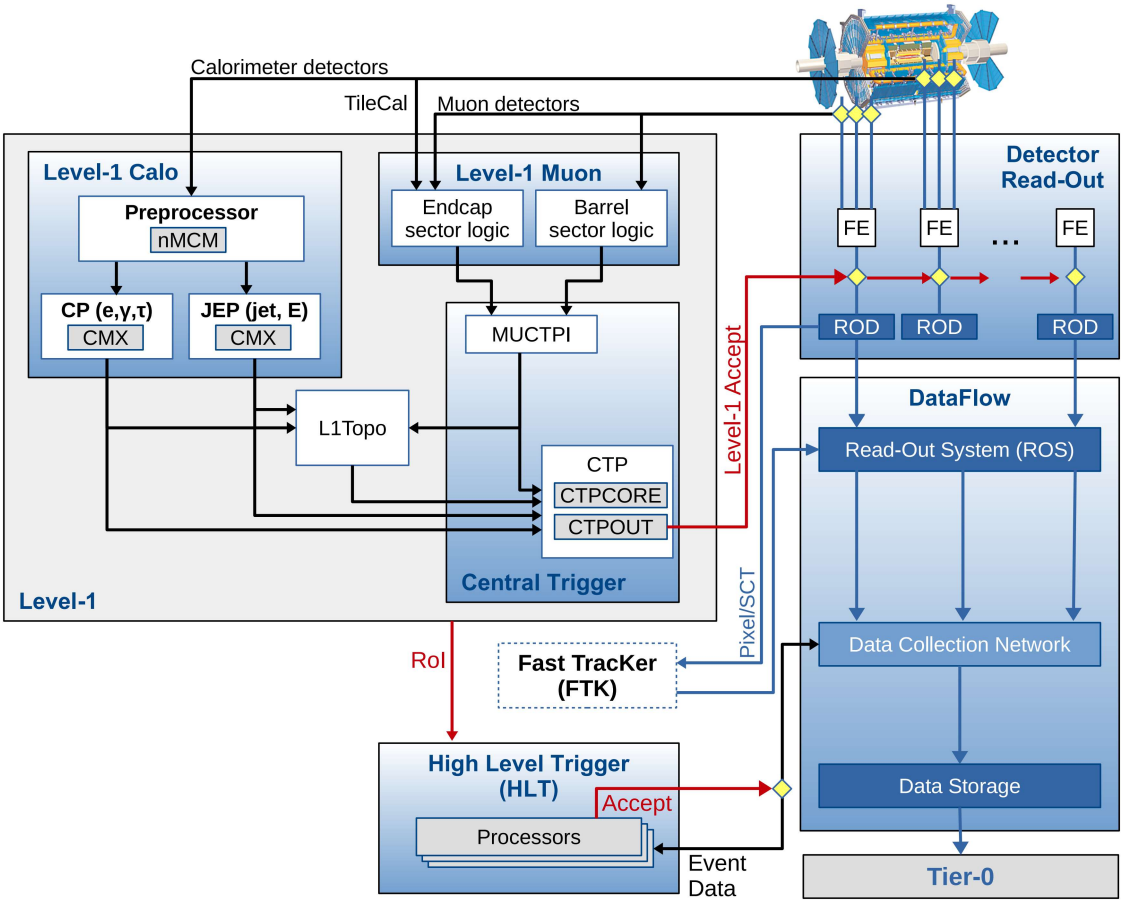
For instance, in the case of jets, the energy of the calorimeter cells is summed in 2×2 trigger towers, with a reduced spatial granularity of $\Delta\eta \times \Delta\phi = 0.2 \times 0.2$, as shown in Fig. 2.32. This is done as a sliding window in 2×2 3×3 and 4×4 areas, and the location of a 2×2 local maximum defines the coordinates of the jet Region of Interest (RoI). The RoIs are then compared to predefined jet energy thresholds, to determine whether the event is interesting or not.

The HLT relies on the full detector information available, applying tighter selection criteria, with reconstruction algorithms that are very close to the offline algorithms. Depending on the trigger selection, the HLT can check for the required objects only within regions selected by the RoI, or in all the detector regions. An HLT can therefore be seeded by an L1 trigger, but not always. For jets, the reconstruction used during Run 2 was the same topological clustering as the one used offline (see Chapter 3), with a calibration that includes most of the offline steps. Having HLT jets that are very close to the offline jets improves the selection efficiency in offline physics analyses. The single-jet triggers are described in more detail in Chapter 4, Section 4.2.1.2.

Considering the large number of physics objects and possible physics analyses, the selection criteria have to be broad enough not to loose interesting events. There are multiple triggers, *i.e.* multiple combinations of L1 selection criteria and HLT selection criteria, for all sorts of objects. This is called the "trigger menu". For instance, the trigger menu used in 2018 is given in [75].



(a)



(b)

Figure 2.31: Trigger and Data Acquisition (TDAQ) system used by ATLAS for Run 2. **(a)** Diagram showing the expected peak rates and bandwidths through each component. **(b)** Diagram that focuses on the components of the Level 1 (L1) trigger system. Reproduced from [74].

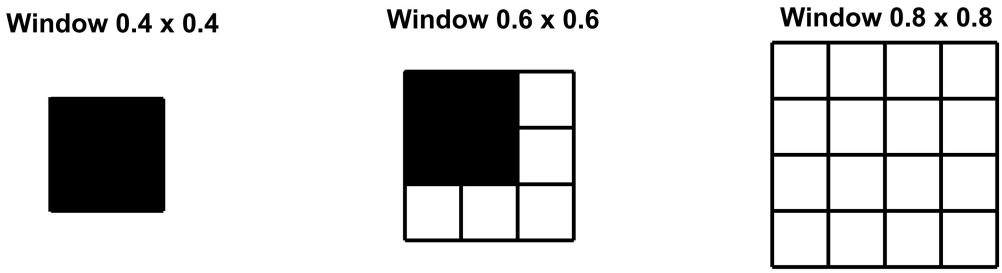


Figure 2.32: Jet trigger algorithms, based on $\Delta\eta \times \Delta\phi = 0.2 \times 0.2$ jet elements and showing Regions of Interest (RoIs) in the shaded areas. In the 0.6×0.6 case, there are four possible windows containing a given RoI. In the 0.8×0.8 case, the RoI is required to be in the center position, in order to avoid the possibility of two jets per window. Reproduced from [65].

2.2.9 Data processing and object reconstruction

The data processing flow of ATLAS is shown in Fig. 2.33.

Once the experimental data are recorded into permanent storage by the online trigger system, they are analyzed offline. In parallel of experimental data, there are simulated datasets [76], used for multiple purposes like calibration or comparison with theory in physics analyses. Simulated datasets (called "MC") are also created with Monte Carlo techniques explained in Chapter 1 Section 1.3.1: a matrix element for the desired physics process is generated, then if necessary showering and hadronization occur, leaving a collection of particles. At this stage, something is missing: pile-up, *i.e.* the effects of having multiple proton-proton collisions per bunch crossing (in-time pile-up), and the effect of the collisions from previous bunch crossings due to the long integration time of the calorimeters (out-of-time pile-up). For simulations to match reality, zero-bias pile-up particles, simulated with PYTHIA8 following the N_{PV} distribution shown in Fig. 2.35a, are overlaid to the hard-scatter particles of the MC simulation (here, PYTHIA8 [7] is used with the A3 set of tuned parameters and the NNPDF2.3LO [11] PDF set). Those MC particles (both from the hard-scatter event and from the additional pile-up) undergo a simulation of the detection process, with the GEANT4 program [77] which contains a very precise description of the ATLAS detector (geometry, materials, digitization).

Both experimental and MC datasets are therefore collections of hits in the tracking systems and energy deposits in the calorimeters. The reconstruction step then occurs, where the physics objects are built. The reconstruction of those objects is briefly described hereafter. Finally, physics analyses can start.

Considering the extremely large amount of data processed (see orders of magnitude for Run 1 in Fig. 2.33), and the large demand for computing power and storage from physics analyses, CERN has made available its own distributed computing infrastructure, the "grid", ran by laboratories from all over the world. ATLAS has developed its own software for data analysis, called Athena, based on a kernel called Gaudi (shared with LHCb).

Track reconstruction

Tracks are reconstructed from the hits in the ID [78], as shown in Fig. 2.34a. Tracks are parameterized by a helicoidal trajectory with five parameters, shown in Fig. 2.34b. The impact parameters d_0 and z_0 characterize the position of the perigee of the track with respect to the beam axis, with d_0 the distance in the transverse plane and z_0 the longitudinal distance. Two angles characterize the direction of the track at the perigee: ϕ is the azimuthal angle and θ is the polar angle. The charge-momentum ratio q/p defines the curvature of the track due to the 2 T solenoidal magnetic field.

There are typically one hit per PD and SCT layer crossed by a charged particle, *i.e.* 6 hits, and 36 hits in the barrel region of the TRT. ATLAS uses an "inside-out" method: the information from the inner sub-detectors (PD and SCT) is used before the outer TRT information is taken into account. The first step is to cluster hits locally to form track seeds. Second, the estimated trajectory, called a "search road", is built: it is a list of the detector modules that are expected to contain hits. Third, a KALMAN filter searches both inwards and outwards along the search road while trying to smooth the trajectory. Fourth, the track quality is checked to reject "fake" ones, and tracks that share hits are resolved. Fifth, an "outside-in" procedure is performed, starting with the remaining TRT hits, to reconstruct tracks from the late photon conversions ($\gamma \to e^+e^-$).

Vertex identification

Following track reconstruction, vertices are reconstructed [79]. A Primary Vertex (PV) is the point where one proton-proton collision occurs. When proton bunches cross each other, multiple collisions occur, and there are therefore multiple PVs every 25 ns, with a mean at $\langle\mu\rangle = 33.7$ during the Run 2 of LHC, as shown in Fig. 2.35a.

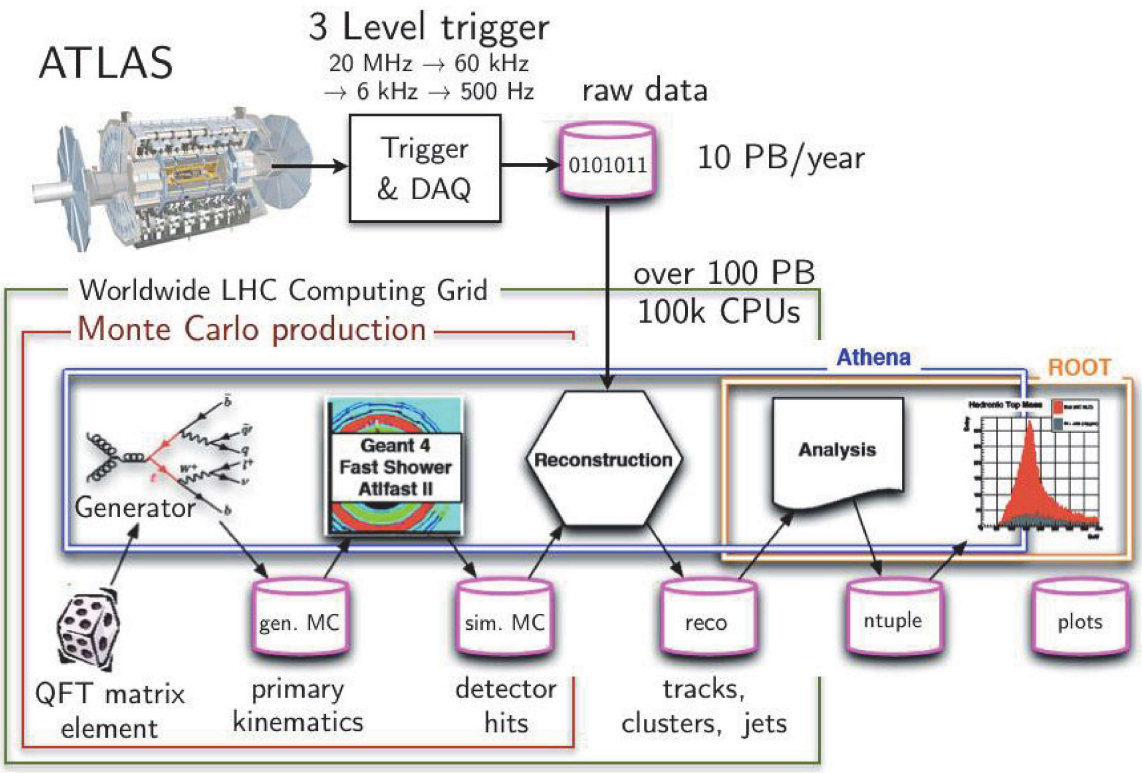
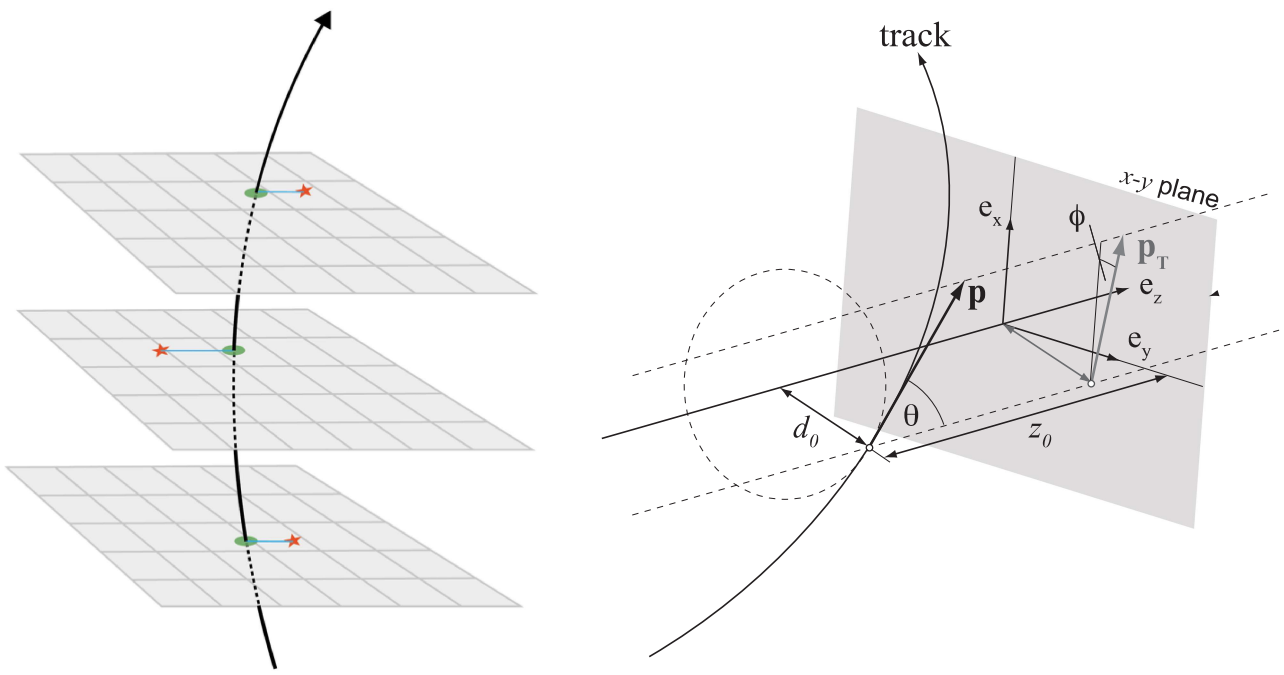
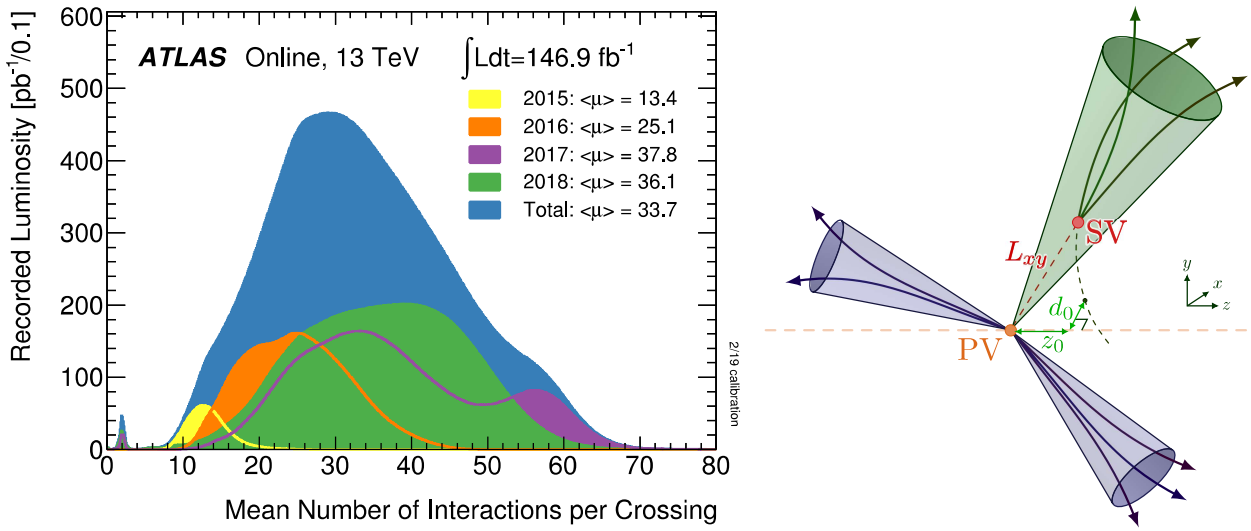


Figure 2.33: Data processing model of ATLAS during the Run 1 operation period, from the detector and simulation (Monte Carlo production) to the physics results. Reproduced from [66].



(a) (b)

Figure 2.34: **(a)** Track reconstructed from hits in the Inner Detector. Reproduced from [57]. **(b)** Track parametrization. Reproduced from [78].



(a) (b)

Figure 2.35: **(a)** Average number of primary vertices per bunch crossing for the different years of the Run 2 data-taking period. Reproduced from [80]. **(b)** Primary and secondary vertices, with particles associated in jets. Charged particles produce tracks, which allow for the reconstruction of the vertices. Reproduced from [70].

The identification of the hardest scattering PV_0 from the other (soft) ones is essential for analyses; the particles coming from the other PVs is called "in-time pile-up".

A Secondary Vertex (SV) is the point where a long-lived particle decays into other particles, as shown in Fig. 2.35b. The reconstruction of SVs is important for particle identification, like b-tagging, because B mesons have a relatively long lifetime ($\tau \approx 1.5$ ps) and therefore their flight distance is of the order of 0.5 mm to a few millimeters depending on the boost.

Vertices are found in two steps. First, a vertex finding step consists in a three-dimensional FOURIER transform of the track path lengths to identify the high density regions, that correspond to vertices. Second, a vertex fitting step iteratively assigns tracks to the closest vertex. The hard-scatter vertex PV_0 is the one with the highest $\sum_{\text{tracks}} p_T^2$.

Electron and photon reconstruction

Electrons and photons are identified and reconstructed in a similar way [81–83]. The path of an electron in ATLAS is shown in Fig. 2.36. Until 2017, they were reconstructed with a fixed-size sliding window algorithm; since then, ATLAS uses a dynamical topological cell clustering (quite similar to the topo-cluster reconstruction to build jets, as described in Chapter 3, Section 3.1.1.1). Proto-clusters are seeded by cells in which the energy deposit has a significance of 4 times the noise threshold, then proto-clusters are selected and grouped into super-clusters. This allows to recover the photons from brehmsstrahlung, that was not possible with the previous algorithm. Photons and electrons are then selected according to isolation criteria to have a good purity, and calibrated.

Electrons are identified by a matching with at least one track in the ID. Tracks can also be tested for pattern recognition to distinguish pion and electron tracks. Electrons and positrons are distinguished by the curvature direction of their tracks. Photons are split into two categories: converted and unconverted photons. Unconverted photons are associated to no track, while converted ones are. Photons are identified based on the shower shape and energy deposition characteristics, to differentiate photons from fakes like $\pi^0 \to \gamma\gamma$.

Jets

Jets are the main topic of this thesis, and therefore Chapter 3 is dedicated to them: reconstruction, selection and calibration.

Muon reconstruction

A muon [84, 85] leaves a track in the MS, but also in the ID, and energy deposits in the calorimeters. To identify a muon, a track in the MS can be used alone, or combined with the ID track and also with the energy deposits in the calorimeter. MS tracks are therefore extrapolated inward to match ID tracks and calorimeter energy deposits, as well as to test the compatibility with the interaction point. Isolation criteria are also used, like for electron or photon identification, followed by a calibration.

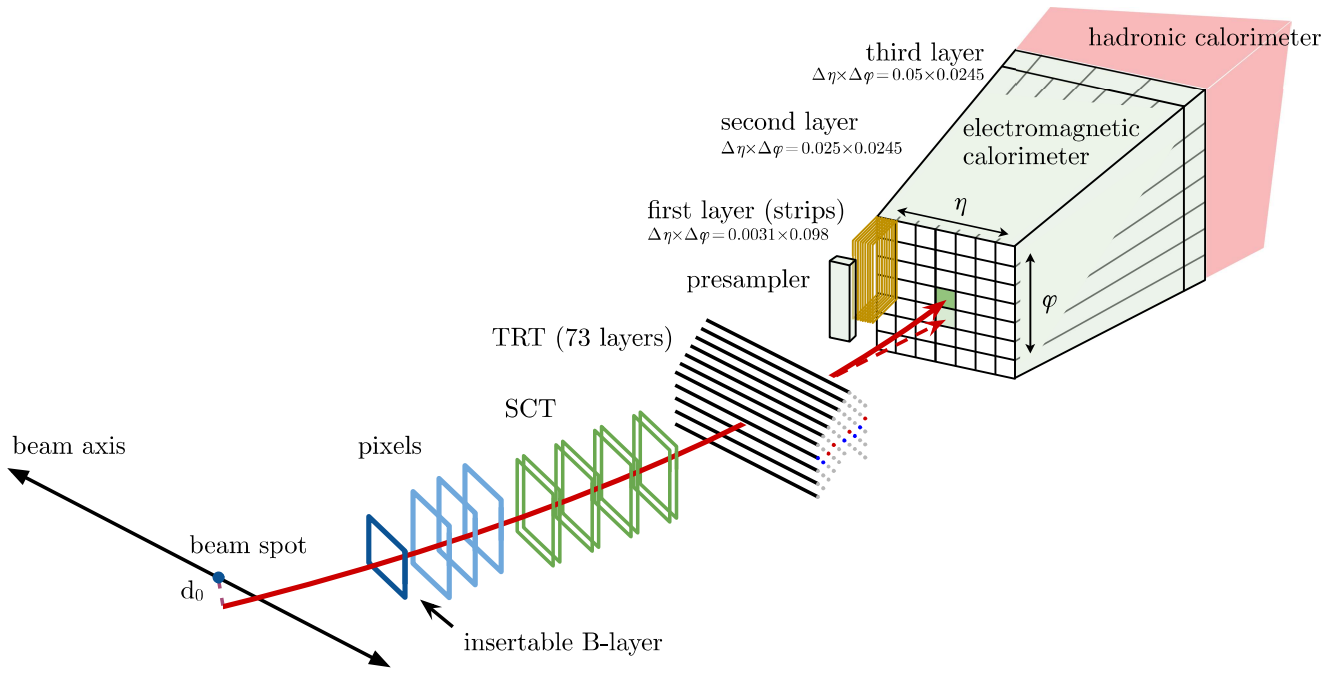


Figure 2.36: Path of an electron (red solid line) in the ATLAS detector. Reproduced from [81].

τ reconstruction

Due to their mean lifetime of 290.3 fs, τ leptons decay before reaching the ATLAS detector. They decay leptonically ($\tau \to l\nu_l\nu_\tau$ with $l \in \{e, \mu\}$) 34 % of the time, and hadronically ($\tau \to \nu_\tau$ + hadrons) 66 % of the time. Leptonic decays produce isolated electrons or muons, and neutrinos cause a momentum imbalance in the transverse plane (missing transverse momentum, see hereafter). Hadronic decays can be classified in two categories, the 1 prong and 3 prong decays, which contain respectively 1 and 3 charged particles (mostly $\pi^\pm$) with respective probabilities of 72 % and 22 %. In 68 % of cases, those hadronic decays contain at least one neutral pion. In the ATLAS detector, hadronic τ decays are reconstructed as jets with the anti-k_t clustering algorithm with radius parameter $R = 0.4$ (see Chapter 1, Section 1.3.2, and Chapter 3), and are required to have $p_T > 10$ GeV, $|\eta| < 2.5$ and one or three charged tracks within a cone of $\Delta R = 0.2$ around the jet axis [86]. The neutrino from the hadronic τ lepton decay cannot be reconstructed and the combination of all visible decay products is referred to as $\tau_{\text{had-vis}}$.

Missing-p_T reconstruction

Some particles do not leave energy deposits in the ATLAS detector, like neutrinos, or particles in models Beyond the Standard Model (BSM) like dark matter candidates. In this case, a large transverse momentum imbalance is observed. After a collision, the vectorial sum of the transverse momenta of all detected particles should be null. This quantity is called the missing transverse momentum $\vec{p}_T^{\text{miss}}$ (note that until recently, it used to be called "missing transverse energy", written $\vec{E}_T^{\text{miss}}$) [87]. It is computed from all hard objects (photons, electrons, muons, hadronic taus, jets), and a "soft term" is also taken into account (that takes into account all the low-energy contributions, like multi-parton interactions, thanks to the tracks that are matched to the primary vertex but not assigned to any hard object):

$$\vec{p}_T^{\text{miss}} = -\left(\underbrace{\sum_{\substack{\text{selected}\\\text{electrons}}} \vec{p}_T^{e} + \sum_{\substack{\text{accepted}\\\text{photons}}} \vec{p}_T^{\gamma} + \sum_{\substack{\text{accepted}\\\text{hadronic } \tau}} \vec{p}_T^{\tau_{\text{had-vis}}} + \sum_{\substack{\text{selected}\\\text{muons}}} \vec{p}_T^{\mu} + \sum_{\substack{\text{accepted}\\\text{jets}}} \vec{p}_T^{\text{jet}}}_{\text{hard term}} + \underbrace{\sum_{\substack{\text{unused}\\\text{tracks}}} \vec{p}_T^{\text{track}}}_{\text{soft term}} \right) \tag{2.24}$$

Chapter 3

The Reconstruction and Calibration of Jets in ATLAS

> There is a theory which states that if ever anyone discovers exactly what the Universe is for and why it is here, it will instantly disappear and be replaced by something even more bizarre and inexplicable. There is another theory which states that this has already happened.
>
> Douglas Adams, *The Restaurant at the End of the Universe*, 1980

Introduction

The particles constituting a jet, see Chapter 1, Section 1.3 (hereafter called a "particle-level jet" or a "truth jet") enter the ATLAS detector and produce tracks in the tracking systems and energy deposits in the calorimeters, see Chapter 2, Section 2.2. Those detector inputs have to be clustered together to get observables and form a reconstructed jet (hereafter called a "reco jet"), then a calibration has to be derived and applied on the jet observables to correct for detector inaccuracies, to obtain the properties of the parton that initiated the jet.

This is done in multiple steps in ATLAS. Tracks and energy deposits are clustered to obtain particle-like objects: Section 3.1.1. Those objects are themselves clustered together to form a jet with jet algorithms: Section 3.1.2. Jet observables then receive a series of corrections based on simulations, that capture the most important effects: Section 3.3.1. A final series of data-driven calibrations aims at correcting the residual differences between experimental data (hereafter called "data") and simulated data (hereafter called "simulation" or "MC"), and at evaluating the uncertainties associated to the calibration: Section 3.3.2. A measurement of the resolution in energy of the jets is performed both for data and for MC: Section 3.3.3.

The work presented in this thesis focuses on small-radius jets (hereafter called "small-R" jets, with $R = 0.4$ the radius parameter) reconstructed from particle-flow objects, either receiving a partial Jet Energy Scale (JES) calibration, in view of the η-intercalibration studies described in Chapter 4; or receiving a full jet energy scale calibration, for the cross-section measurements in Chapter 5. In the latter case, the numerous uncertainties on the JES and JER are propagated through the analysis; they are therefore listed and briefly described in this chapter.

I also ran the first steps of the η-intercalibration framework for large-radius jets (hereafter called "large-R" jets, with $R = 1.0$).[1] Those jets are reconstructed using the "unified flow objects" algorithm with the "constituent subtraction" and "soft-killer" pile-up rejection algorithms, which are therefore also briefly described hereafter.

[1] The work described in this thesis focuses on small-R jets produced with proton-proton collisions at a center-of-mass energy of $\sqrt{s} = 13$ TeV during the run 2 data-taking period of the LHC (from 2015 to 2018), in a high-pile-up environment (the average number of collisions per bunch-crossing during run 2 was $\langle\mu\rangle = 33.7$), as described in Chapter 2; therefore, jets produced in low-pile-up runs, during run 1, or originating from heavy ion collisions, are not described here. The η-intercalibration and the jet cross-section measurements are made inclusively in jet flavor (*i.e.* in the type of parton that originates the jet), therefore I have not used any flavor-tagging.

Other types of jets are commonly used in ATLAS, like track-jets [88], built solely from tracks in the tracking systems.

There are calibrations dedicated to jets produced in ATLAS that I did not use and which are therefore not described here: the b-JES [89, 90], dedicated to jets originating from b quarks, or the calibrations dedicated to the mass of the jets [88, 91].

3.1 Jet reconstruction

The jet clustering algorithms require the four-momentum of particles or particle-like inputs in order to deduce the properties of the parton that initiated the jet. The MC simulations provide both the particles originating from the parton, and the simulated tracks and energy deposits originating from the simulation of the interaction of these particles with the ATLAS detector, from which one can build respectively particle-level jets ("truth jets") or reconstructed jets ("reco jets"). Experimental data only provides tracks and energy deposits, from which reco jets are built.

As described in Chapter 1, Section 1.2.2, only a few particles have a long enough lifetime to reach the ATLAS detector:

- Electrons, muons and charged hadrons produce tracks in the tracking systems;
- Electrons, photons and charged hadrons produce energy deposits in the electromagnetic calorimeters;
- Hadrons produce energy deposits in the hadronic calorimeters;
- Muons produce tracks in the muon chambers;
- Neutrinos do not interact with the detector.

The detector inputs associated to electrons, photons and muons originating from the primary vertex are excluded from the object reconstruction used to build jets. All the other inputs, mainly tracks in the tracking systems and energy deposits in the calorimeters, are used to build particle-like objects for the jet reconstruction.

On top of that, the algorithms described hereafter serve the very important role of mitigating the effect of pile-up, by not adding to the particle-like clusters the tracks or energy deposits that most probably come from pile-up.

3.1.1 Particle-like object reconstruction

Multiple algorithms are used in ATLAS to build particle-like objects from the different detector inputs.

The Particle Flow Objects (PFOs), described hereafter, are the standard nowadays for small-R jets, although I have also run on jets built from Topological clusters when starting working on the jet cross-section analyses. I have run part of the η-intercalibration framework for the large-R analysis which are built from Unified Flow Objects, using pile-up mitigation techniques called "Constituent Subtraction" and "Soft Killer", that are therefore also described here.

The different algorithms can be viewed as sequential improvements of the previously existing algorithms:

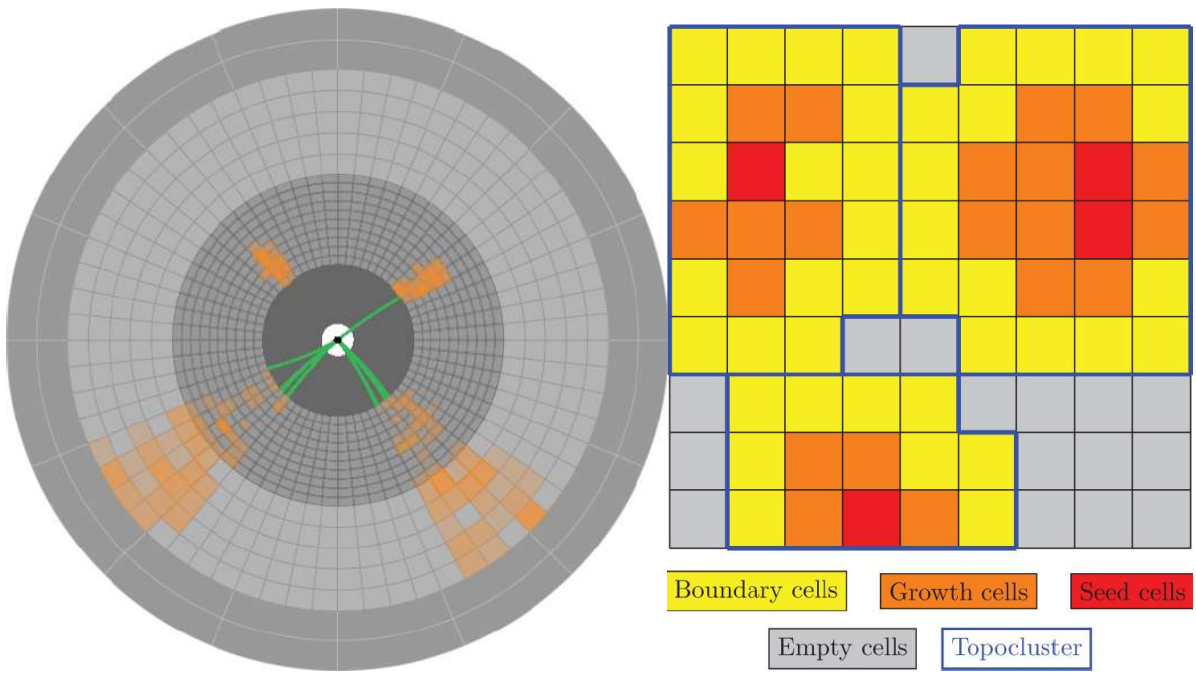
- Energy deposits in the calorimeters are clustered together with the topological clusters, while reducing the noise from pile-up energy deposits;
- Then the information of tracks is used to split or merge the topological clusters and improve the resolution with the Particle Flow algorithm, thus again improving the pile-up separation;
- Then the particle flow objects are split with the Track-CaloCluster algorithm to form Unified Flow Objects, improving the mass resolution at high p_T;
- Finally the Constituent Subtraction and Soft-Killer algorithms reduce the pile-up dependency.

3.1.1.1 Topological clusters

A hadron entering the ATLAS detector will deposit its energy in multiple calorimeter cells: in different calorimeter layers, and in different cells of one single layer, as shown by the illustration of Fig. 3.1a. The fine structure of the calorimeter system of ATLAS allows for a three-dimensional reconstruction of particle showers.

Energy deposits in the calorimeter cells are clustered together "topologically" based on the amplitude of the energy deposits: adjacent cells in the (η, ϕ) plane that produce a signal significantly different from noise are considered to have received energy deposits from particles originating from the same hadronic shower, thus the signals coming from these cells are added to form one single object called a "topo-cluster" [92].

Topo-clusters allow for the full reconstruction of the kinematic properties of the particles or ensemble of particles that initiated the shower, including location, direction, shape and energy, thus giving a full four-momentum as well as three-dimensional information on the shower shape. Topo-clusters allow to efficiently extract the signal from a background of electronic noise and other sources of fluctuations such as pile-up. Indeed, signals in cells that are not clustered in topo-clusters are discarded from further jet, particle and missing momentum reconstruction, as they probably originate from such sources of noise and fluctuations.



(a) (b)

Figure 3.1: **(a)** Schematics of the tracks in the tracking systems (in green) and energy deposits in the calorimeters (in orange) of a few particles originating from a single hard scatter. **(b)** Illustration of the topological clustering algorithm: cells with energy $|E_{\text{cell}}| > 4\sigma_{\text{noise,cell}}$ (in red) seed the topo-clusters, then the neighboring cells with $|E_{\text{cell}}| > 2\sigma_{\text{noise,cell}}$ (in orange) are added to the topo-clusters, finally the adjacent "boundary" cells are added (in yellow). The final topo-clusters are illustrated by the blue boundaries. Reproduced from [93].

Building topo-clusters

To build topo-clusters, the significance of a signal in a cell is defined as the ratio between the cell signal $E_{\text{cell}}^{\text{EM}}$ and the average (expected) noise $\sigma_{\text{noise,cell}}$:

$$\frac{E_{\text{cell}}}{\sigma_{\text{noise,cell}}} \tag{3.1}$$

where the average expected noise is defined as:

$$\sigma_{\text{noise,cell}}^{\text{EM}} = \sqrt{\left(\sigma_{\text{noise,cell}}^{\text{electronic}}\right)^2 + \left(\sigma_{\text{noise,cell}}^{\text{pile-up}}\right)^2} \tag{3.2}$$

where $\sigma_{\text{noise,cell}}^{\text{electronic}}$ is the electronic noise and $\sigma_{\text{noise,cell}}^{\text{pile-up}}$ is the expected noise from pile-up, taking the (expected) average number of interactions per bunch-crossing $\langle\mu\rangle$ into account.

Topo-clusters are "grown" sequentially, as illustrated by the cartoon of Fig. 3.1b:

1. Cells with a high signal significance $|E_{\text{cell}}| > 4 \cdot \sigma_{\text{noise,cell}}$ serve as seeds of the proto-cluster;
2. The cluster is grown by adding neighboring cells with $|E_{\text{cell}}| > 2 \cdot \sigma_{\text{noise,cell}}$ to the proto-cluster. Here, "neighboring" is defined as two calorimeter cells being directly adjacent in a given sampling layer, or, if in adjacent layers, having at least partial overlap in the (η, ϕ) plane. Proto-clusters can merge at this step.
3. The neighboring cells are added, without requirement on the significance. This allows the retention of cells with signals that are close to the noise levels while preserving the noise suppression feature of the clustering algorithm. Proto-clusters can also merge at this step.
4. Negative-energy topo-clusters are discarded.
5. Topo-clusters are split. To do so, local maxima are spotted in a topo-cluster based on the criterion $E_{\text{cell}} > 500$ MeV. The topo-cluster is divided into multiple sub-topo-clusters by associating the cells to the nearest local maximum (with a geometrical weighting in case a cell is adjacent to two maxima).

Computing the properties of the topo-clusters

The topo-cluster properties can then be computed from its constituting cells: from the energy and position of the cells. The complete formulas are given in [92].

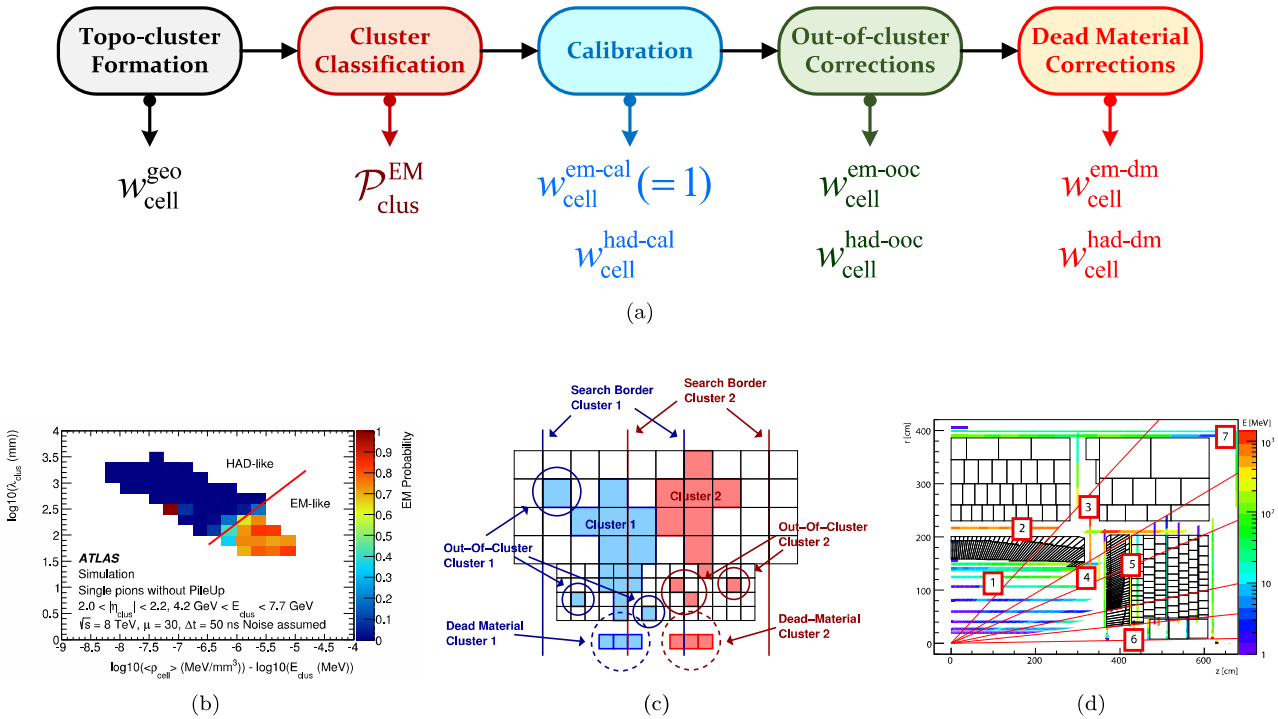


Figure 3.2: **(a)** Overview of the local hadronic cell-weighting (LCW) calibration scheme for topo-clusters. **(b)** Distribution of the likelihood for reconstructed topo-clusters to originate from an electromagnetic shower as a function of the shower depth λ_{clus} and the normalized cluster signal density $\rho_{\text{clus}}/E_{\text{clus}}{}^{\text{EM}}$, with ρ_{clus} being the energy-weighted average of the signal density. The red line indicates the boundary of 50 % likelihood, below which the topo-cluster is classified as mostly electromagnetic and above which it is classified as mostly hadronic. **(c)** Illustration of the assignment scheme for cells inside the calorimeter with true signal not captured in a topo-cluster in the context of the out-of-cluster correction and for dead material cells outside the calorimeter for the dead material correction. **(d)** The average energy loss in the virtual dead material cells for charged 100 GeV pions. The numbers 1 to 7 indicate the different regions, with region 8 (not displayed) being everywhere outside regions 1-7. The dead material cells are superimposed on a schematic (r, z) view showing a quarter of the ATLAS calorimeter system with its read-out segmentation. Reproduced from [92].

- The cluster directions η and ϕ are calculated as signal-weighted barycenters.
- The total energy of the cluster is a sum of the cell energies using geometrical weights.
- The momentum is then reconstructed thanks to the energy and direction of the cluster, by assuming that the topo-cluster is a massless pseudo-particle.
- Finally, the signal distribution inside a topo-cluster provides valuable information for signal characterization. Many observables can be constructed, called "cluster moments", some of which are used at later stages (for the local cell weighting, or for instance in the global sequential calibration, which is a step in the jet calibration, see Section 3.3.1.3).

Local hadronic Cell Weighting

The Local hadronic Cell Weighting (LCW) aims at correcting the energy of the topo-clusters on the basis of their geometrical and signal moments, for three effects, as shown in the flow chart of Fig. 3.2a:

- The non-compensating nature of calorimeter of ATLAS: the signal of the calorimeter to a hadron is smaller than the signal of an electron or photon depositing the same energy. Each topo-cluster receives a calibration based on the probability for it to be generated by an electromagnetic shower, and shown in Fig. 3.2b.
- Energy lost outside the topo-cluster. Due to the noise suppression of the topological clustering, a significant amount of energy is lost, mainly at the margins of the topo-cluster, as shown in Fig. 3.2c.
- Energy lost in dead material. The region in front of or between calorimeter modules is divided into virtual cells. MC simulations of 100 GeV single pions are used to determine the average energy deposited in those virtual cells, shown in Fig. 3.2d.

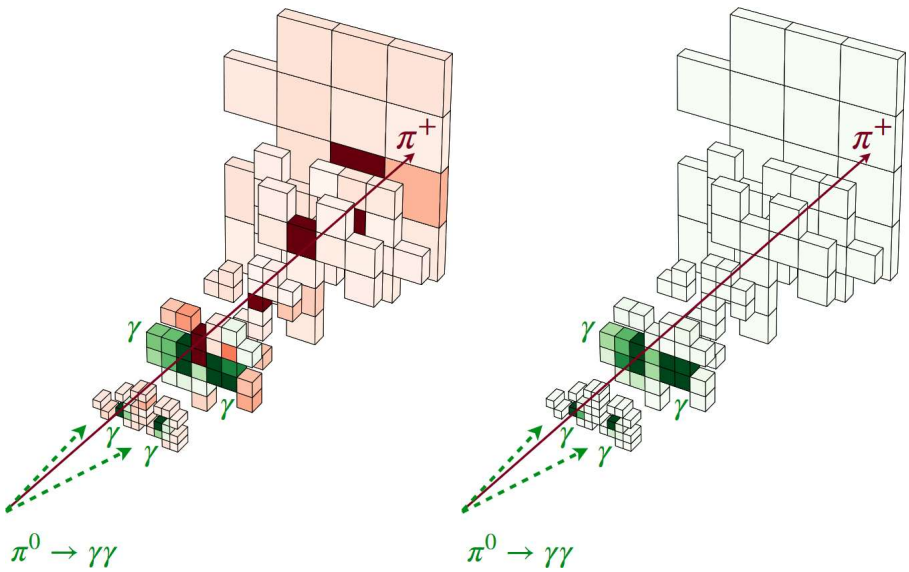


Figure 3.3: A 3-D display of the energy deposits of a π^+ and $\pi^0 \to \gamma\gamma$ in the calorimeters (left) and of π^0 only (right). The π^+ track and its extrapolation to the calorimeters is also displayed. Cells where the fraction of deposited energy from the π^+ is dominant are illustrated in red. The neutral energy deposits originating from the π^0 are otherwise illustrated in green. This illustrates the ideal result of the particle flow algorithm: the energy depositions made by the π^+ in the calorimeter would be removed, to only use the track for this particle, thus only leaving the contributions of neutral particles in the calorimeter inputs. Reproduced from [95].

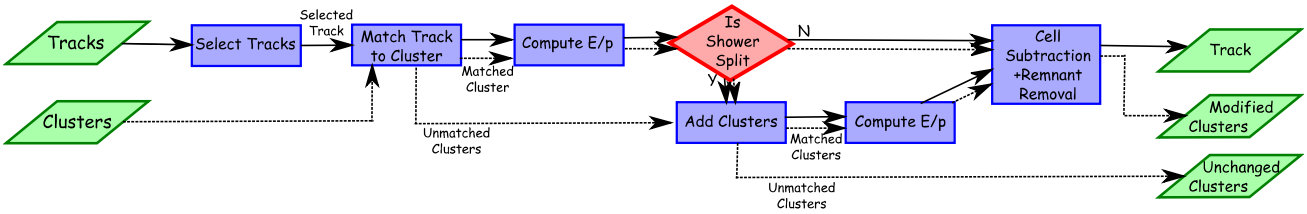


Figure 3.4: A flow chart of how the particle flow algorithm proceeds, starting with track selection and continuing until the energy associated with the selected tracks has been removed from the calorimeter. At the end, charged particles, topo-clusters which have not been modified by the algorithm, and remnants of topo-clusters which have had part of their energy removed remain. Reproduced from [94].

3.1.1.2 Particle Flow Objects

It it common for a particle to deposit energy in multiple topo-clusters, or for multiple particles to deposit energy in one single topo-cluster, as exemplified in Fig. 3.3 where a π^+ and a π^0 deposit energy in the same topo-clusters. To improve the jet reconstruction, it is possible to use the tracks made by the charged particles in the tracking systems, as they have better resolution at low p_T and can separate pile-up, by matching them to the topo-clusters.

This "particle flow" algorithm, hereafter also called "PFlow", improves on the topo-clusters described above [94].

Algorithm

Tracks that have a better energy resolution than the expected energy deposit in the calorimeter are selected, and are matched to topo-clusters. Tracks which are expected to be better measured in the tracker are selected and matched to topo-clusters. To avoid double-counting a particle that would have made both a track and energy deposits in the calorimeters, the expected energy of the track is subtracted from the topo-cluster(s) associated to this track. Only the track is kept for this particle; the remnants of the associated topo-cluster(s) are only kept as separate objects if they are significant. This leaves only two objects: the tracks come from charged particles, while the remnants of the topo-clusters come from neutral particles. The majority of tracks are generated by charged pions, particularly at low energy, therefore the pion mass hypothesis is assumed for all tracks.

The particle flow algorithm, summed up in the flow chart of Fig. 3.4, consists of the following steps:

1. Quality criteria are used to ensure the selection of well-measured tracks.
2. The track is matched to topo-clusters with two requirements: the pion must have deposited a large enough fraction of its energy in the topo-cluster: $E^{\text{cluster}}/p^{\text{track}} > 0.1$, and the extrapolated track and the barycenter

of the topo-cluster must be geometrically close with the criterion $\Delta R' < 1.64$, with $\Delta R' = \sqrt{\left(\frac{\Delta\phi}{\sigma_\phi}\right)^2 + \left(\frac{\Delta\eta}{\sigma_\eta}\right)^2}$ where σ_ϕ and σ_η represent the angular topo-cluster widths.

3. The expected energy deposited by the particle in the calorimeter is evaluated as $E^{\text{deposited}} = p^{\text{track}} \langle \frac{E^{\text{calorimeter}}}{p^{\text{track}}} \rangle$, with $\langle \frac{E^{\text{calorimeter}}}{p^{\text{track}}} \rangle$ determined using single-particle samples without pile-up by summing the energies of topo-clusters in a ΔR cone of size 0.4 around the extrapolated track position. The spread of the expected energy position $\sigma(E^{\text{cluster}})$ is also determined.
4. The algorithm determines whether the particle shower deposited the energy in one topo-cluster or more, based on the significance of the difference between the expected energy and the energy of the matched topo-cluster $S\left(E^{\text{cluster}}\right) = \frac{E^{\text{cluster}} - E^{\text{deposited}}}{\sigma(E^{\text{deposited}})}$: if $S\left(E^{\text{cluster}}\right) < -1$, a split shower recovery procedure is run. In this case, all the topo-clusters within a cone of $\Delta R = 0.2$ around the extrapolated track position are considered to be matched to the track.
5. The energy subtraction is performed on the set of matched topo-clusters. If the expected deposited energy $E^{\text{deposited}}$ exceeds the total energy of the set of matched topo-clusters, the topo-clusters are simply removed. Otherwise, subtraction is performed cell by cell, starting from the extrapolated track position in the calorimeter layer of highest energy density and then moving outwards in (η, ϕ) rings.
6. If the energy in the remnants of the topo-clusters is consistent with the width of the $E^{\text{calorimeter}}/p^{\text{track}}$ distribution, *i.e.* if it is less than $1.5\sigma(E^{\text{deposited}})$, it is assumed that the remnant energy originates from shower fluctuations, and so the energy in the remaining cells is removed.

Figure 3.5 illustrates how the particle flow algorithm deals with the different cases encountered: one or more particles, charged or not, producing one or more topo-clusters.

The Particle Flow Objects (PFOs, or PFlow objects) are the selected tracks and remaining topo-clusters.

Benefits

The particle flow objects have multiple benefits. The positional resolution of tracks is excellent, thus improving the overall position resolution of jets. It also allows for a considerable reduction of pile-up effects, by allowing to exclude from the jet reconstruction the tracks that are inconsistent with the primary vertex, as they likely come from pile-up. The jet energy resolution is also improved at low p_T, due to the reduced impact of pile-up, and to the use of well-measured tracks [94].

3.1.1.3 Constituent-level pile-up rejection

Pile-up contributions can be further removed from the particle-like input, prior to jet reconstruction. The PFlow algorithm provides a very efficient and simple way to mitigate the effect of pile-up, called the "Charged Hadron Subtraction" (CHS) [96]. However, it only removes pile-up contributions originating from charged particles. Algorithms called "Constituent Subtraction" (CS) and "Soft Killer" (SK) have proven to improve the energy resolution of low-p_T jets and to reduce the dependency to pile-up, both for large-R jets (see [97]) and for small-R jets (see [96]). Figure 3.6 shows an example event display with all the constituents shown before pile-up mitigation (Fig. 3.6a), after Constituent Subtraction (Fig. 3.6b) and after Constituent Subtraction and Soft Killer (Fig. 3.6c). CS and SK have however only been used for large-R jets in ATLAS until now, because these algorithms are fairly new, because large-R jets are more sensitive to pile-up due to the increased cone size and because the jet mass (not used for small-R jets) is sensitive to high-angle soft radiations.

Charged Hadron Subtraction

Particle Flow Objects (PFOs) are divided into two categories: charged PFOs (tracks) and neutral PFOs (modified or unmodified topo-clusters). Tracks have a very good angular resolution, allowing to identify whether the charged PFO originates from the primary vertex (PV) or from pile-up, by requiring $|z_0 \sin(\theta)| < 2$ mm, where z_0 is the distance of closest approach of the charged PFOs to the hard-scatter primary vertex along the z axis, and θ is its polar angle. Charged PFOs that do not satisfy this requirement are discarded.

Constituent subtraction

Pile-up is essentially constituted of soft contributions (*i.e.* low energy particles), approximately uniformly distributed in the (η, ϕ) plane. Therefore, at first order, pile-up has a constant density ρ and contributes uniformly onto the area of the constituent.

The pile-up transverse momentum density ρ is estimated as the median of the $\frac{p_T}{A}$ distribution of all k_t jets with $R = 0.4$ in the event, with A the jet area determined by adding a large number of soft particles (called ghosts) to the event: the area of the jet corresponds to the region in the (η, ϕ) plane where the ghosts are clustered into the jet.

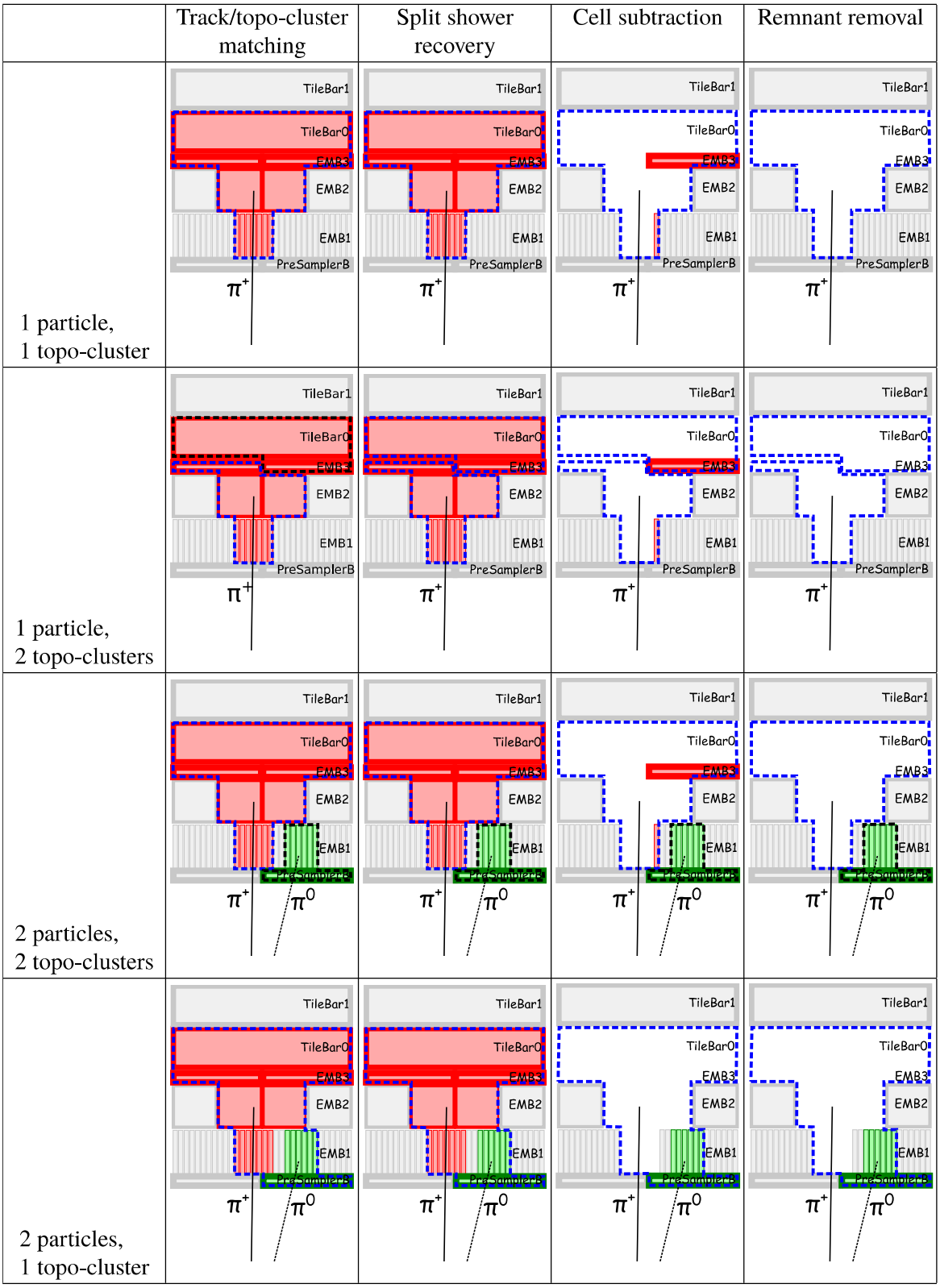
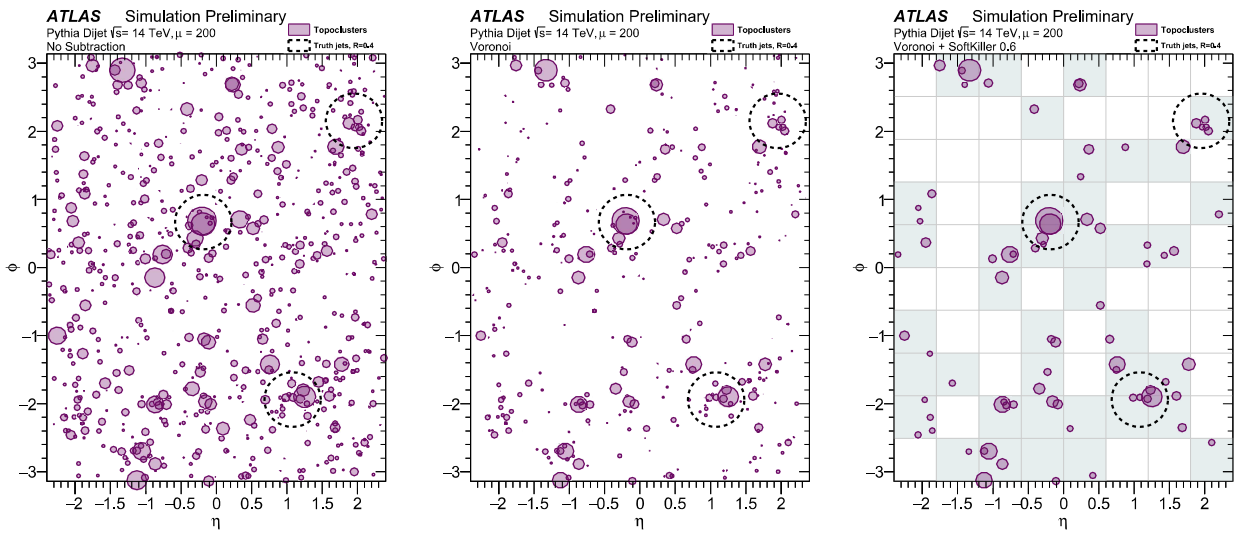


Figure 3.5: Idealized examples of how the algorithm is designed to deal with several different cases. The red cells are those which have energy from the π^+, the green cells energy from the photons from the π^0 decay, the dotted lines represent the original topo-cluster boundaries with those outlined in blue having been matched by the algorithm to the π^+, while those in black are yet to be selected. The different layers in the electromagnetic calorimeter (Presampler, EMB1, EMB2, EMB3) are indicated. In this sketch only the first two layers of the Tile calorimeter are shown (TileBar0 and TileBar1). Reproduced from [94].



(a) (b) (c)

Figure 3.6: Example event display **(a)** before pile-up removal and **(b)** after VORONOI Subtraction is applied to constituents (which is a pile-up mitigation technique close to the Constituent Subtraction, described in the same paper [97], that mostly differs on the way the area of the constituent is defined: the VORONOI subtraction uses the VORONOI area of the constituent), and **(c)** after VORONOI Subtraction and Soft Killer are applied. The size of the markers corresponding to the constituents is proportional to their p_T, and the reconstructed jets (with dashed lines) are those which are constructed after constituent-level pile-up removal has been applied. Grey squares indicate grid spaces which do not contain clusters after applying SoftKiller. Reproduced from [97].

The Constituent Subtraction (CS) algorithm corrects the transverse momentum of each constituent (which can be a topo-cluster or a particle-flow object) by subtracting the estimated contribution of pile-up.

Massless low-p_T virtual constituents ("ghosts") are added to the event in $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ grid, with an assigned transverse momentum of $p_T^g = A_g \times \rho$, with $A_g = 0.01$ the "area" of the ghost.

Then, the CS algorithm iteratively subtracts the contributions of the ghosts from the constituents (with a geometrical ordering); in case a constituent has less transverse momentum than a ghost, it is simply removed (and the ghost's p_T is decreased by the associated quantity). This is performed until no constituent-ghost pair with a ΔR less than 0.25 is left.

Soft Killer

The Soft Killer (SK) is a hard cut on the constituents' p_T to decide whether they are retained or rejected. The SK therefore deals with local pile-up fluctuations, and works particularly well in combination with the Constituent Subtraction since the CS amounts to a first order pile-up calibration.

The SK sets the median flow density of each event to zero by removing the softest constituents.

To that extent, the (η, ϕ) space is divided into a 0.6×0.6 grid, and a p_T cut is chosen such that half of the grid spaces are empty.

3.1.1.4 Unified Flow Objects

The PFlow algorithm of ATLAS has a great performance at low p_T (roughly below 50 GeV), but can be improved in the highly-boosted regime (roughly above 500 GeV). Indeed, it relies on tracks providing a very good measurement of the momentum of a particle, but high energy particles have a trajectory in the tracking systems that is less bent than lower energy particles, thus reducing the momentum resolution. Furthermore, in the highly-boosted regime, objects can be merged into the same jet (or even in the same topo-cluster), thus reducing the energy and mass resolution.

An alternative "particle flow" algorithm has been devised in ATLAS to improve the jet substructure, flavor tagging, and the energy and mass resolutions in the high energy regime: the Track-CaloCluster (TCC) algorithm. This algorithm has recently been merged with the PFlow algorithm and with the pile-up mitigation techniques described above: the Unified Flow Objects (UFOs). The UFO algorithm is now the standard one used in ATLAS for large-R jets [98], and will also become standard for small-R jets as well [96] for Run 3 precision analyses.

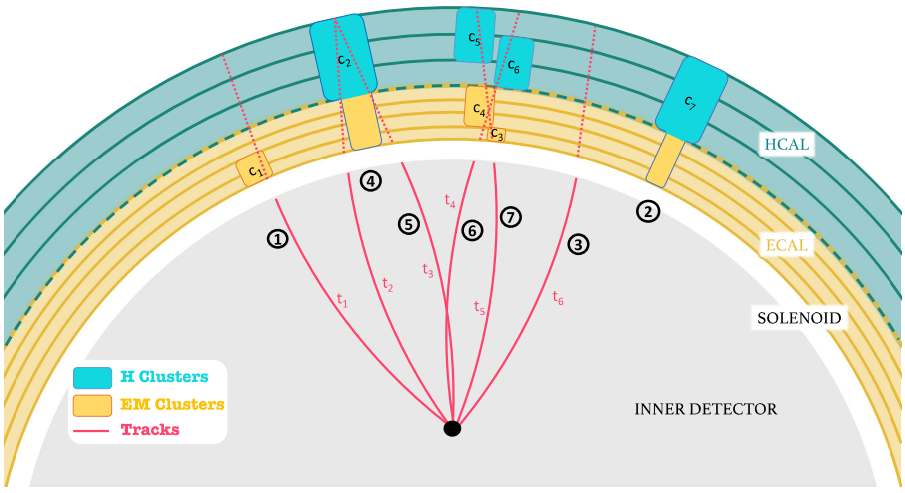


Figure 3.7: A schematic demonstrating the creation of seven TCC objects representing (1) a simple track-cluster match, (2) a topo-cluster without a matching track, (3) a track without a matching cluster, (4) and (5) are each tracks matching a single cluster but sharing that cluster's energy, and (6) and (7) showing a much more complex scenario with multiple track-cluster matches. Reproduced from [99].

Track-CaloClusters

The goal of the Track-CaloCluster (TCC) algorithm is to take advantage of the very good energy resolution of the calorimeter of ATLAS and of the very good angular resolution of the tracks of ATLAS, both being especially good at high energy. The typical extrapolated track angular uncertainty is significantly smaller than the size of the average topo-clusters, especially for moderate and high p_T tracks (roughly above 10 GeV). This would allow to resolve many (charged) particles; but the energy of these particles should not be double-counted (once in the tracks, and once in the topo-clusters).

The difficulty is that topo-clusters and tracks are at different energy scales, as the tracks only represent the electromagnetic energy, from charged particles only, while the topo-clusters are corrected on average to the hadronic scale. Therefore, the goal of the TCC algorithm is to use the angular information from tracks and the information of energy and mass coming from the topo-clusters, while never directly comparing tracker and calorimeter energy measurements.

The idea behind the TCC algorithm is to match tracks to topo-clusters and to associate part of the energy and mass of the matched topo-clusters to the tracks. If multiple tracks are matched to a topo-cluster (or more than one), the energy and mass of that topo-cluster is split between the multiple tracks. There is therefore one TCC object per track. The four-momentum of the TCC object is therefore built from the η and ϕ angles of the track, and from a weighted combination of the p_T of the topo-clusters matched to that track (and also of their mass, though in case of topo-clusters the mass is assumed to be 0). In case no topo-cluster is matched to a track, the TCC object generated from that track is therefore only the track itself, with its whole four-momentum. Similarly, if no track is matched to a topo-cluster, the topo-cluster is used as a TCC object. The different situations are showcased in Fig. 3.7.

The matching is done with two criteria: the extrapolation uncertainty of the track must be smaller than the topo-cluster's width (otherwise, the topo-cluster's angular information is better than the track's angular information), and the angular separation must satisfy $\Delta R < \sqrt{\sigma^2_{\text{cluster}} + \sigma^2_{\text{track}}}$, with σ_{cluster} the topo-cluster's width and σ_{track} the track extrapolation uncertainty.

The weights used for the combination of the p_T and mass of the topo-clusters matched to one or more tracks are given in [99]. They evaluate the contribution of a track to a topo-cluster compared to all the other tracks that contribute to the topo-cluster; they also evaluate the contribution of a topo-cluster from a track compared to all topo-clusters that get a contribution from this track.

This way, tracker and calorimeter energy measurements are never directly compared: scale factors derived from the tracking systems are used to weight the energy split from the calorimeters.

The main difference with the PFlow algorithm is that PFlow uses tracking to correct the full four-vector of a topo-cluster, while TCCs do not rely on the measurement of the energy made by the tracking systems, which is worse than the calorimeter measurements at medium and high p_T.

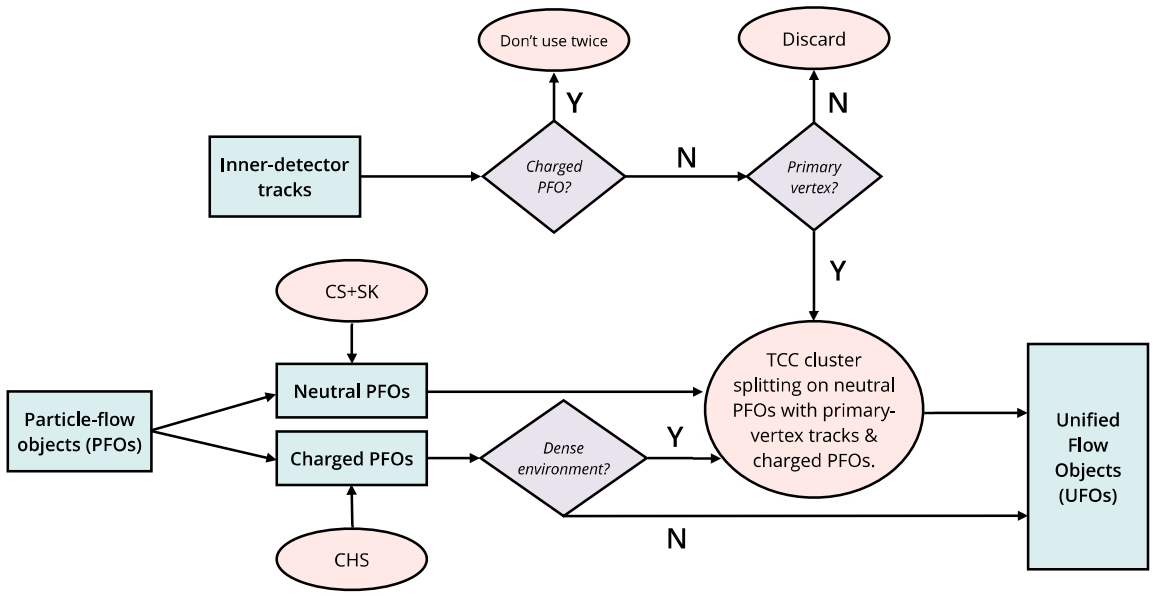


Figure 3.8: An illustration of the Unified Flow Object reconstruction algorithm. Reproduced from [98].

Unified Flow Objects: combinations of PFOs and TCCs
The PFlow and TCC algorithms improve the performance upon simple topo-clusters by using the information of tracks, but they do it differently, and are more performant in different regimes: while PFOs are less sensitive to pile-up and improve the angular and energy resolutions at low p_T, TCCs improve the tagging performance, energy resolution and mass resolution at high p_T. The Unified Flow algorithm aims at combining the two approaches to benefit from both in the phase space regions where they excel.

The Unified Flow algorithm uses as inputs the Particle Flow Objects (PFOs), with constituent pile-up rejection techniques applied: the Charged Hadron Subtraction (CHS) is applied to the charged PFOs (which are tracks), while the Constituent Subtraction (CS) and Soft-Killer (SK) are applied to neutral PFOs (modified or unmodified topo-clusters). Then, the tracks that have not been used for the PFlow subtraction (either because they have a high momentum or because they were located in a dense environment) and that are matched to the hard-scatter vertex are used for the TCC algorithm to split the neutral PFOs. A schematics of the algorithm is given in Fig. 3.8.

3.1.2 Jet reconstruction

3.1.2.1 Jet algorithm

After having built the particle-like inputs from the calorimetric and tracking information, a jet algorithm is run to cluster them into jets, as described in Section 1.3.2.

The main jet algorithm used in ATLAS is the anti-k_t algorithm, either:

- with a radius parameter of $R = 0.4$ for small-radius jets, which are the main focus of this thesis;
- or with a radius parameter of $R = 1.0$ for large-radius jets.

These choices result from an optimization discussed in Section 1.3.2, based on the physics studied and the energy of the processes of interest. Unified choices between the experiments (like ATLAS and CMS) also allow for easier comparisons and combinations of the results.

Note that other algorithms are used in ATLAS for specific studies, like the k_t algorithm (for instance to evaluate the pile-up density thanks to "ghosts", i.e. infinitely soft particles, since this algorithm clusters soft elements with a more unified background more than the anti-k_t algorithm), or like the Cambridge-Aachen algorithm (for instance in the "Soft Drop" approach of large-R jet grooming, since this algorithm provides an angle-ordered jet clustering history). Other radius parameters are also used in ATLAS, like $R = 0.2$ or $R = 0.6$.

3.1.2.2 Large-R jet grooming

Jets are constructed from particle-like inputs, themselves built with algorithms designed to mitigate the effect of pile-up as much as possible. It it possible to reduce the pile-up dependency even further at the jet level, with

techniques called "grooming" algorithms, that consist in removing the regions of the jet that probably originate from pile-up. Multiple such techniques have been tested and are described in [98]: Trimming, Pruning, Soft-Drop, and variations of the Soft-Drop. The one currently used for large-R jets (not used for small-R jets yet) is the Soft-Drop.

The Soft-Drop technique aims at removing soft and wide-angled radiations from a jet. The jet is re-clustered with the Cambridge-Aachen jet algorithm, that provides an angle-ordered jet clustering history, then the clustering sequence is traversed starting from the widest-angled radiation towards the jet core: at each step, a condition is tested, that decides whether the lower-p_T branch should be dropped or not.

3.2 Quality criteria

To ensure that the jets come from the hard-scatter collision of interest and are of good quality, a few selection criteria can be used on all the jets reconstructed from the jet algorithms described before.

3.2.1 Pile-up rejection

Pile-up is the general term qualifying the energy deposits and tracks not originating from the hard-scatter vertex of interest. It is categorized as in-time pile-up and out-of-time pile-up. In-time pile-up qualifies the energy deposits and tracks coming from other proton-proton collisions occurring during the same bunch crossing as the event of interest. Out-of-time pile-up qualifies all the other signals that do not come from the event of interest; for instance, some parts of the detector and/or electronics have an integration time that is longer than the time between bunch crossings, as is the case for the liquid argon calorimeters.

Pile-up affects the determination of many observables, especially jet energy, p_T and mass, but also sub-structure variables or flavor tagging. Pile-up was already at a high level during LHC's Run 2, with an average number of primary vertices per bunch-crossing of $\langle\mu\rangle = 33.7$, even reaching more than 70 primary vertices. The number of primary vertices has already increased during the 2022 data-taking period of Run 3 ($\langle\mu\rangle = 42.5$) and is planned to go even higher during the future data-taking periods, especially during the "High-Luminosity LHC" (HL-LHC), scheduled for 2029, that will increase the instantaneous luminosity by a factor 5 to 7.5. Pile-up resilient techniques are important for the analysis of the data-sets already recorded and crucial to avoid a degradation of the performance in the future analyses.

The effect of pile-up is mitigated in multiple steps, some of which have already been described so far:

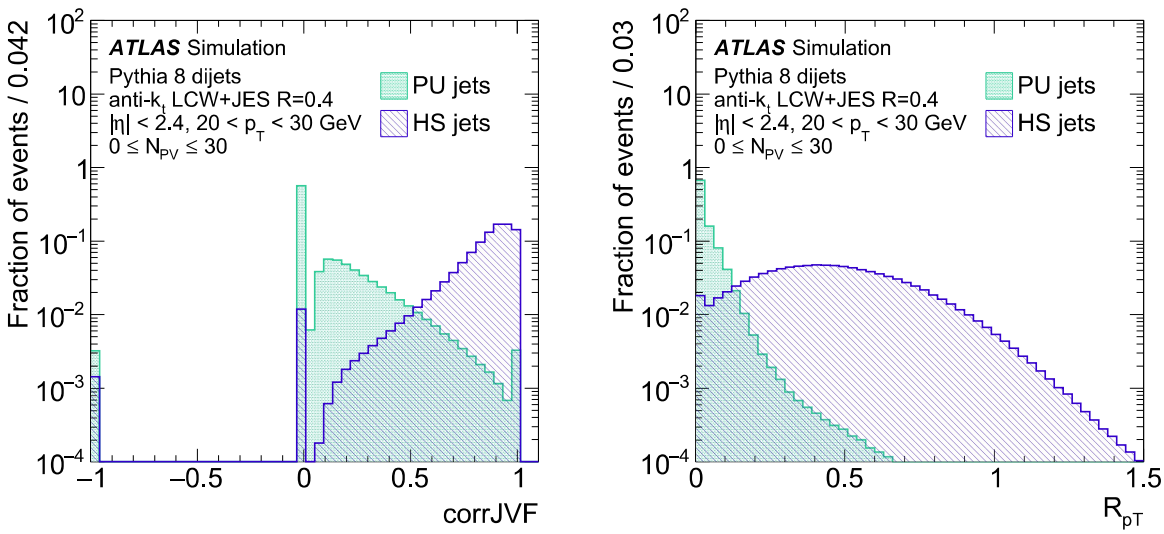
- During the creation of particle-like objects: topo-clusters reject pile-up contributions in the calorimeters while the PFlow algorithm coupled to the Charged Hadron Subtraction use the great angular and timing resolution of tracks to identify PFOs that do not originate from the hard-scatter vertex. The Constituent Subtraction and Soft-Killer algorithms are also used in the context of large-R jets to reject pile-up contributions from topo-clusters.
- During the creation of jets, where the main jet algorithm used is anti-k_t, which (among other properties) is IRC-safe and soft-resilient. In the context of large-R jets, further subjet-level pile-up rejection is done with jet grooming.
- During the jet energy calibration, where the effect of pile-up is mitigated by the jet area correction and the residual calibration.
- At the jet level, with the Jet Vertex Tagger (JVT) described in this section, that allows to reject entire jets that have a high probability to come from pile-up [100].

The Jet Vertex Tagger (JVT) is a two-dimensional likelihood based on the corrJVF and R_{p_T} variables that characterizes the relative probability for a jet to originate from signal, in the central and low-p_T phase-space region: $20 < \mathrm{p_T} < 50$ GeV and $|\eta| < 2.4$.[2] The two variables used are defined as follows:

$$\mathrm{corr}JVF = \frac{\sum_m p_{T,m}^{\mathrm{track}}(\mathrm{PV}_0)}{\sum_l p_{T,l}^{\mathrm{track}}(\mathrm{PV}_0) + \frac{\sum_{n\geq 1}\sum_l p_{T,l}^{\mathrm{track}}(\mathrm{PV}_n)}{k \cdot n_{\mathrm{track}}^{\mathrm{PU}}}} \tag{3.3}$$

$$R_{p_T} = \frac{\sum_m p_{T,m}^{\mathrm{track}}(\mathrm{PV}_0)}{p_T^{\mathrm{jet}}} \tag{3.4}$$

[2]The JVT method is based on tracks, and therefore is limited by the acceptance of the tracker; new techniques based on missing transverse momentum are used to identify pile-up jets in the forward region: if a forward jet can resolve the momentum imbalance of a pile-up vertex, it is probably a pile-up vertex [101]. This method, called the forward Jet Vertex Tagger ($fJVT$), cannot be used in the analyses presented in this thesis, and is therefore only mentioned for completeness.



(a) (b)

Figure 3.9: Distributions of **(a)** corrJVF and **(b)** R_{p_T} for pile-up (PU) and hard-scatter (HS) jets with $20 < \mathrm{p_T} < 30$ GeV. Reproduced from [100].

with PV_0 the primary vertex, PV_n with $n > 0$ the pile-up vertices, $p_{T,m}^{\mathrm{track}}(\mathrm{PV}_n)$ the transverse momentum of tracks associated to the jet of interest and coming from the vertex n (either the primary vertex or a pile-up vertex), $n_{\mathrm{track}^{\mathrm{PU}}}$ the total number of pile-up tracks per event (*i.e.* all the tracks associated with vertices other than the hard-scatter vertex), $k = 0.01$ the scaling factor capturing the dependency of $\langle p_T^{\mathrm{PU}} \rangle$ with $n_{\mathrm{track}}^{\mathrm{PU}}$, and p_T^{jet} the fully calibrated transverse momentum of the jet of interest (including the pile-up subtraction).

Both variables aim at capturing the fraction of the jet's p_T that comes from tracks matched to the hard-scatter vertex: $\sum_m p_{T,m}^{\mathrm{track}}(\mathrm{PV}_0)$. corr$JVF$ compares it to the total p_T of the jet coming from all the track matched to any vertex (with a correction for the dependency to the number of primary vertices: the more primary vertices, the higher the denominator, which would make this variable dependent on the number of primary vertices N_{PV}); corrJVF is therefore interpreted as the fraction of the jet's p_T that originates from the primary vertex. R_{p_T} compares it to the total p_T of the calibrated jet, including information from the calorimeter, and is therefore equivalent to the fraction of the p_T of the jet that comes from charged particles from the primary vertex. The distributions of these variables are shown in Fig. 3.9.

The track-to-vertex association uses the vertex reconstruction that assigns tracks to vertices. In case a track is attached to more than one vertex, priority is given to the vertex with higher $\sum \left(p_T^{\mathrm{track}}\right)^2$. Then, if a track is not associated with any vertex but satisfies $|\Delta z| < 3$ mm with respect to the hard-scatter primary vertex, it is assigned to the hard-scatter primary vertex (to take the B meson decay into account).

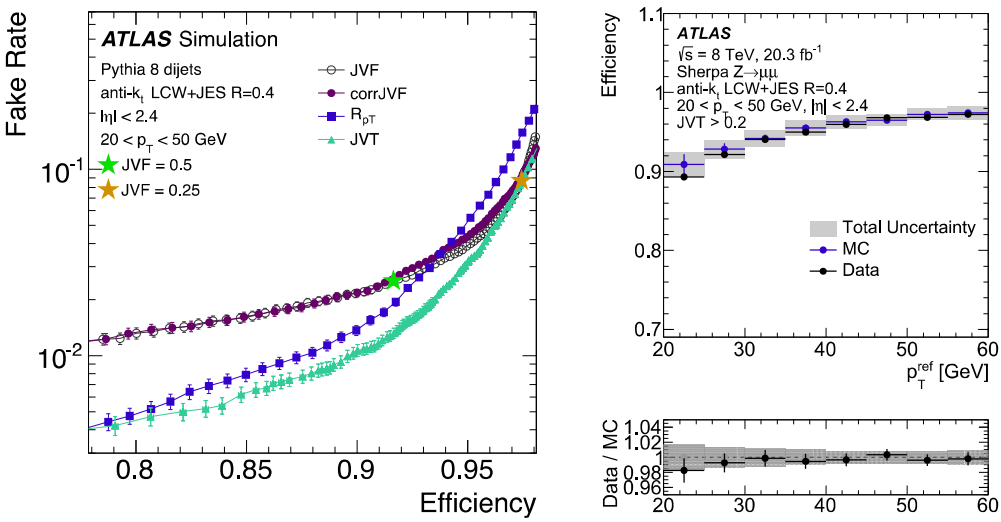
The likelihood is evaluated on an MC sample of $Z(\rightarrow ee)$+jets sample, where the EMTopo jets geometrically matched to truth particles from the hard-scatter are considered hard-scatter jets, and the others are considered pile-up jets (with the geometrical matching being $\Delta R < 0.4$). The fake rate versus efficiency curve of the JVT selection criterion is shown in Fig. 3.10a. The standard working point of the JVT rejects jets below $JVT = 0.2$, and the tight working point rejects jets below $JVT = 0.5$. The efficiency curve for the 0.2 working point as a function of p_T is shown in Fig. 3.10b.

3.2.2 Event-level selection

Another measure to ensure good quality of the events is the event-level cleaning, described in [102]. This cleaning is a set of identification criteria, that aim at rejecting the following backgrounds:

- Beam-induced background due to proton losses upstream of the interaction point that induce muons which reach the ATLAS detector.
- Cosmic-ray showers produced in the atmosphere, overlapping with collision events (mostly muons).
- Calorimeter noise from large scale coherent noise or isolated pathological cells.

The study to identify the variables able to discriminate between signal and noise have been performed on MC simulated samples of dijet events; then the selection efficiency has been studied *in situ* on data events. A signal



(a) (b)

Figure 3.10: **(a)** Fake rate from pile-up jets versus hard-scatter jet efficiency curves for JVF, corrJVF, R_{p_T}, and JVT. The JVF working points 0.25 and 0.5 are indicated with gold and green stars (the working points for the JVT are 0.2 and 0.5). **(b)** Efficiency measured in $Z(\rightarrow \mu\mu)$+jets events as a function of $p_T^{\text{ref}} = p_T^Z$ in data and MC simulation for $JVT > 0.2$. Reproduced from [100].

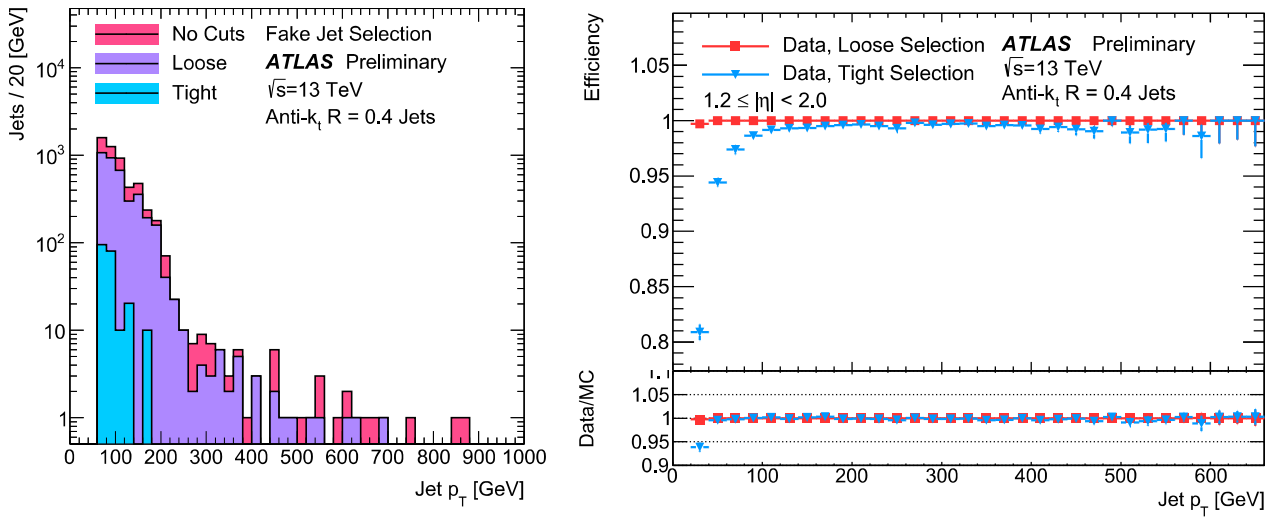
event, called a "good" event, is required to have the two leading jets with $p_T > 70$ GeV that are back-to-back in the transverse plane ($\Delta\phi_{j-j} > 3.0$) and balanced in transverse momentum ($\frac{|p_T^1 - p_T^2|}{p_T^1 + p_T^2} < 0.3$). A noise event, called a "bad" or a "fake" event, is required to have at least one jet with $p_T > 70$ GeV, $H_T^{\text{miss}} = \left|\vec{H}_T^{\text{miss}}\right| > 70$ GeV (with $\vec{H}_T^{\text{miss}} = -\sum_{\text{jets}} \vec{p_T}$ where all jets with p_T greater than 20 GeV are considered), the direction of $\vec{H}_T^{\text{miss}}$ should be opposite to the transverse component of the jet momentum ($\Delta\phi_{H_T^{\text{miss}}-j} > 3.0$), and the leading jet is required to be out-of-time ($|t_{\text{jet}}| > 6$ ns, with the jet time being the energy-squared weighted average of the time of the energy deposits in the jet). These samples enriched with good and bad events are also used to obtain the working point, *i.e.* the exact values of the selection criteria using those discriminating variables, as well as the associated efficiency of selection of good jets and efficiency of rejection of bad jets.

The variables used to discriminate between noise and signal are the following ones:

- The quadratic difference Q_{cell}^{LAr} between the actual and expected pulse shapes in the liquid argon calorimeter provides a measure of a pulse quality and is used to define:
  - $\langle Q \rangle$: The average jet quality, defined as the energy-squared weighted average of Q_{cell}^{LAr};
  - f_Q^{LAr} and f_Q^{HEC}: The fraction of the jet energy in the LAr calorimeter cells (HEC calorimeter cells respectively), see Chapter 2, of a jet with poor signal shape quality, defined as Q_{cell}^{LAr};
- E_{neg}: The sum of all the negative energy deposits in the jet;
- f_{EM} and f_{HEC}: The electromagnetic fraction and the HEC energy fraction are the ratio of the energy deposited in the electromagnetic calorimeter (HEC calorimeter respectively) to the total energy of the jet;
- f_{max}: The ratio of the maximum energy deposited in a single calorimeter layer to the total energy of the jet;
- f_{ch}: The jet charged fraction is the ratio of the scalar sum of the p_T of the tracks coming from the primary vertex associated to the jet divided by the jet p_T.
- $f_{\text{ch}}/f_{\text{max}}$ is also found to discriminate between good and fake events.

Those variables discriminate between the fake events coming from the different sources and the good events, as follows:

- The calorimeter noise is characterized by a bad signal shape, hence by the variables $\langle Q \rangle$, f_Q^{LAr} and f_Q^{HEC}. The sporadically noisy cells can generate large fake energy deposits, hence the use of E_{neg}.
- The beam-induced background and the calorimeter noise tend to be more localized longitudinally in the calorimeters than jets from proton-proton collisions, hence the use of the variables f_{EM}, f_{HEC} and f_{max}. The track-based variables f_{ch} and $f_{\text{ch}}/f_{\text{max}}$ are also good at identifying beam-induced background.
- The cosmic-ray showers are identified by combinations of these variables: f_{max}, f_{EM} and f_{ch}.



(a) (b)

Figure 3.11: Jet transverse momentum distribution of **(a)** the fake-enriched sample (inclusive in η) and **(b)** the good-enriched sample (in the $1.2 \leq |\eta| < 2.0$ bin). Reproduced from [102].

The selection criteria using those variables are detailed in [102]: the two working points are called "BadLoose" and "BadTight".

The efficiency of those selections has been studied *in situ* with a dijet technique: for dijet events that are back-to-back in the transverse place, if one jet passes tight requirements, it is looked whether or not the second jet passes the BadLoose or "BadTight" requirements. The "BadLoose" working point is designed to provide a high selection efficiency of good events while rejecting as many fake events as possible; its efficiency for good jets is above 99.5 % (respectively 99.9 %) for $p_T > 20$ GeV (respectively $p_T > 100$ GeV). The "BadTight" working point is designed to further reject the background events; its efficiency for good jets is above 95 % (respectively 99.5 %) for $p_T > 20$ GeV (respectively $p_T > 100$ GeV). The efficiency of selection of good jets and of rejection of bad jets is shown in Fig. 3.11 for both working points.

3.2.3 Timing selection

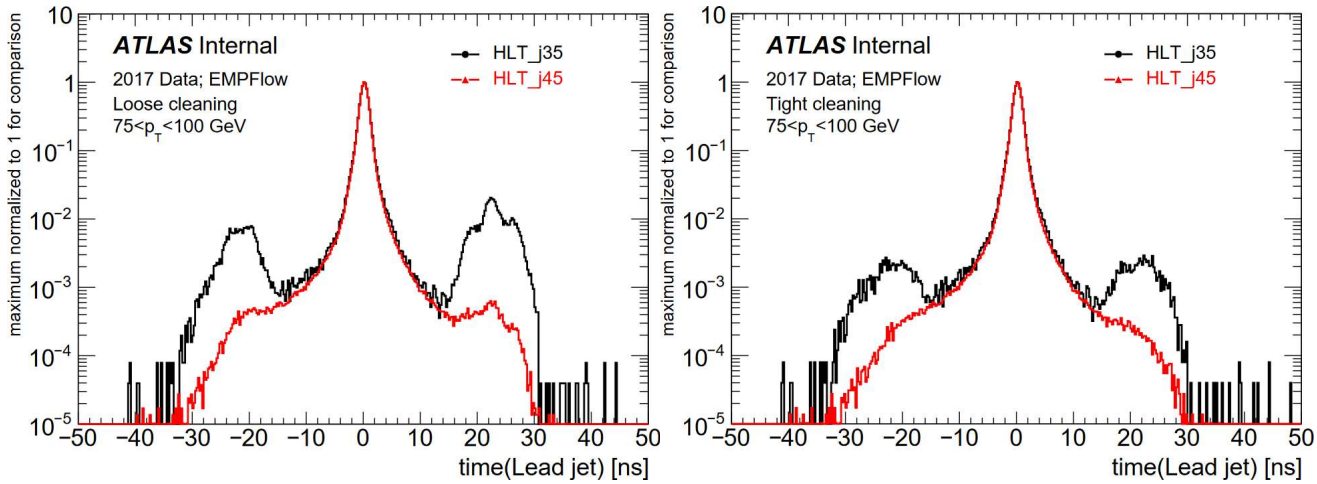
The calorimeter has a large charge collection time (hundreds of nanoseconds) compared to the bunch-crossing interval (25 ns). Jets reconstructed from energy depositions of neighbouring bunch crossings therefore constitute noise for the event of interest, and are called out-of-time pile-up. They are already partially identified by the JVT, but can be further rejected by a timing selection.

The time of a LAr calorimeter cell is defined with respect to the event time recorded by the trigger (which is synchronized with the bunch-crossing time) from the signal shape of the LAr calorimeter response. The jet time t_{jet} is defined as the energy-square-weighted average of the times of the LAr calorimeter cells reconstructed within the jet.

The timing of the leading jet has been studied in [103] to determine a selection criterion: a jet is considered out-of-time if $|t_{\text{jet}}| > 10$ ns. The jet time distributions can be observed in Fig. 3.12 for EMPFlow jets, both for the BadLoose and BadTight working points of the jet cleaning, for two triggers (HLT_j35, that is random at L1 and thus integrating over multiple bunch crossings, and HLT_j45 that is L1-calo based thus integrating over only one bunch crossing; see Sections 2.2 and 4.2.1.2 for information on triggers).

3.2.4 Jet-level selection

The "standard" event-level cleaning described previously in Section 3.2.2 works very well in the central and low-p_T region, but is not enough for the jet cross-section measurements, for multiple reasons. The standard cleaning is not efficient enough at removing the fake jets in the high η or high p_T regions, it has been optimised with a (relatively) small sample of MC events thus not allowing for an *in situ* estimation of the efficiencies, and it has only been optimized for EMTopo jets (although the event-level cleaning is possible for PFlow jets by using the associated EMTopo jets). The recent jet-level cleaning [104] addresses all these issues with a refined set of selection criteria.



(a) (b)

Figure 3.12: Leading jet timing distributions for EMPFlow jets from the HLT_j35 and HLT_j45 triggers, for **(a)** the BadLoose event cleaning and **(b)** the BadTight event cleaning, both for jets with $75 < p_T < 100$ GeV where the two triggers are fully efficient. Reproduced from [103].

This new cleaning also stems from a deeper understanding of the different sources of noise, either coming from physics (electrons or electroweak processes) or of non-collision origin, as can be seen in Fig. 3.13a.

Similarly to the standard event-level cleaning, a simple set of selection criteria based on noise-sensitive variables is derived, depending on the region of the detector. The noise-sensitive variables used are f_{EM}, f_{max}, f_{ch}, $\langle Q \rangle$, f_Q^{LAr}, E_{neg}, already defined in Section 3.2.2. The full set of selection criteria is given in [104], both for EMTopo and PFlow jets (that require different selection since PFlow contains tracking information: its internal substructure variables can be quite different).

The jet-level cleaning rejects the non-collision background and the physics-originating background to give a smoothly falling p_T spectrum, as can be observed in Fig. 3.13b. This figure shows the good-enriched and fake-enriched samples, before and after the jet-level cleaning: the fake events are strongly suppressed compared to the good events by three orders of magnitude in most phase space regions (five orders of magnitude in the central region $|y| < 1.5$), and the unphysical high-p_T tails are removed.

The efficiency of this selection has been studied *in situ* with a tag-and-probe method on dijet events: the tag jet (which is the sub-leading jet) is required to pass the jet-level cleaning, and the efficiency is computed as the ratio of the number of leading jets passing the selection divided by the total number of leading jets. An example of efficiency *vs.* p_T for one rapidity bin is shown in Fig. 3.13c. For the central region $|y| < 2.0$, the inefficiency is of about 0.5 to 1 %. For the $2.0 \leq |y| < 2.5$ region, the inefficiency is of 0.5 % at low p_T to 1 % at high p_T. For the $2.5 \leq |y| < 3.5$ region, the inefficiency is less than 1 % at low p_T and reaches almost 3 % at the highest p_T. For $|y| > 3.5$, the inefficiency is around 1 %.

While the standard event-level cleaning performs very well in the central region of the detector and at low and intermediate p_T, the new jet-level cleaning removes more fake jets, as can be seen in Fig. 3.13d, especially at high p_T or in the forward region (especially in the $3.0 \leq |y| < 3.5$ region).

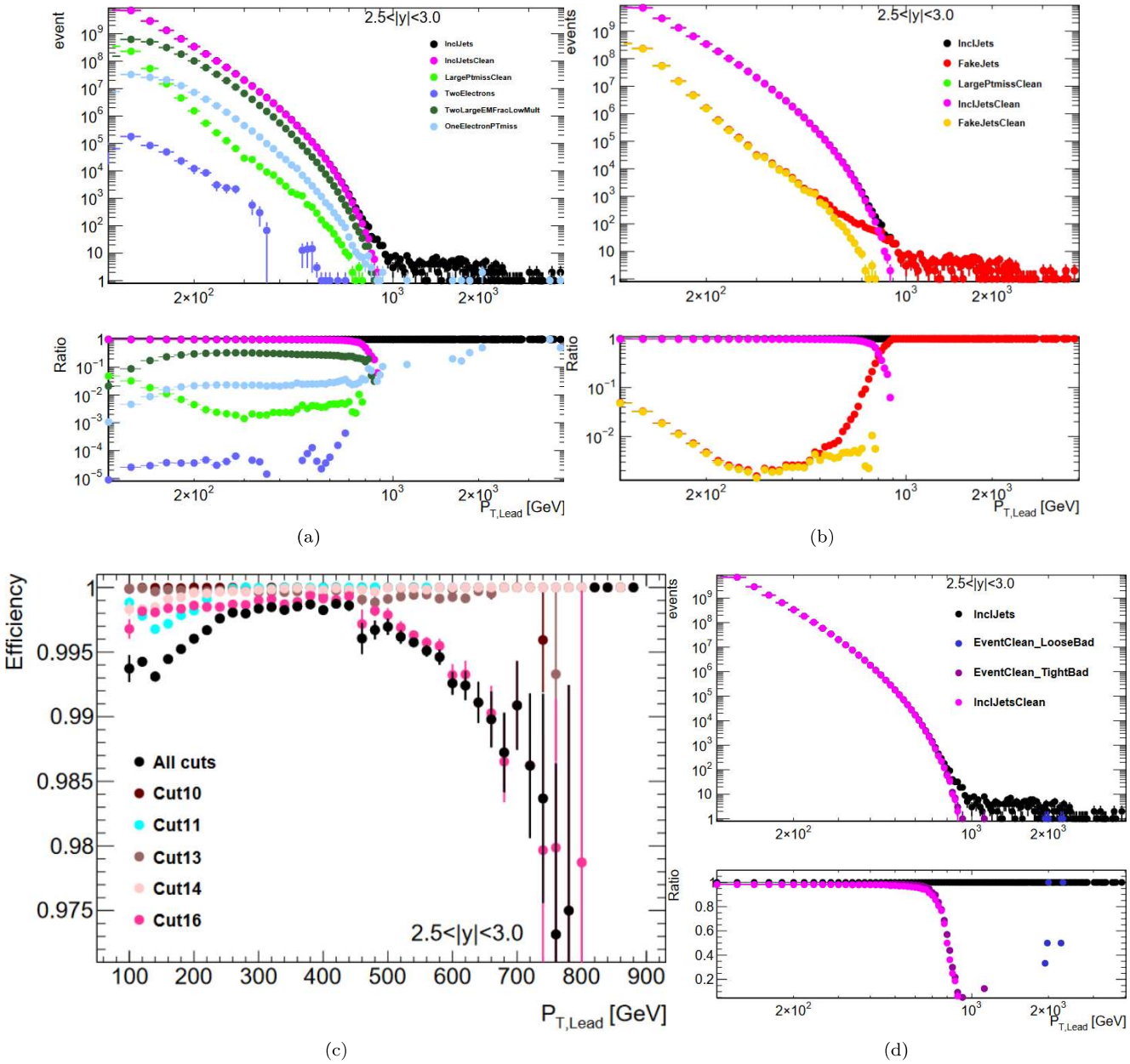


Figure 3.13: Example of one rapidity bin: $2.5 \leq |y| < 3.0$. **(a)** Number of events as a function of the leading EMTopo jet p_T, either for an inclusive event selection (black), for event classes characterizing the physics background (large p_T^{miss} in light green, two electrons in dark blue, two jets with a large fraction of their energies in the EM calorimeter and a low multiplicity of tracks in dark green, one electron and p_T^{miss} in light blue), or for an inclusive event selection after the jet-level cleaning is applied (pink). **(b)** Leading EMTopo jet spectrum with the standard BadLoose cleaning (black) and the updated jet cleaning (pink), together with the fake jet background class before (red) and after application of the new jet cleaning (yellow). **(c)** Jet cleaning efficiency as a function of the leading PFlow jet p_T for all selection criteria applied (black) or for individual selection criteria. **(d)** Inclusive EMTopo jet spectrum with the standard BadLoose (blue) and BadTight (purple) jet cleanings, and the new jet cleaning (pink). Reproduced from [104].

3.3 Jet calibration

The particles stemming from the hadronization of a quark or gluon enter the ATLAS detector and produce tracks and energy deposits, clustered together to form objects of interest for physics analyses, jets, reconstructed following the procedures previously described. However, physics observables of the jet reconstructed from detector inputs or directly from the (inaccessible) particles might differ, for a number of reasons: detector effects like the response difference between the different types of particles, operation conditions like the number of collisions per bunch crossing, accidental events like cosmic muons, and many more. This is quantifiable and measurable thanks to the multiple techniques described in this section.

A "calibration" is a comparison: it is the quantification of the differences between the measurements of probe and reference objects, like between the "reco" and the "truth" quantities, with an estimation of associated uncertainties. It is usually followed by a "correction" to bring the reco quantities to the truth ones by correcting the measured data (on average, within uncertainties). For instance, the purpose of the Jet Energy Scale is to achieve $\left\langle \frac{p_T^{\text{reco}}}{p_T^{\text{truth}}} \right\rangle = 1$. The goal of the calibration is to achieve a small bias with realistic estimations of the uncertainties.

The calibration in ATLAS is a chain of multiple techniques that aim at correcting the jet energy, η and p_T (with the Jet Energy Scale, or JES), the jet mass (with the Jet Mass Scale, or JMS), or at measuring the resolution in those quantities (with the Jet Energy Resolution or JER and the Jet Mass Resolution or JMR) [72, 89, 105–107]. The focus of this thesis being on small-R jets, only the JES and JER obtained for small-R jets will be described thoroughly.

A few definitions will be useful for the rest of this chapter and the next two chapters:

- The "response" of the detector qualifies the measured inequality for a quantity X between the reco level and the truth level, on average: $\mathcal{R} = \left\langle \frac{X^{\text{reco}}}{X^{\text{truth}}} \right\rangle$. The word "response" is reserved to the case where such an inequality is caused by detector effects only (for instance, as will be discussed at length in Chapter 4, Section 4.1 the intercalibration factor c can be affected both by detector effects and physics effects, in which case the use of the word "response" is deemed inappropriate). The inverse of a response $\mathcal{C} = \frac{1}{\mathcal{R}}$ can therefore be used as a correction factor applied to X: $X^{\text{reco calibrated}} = \mathcal{C} X^{\text{reco uncalibrated}}$.
- The "scale" of a quantity X is the correction to be applied to this quantity, as illustrated in Fig. 3.14. It can be one single value, or a function depending on multiple parameters, for instance the jet's p_T and η. The scale is typically obtained through two different processes: either by a direct measurement of $\left\langle \frac{X^{\text{reco}}}{X^{\text{truth}}} \right\rangle$ in MC events which typically capture the largest detector effects, or *in situ* (*i.e.* data-driven) by using physical equalities. There are multiple techniques, differing by the event topology used, the variables as a function of which the calibration is derived or applied, the variables over which the technique averages. Some techniques calibrate the whole phase space at once, like the MC JES, while some others use one part of phase space that is already calibrated to calibrate another part of phase space, like the multijet balance.
- The "resolution" quantifies the spread of the measurements around the mean value, and is typically measured by the width of the first Gaussian of the distribution, as illustrated in Fig. 3.14.
- The "unfolding" is a statistical deconvolution of detector effects (resolution primarily, and residual miscalibrations) from calibrated data. The JES correction is applied on a jet-by-jet basis, and aims at bringing the reco quantity to the truth scale only on average, while the spread due to the resolution cannot be correct on a jet-by-jet basis and has to be corrected on average at the "unfolding" stage.
- An "uncertainty" expresses the dispersion of the values of the measured quantity [108]. They are of two types, referred to as:
  - "Statistical uncertainties", that are based on the statistical analysis of repeated measurements. They decrease when the number of measurements increase due to the central limit theorem.
  - "Systematic uncertainties", that are obtained by non-statistical procedures, like scientific judgment or other information concerning the possible values of the quantity.

To people more familiar with metrology terminologies, the following correspondences can be used:

- The "scale" and "resolution" quantify the two parts of accuracy, respectively trueness and precision. The "scale" measures the systematic error (or the statistical bias), and the "resolution" quantifies the spread due to the statistical variability of the measurement by the detector of the jet's quantity of interest.
- The "scale correction" makes the measurement true (within uncertainties), and the "unfolding" corresponds to the statistical correction of the imprecision of the detector (analogous to the deconvolution of the point spread function in optics).
- The "statistical uncertainties" are the type A uncertainties, and the "systematic uncertainties" are the type B uncertainties. There are uncertainties on the knowledge of both the "scale" and the "resolution".
- The "precision" qualifies all the uncertainties; it is therefore not the metrologist's precision.

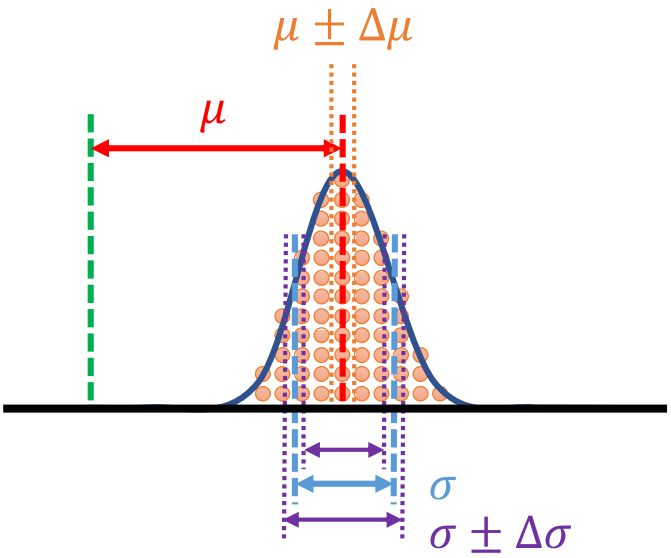


Figure 3.14: Illustration of the scale and the resolution for an observable like the jet energy. The true value is represented by the green bar, and the circles represent experimental measurements. The scale μ corrects for the systematic bias, such that the mean of the corrected measurements is equal to the true value within uncertainties $\Delta\mu$. The spread of the measurements around the mean is due to the resolution, represented by the blue curve, that is characterized by σ and associated uncertainties $\Delta\sigma$.

3.3.1 The calibration chain of small-R jets

The calibration of the Jet Energy Scale (JES) for small-R jets proceeds in multiple steps [72], shown in the flow chart of Fig. 3.15:

1. After the jet reconstruction, two calibrations dedicated to reducing the effect of pile-up are applied: a correction based on the pile-up density and the jet area, and an MC-based residual correction that removes the residual dependency of the JES to pile-up-sensitive variables.
2. The absolute MC-based calibration, that measures the response of jets in MC and corrects their energy and direction in η.
3. The MC-based Global Sequential Calibration (GSC), that reduces the dependency of the JES on visible features of the jets, like the quark/gluon nature. It improves the Jet Energy Resolution (JER), without affecting the JES.
4. The *in situ* (*i.e.* data-driven) calibrations, that aim at measuring and correcting for the residual differences between data and MC, using physics equalities in multiple event topologies. Indeed, the previous calibration steps are not perfect, and MC does not perfectly represent data (theory is incomplete, the knowledge of the detector is imperfect, the simulations are truncated to a certain degree of precision, *etc.*) These *in situ* calibrations are only applied onto data. The secondary task of these calibrations is to measure data-driven uncertainties on the JES.

Recent improvements of the calibration [89] have been used in the studies presented in this thesis, and will therefore be presented in the following section.

Some calibrations, not included in the flow chart of Fig. 3.15, are dedicated to specific phase space regions or specific uses:

- The E/p calibration, that is dedicated to the calibration of the very high energy phase space by comparing the response of single pions in the tracker and calorimeter. This calibration uses data and is therefore also *in situ*, strictly speaking, though "*in situ* " is usually used in ATLAS to refer to the tag-and-probe methods.
- The "flavor uncertainties" are a calibration without correction, they are a measurement of the modeling differences between multiple MC generators: the different models have different fragmentation patterns and hadron contents, ultimately resulting in different detector responses.

More calibrations are used in ATLAS for small-R jets, but I have not used them in my analyses, and are therefore not described in this thesis; most notably, I have not used the new b-JES calibration that is based on an *in situ* γ+jet balance (due to the new flavor uncertainties already taking the possible miscalibration of b jets into account) [89].

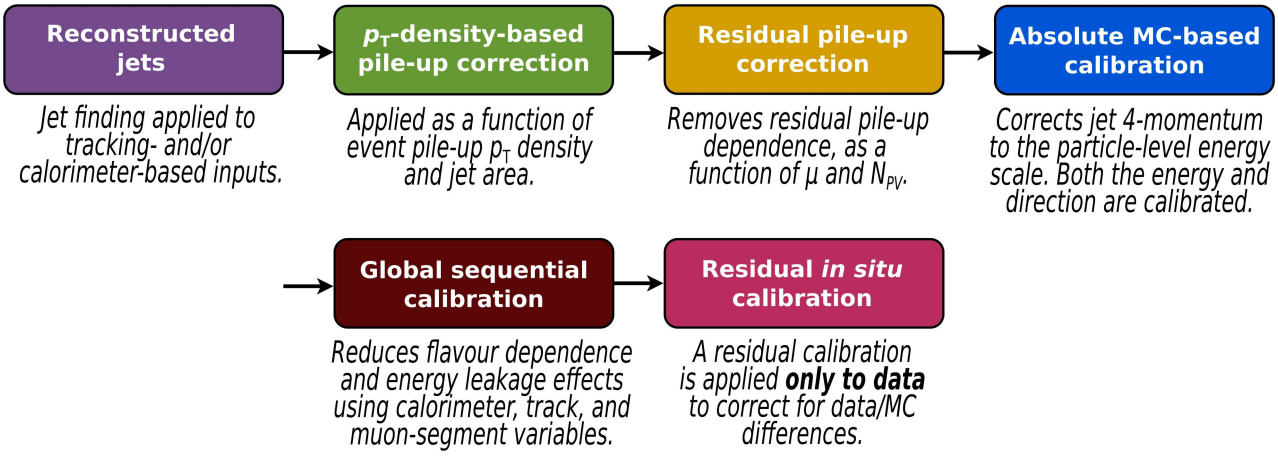


Figure 3.15: Stages of Jet Energy Scale (JES) calibration. Each one is applied to the four-momentum of the jet. Reproduced from [72].

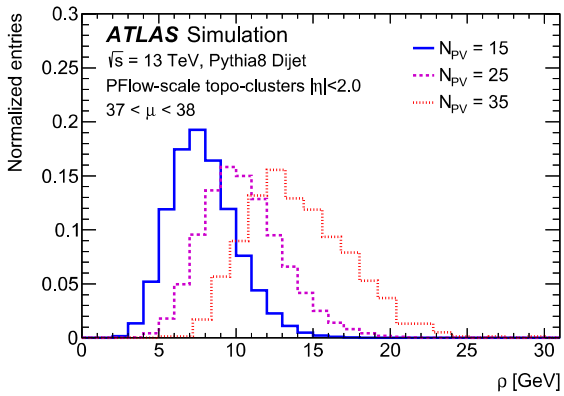


Figure 3.16: Median p_T density $\langle\rho\rangle$ for representative values of the number of primary vertices N_{PV}: 15, 25, 35. Reproduced from [72].

3.3.1.1 Pile-up corrections

The p_T-density-based pile-up correction

As a first approximation, pile-up is a diffuse and rather homogeneous contribution to all jets, due to independent energy deposits in the calorimeter. This first calibration estimates the pile-up p_T density $\langle\rho\rangle$ and corrects the p_T of the jets based on their area A:

$$p_T^{\text{PU density corrected}} = p_T^{\text{reco}} - \langle\rho\rangle A \tag{3.5}$$

The jet area A is a measure of the susceptibility of the jet to pile-up, and is calculated by "ghost-association": prior to jet reconstruction, the (η, ϕ) plane is uniformly covered with infinitely soft particles called "ghosts" which are clustered into the jets by the jet algorithm. The jet area is then the number of ghosts clustered in the jet.

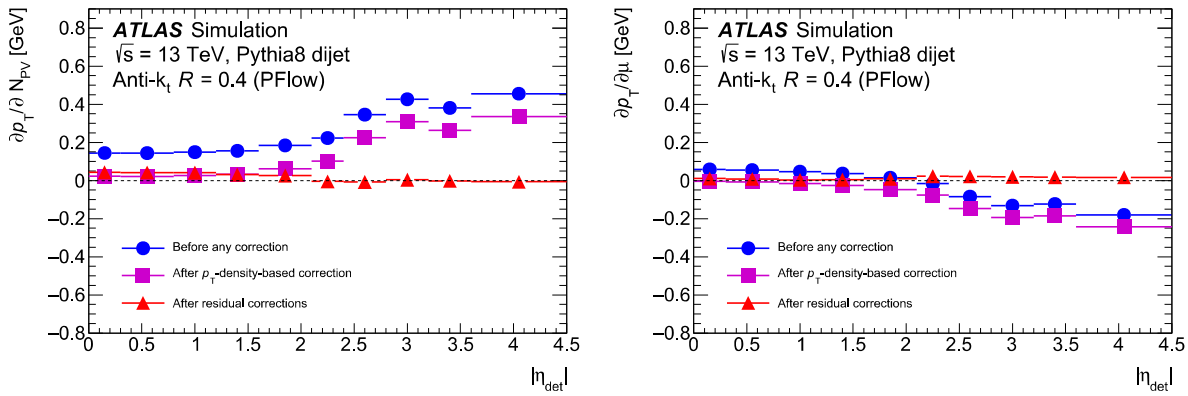
The pile-up density is estimated as the median over all the jets of their p_T divided by their area:

$$\langle\rho\rangle = \underset{i\in\{\text{jets}\}}{\text{median}} \left(\frac{p_T^i}{A_i}\right) \tag{3.6}$$

The median is used to reduce the impact of the hard-scatter jets that populate the high-p_T tails of the distribution. The jets used to estimate the pile-up density are k_t jets (see Chapter 1, Section 1.3.2.2) reconstructed from positive-energy topo-clusters with $|\eta| < 2.0$; the k_t algorithm is chosen due to its tendency to naturally reconstruct jets including an uniform soft background, and only the central region is used due to the lower occupancy in the forward region. The pile-up density is shown in Fig. 3.16 for representative values of the number of primary vertices N_{PV}.

Improvements of the p_T-density-based pile-up correction

To reduce the dependency of the p_T-density-based pile-up correction to the hard-scatter jets, the "side-band" can



(a) (b)

Figure 3.17: Dependency of the jet p_T to **(a)** in-time pile-up (N_{PV} averaged over μ) and **(b)** out-of-time pile-up (μ averaged over N_{PV}) for $p_T^{\text{truth}} = 25$ GeV jets without the p_T-density-based correction (p_T^{reco}, blue), with the p_T-density-based correction ($p_T^{\text{PU density corrected}}$, purple) and with both the p_T-density-based correction and the residual pile-up correction ($p_T^{\text{corrected}}$, red). Uncertainties are taken from the fit results and are too small to be visible. Reproduced from [72].

be used. Instead of reconstructing jets with the normal inputs, *i.e.* charged Particle Flow Objects (PFOs) with $|z_0 \sin(\theta)| < 2$ mm and neutral PFOs, the jets are reconstructed with inputs less likely to come from the hard-scatter vertex, *i.e.* charged PFOs with 2 mm $< |z_0 \sin(\theta)| < 4$ mm.

This results in a reduction in the dependency of the correction on the hard-scattering topologies, thus improving the topology uncertainty (described hereafter) by a factor of 3.

Residual pile-up correction

A pile-up dependency is still observed after the p_T-density-based pile-up correction. Indeed, the pile-up p_T density is computed in the central region and does not fully describe the pile-up sensitivity in the forward calorimeter region or in the higher-occupancy core of high-p_T jets.

The residual dependency is observed to vary linearly with the number of primary vertices in the event N_{PV} and the mean number of primary vertices μ (which are respectively sensitive to in-time pile-up and out-of-time pile-up) when geometrically matching reco jets and truth jets within $\Delta R = 0.3$. The residual dependency on in-time pile-up and out-of-time pile-up is shown respectively in Fig. 3.17a and in Fig. 3.17b, in the purple histograms. Therefore, the p_T of the jet is corrected as follows:

$$p_T^{\text{corrected}} = p_T^{\text{PU density corrected}} - \alpha(\eta)\,(N_{PV} - 1) - \beta(\eta)\mu \tag{3.7}$$

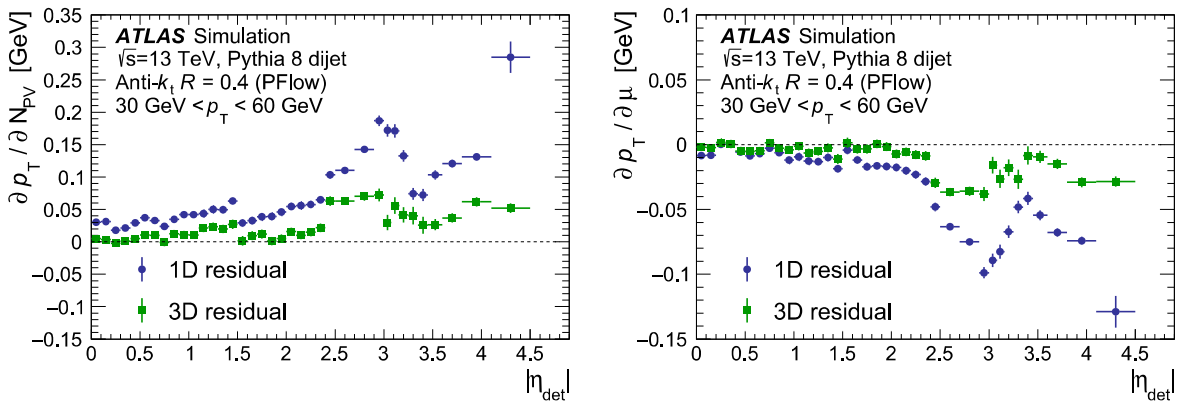
where α and β are the slopes of the $p_T^{\text{PU density corrected}}$ *vs.* N_{PV} and $p_T^{\text{PU density corrected}}$ *vs.* μ distributions respectively, binned in η: $\alpha(\eta) = \frac{\partial p_T^{\text{PU density corrected}}}{\partial N_{PV}}(\eta)$ and $\beta(\eta) = \frac{\partial p_T^{\text{PU density corrected}}}{\partial \mu}(\eta)$. Bot the α and β coefficients are observed to have a logarithmic dependency with respect to p_T^{truth}, and are therefore logarithmically fitted in each η bin in the range $20 < p_T^{\text{truth}} < 200$ GeV. In each η bin, the nominal values of α and β are those of the fit at 25 GeV.

Improvements of the residual pile-up correction

The residual pile-up correction has recently been improved in [89], by taking into account the correlations between more variables: N_{PV}, μ, and jet properties p_T and η. This correction, called the "3D residual correction" (as opposed to the "1D residual correction" described previously), corrects the jet's p_T as follows:

$$p_T^{\text{corrected}} = p_T^{\text{PU density corrected}} - \Delta p_T(N_{\text{PV}}, \mu, \eta, p_T^{\text{PU density corrected}}) \tag{3.8}$$

This improved residual pile-up correction results in a reduction of the pile-up dependency above 30 GeV, as shown in Fig. 3.18. It also simultaneously corrects for pile-up effects and detector effects, thus resulting in smaller corrections factors in the MC JES calibration step. Finally, it allows to remove the "p_T dependency" uncertainty described hereafter.



(a) (b)

Figure 3.18: Dependency of p_T on **(a)** N_{PV} and **(b)** μ after the 1D or the 3D residual pile-up corrections. Reproduced from [89].

Uncertainties

There are four systematic uncertainties associated to this correction, three of which are measured *in situ*:

- The "p_T term" uncertainty is obtained from the logarithmic fit over the p_T range.
- The "offset N_{PV}" and "offset μ" uncertainties estimate the difference of α and β between MC and data thanks to two *in situ* methods. First, the jets reconstructed solely from tracks provide a pile-up-independent estimation of p_T in the tracker-acceptance region $|\eta| < 2.1$. Second the jets recoiling against Z bosons (decaying to two leptons) are used in the $2.1 < |\eta| < 4.5$ region where the p_T of the Z boson is used as a proxy for the p_T of the jet. In both cases, the measurement of α and β is performed onto the pile-up-independent sets of jets, and the biggest discrepancy between MC and data is taken as the uncertainties.
- The "ρ topology" uncertainty accounts for the dependency on the topology of the hard-scattering event in the determination of $\langle\rho\rangle$. Two very different topologies are used, dijet events and $Z(\rightarrow \mu\mu)$+jets events, in which the pile-up p_T-density is estimated, both in data and in MC (for multiple MC generators). The uncertainty is expressed as a function of μ as:

$$\Delta\langle\rho\rangle(\mu) = \left(\langle\rho\rangle_{\text{MC}}^{\text{dijet}} - \langle\rho\rangle_{\text{MC}}^{Z+\text{jet}}\right) - \left(\langle\rho\rangle_{\text{data}}^{\text{dijet}} - \langle\rho\rangle_{\text{data}}^{Z+\text{jet}}\right) \tag{3.9}$$

  and the uncertainty that is kept and propagated as a JES uncertainty is $\Delta\langle\rho\rangle(\langle\mu\rangle)$, with $\langle\mu\rangle = 33.7$ the average number of interactions per bunch crossing over the full run 2.

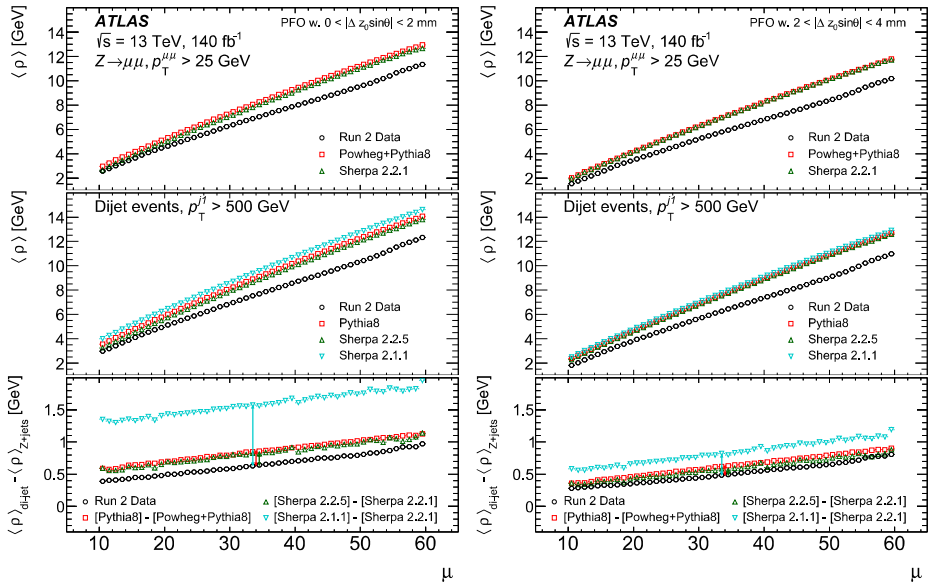
The ρ topology uncertainty has been improved by a factor of 7 thanks to two changes: the side-band technique described previously, and the use of SHERPA MC generators that have the same Multi-Parton Interaction (MPI) models for the dijet and Z+jet samples. This improvement can be seen in Fig. 3.19.

3.3.1.2 Absolute MC-based calibration

After the pile-up corrections, the energy and η direction of the jets are corrected to the "absolute" scale (*i.e.* the truth scale) thanks to an MC-based calibration. This correction is necessary to account for multiple effects: the energy scale is affected by the non-compensating calorimeter response, the energy losses in passive materials and by out-of-cone effects; the η reconstruction is affected by the transition between different calorimeter technologies and by sudden changes in calorimeter granularity.

The truth and reco jets are geometrically matched within $\Delta R = 0.3$, and the response $\mathcal{R}$ is computed as the mean of the $\frac{E^{\text{reco}}}{E^{\text{truth}}}$ distribution: it is extracted by a Gaussian fit on the core of the distribution, in bins of E^{truth} and η_{det}. The response is computed in bins of E^{truth} instead of bins of E^{reco}, to avoid biasing the measurement and because it is a more Gaussian; it therefore has to be re-parameterized in bins of E^{reco}. This is performed through a "numerical inversion": in each $\left(E^{\text{truth}}, \eta_{\text{det}}\right)$ bin, the mean of the E^{reco} distribution is also determined ($\langle E^{\text{reco}}\rangle$), then a parametric fit of the $\left(\langle E^{\text{reco}}\rangle_j, \mathcal{R}_j\right)_{j\in\{E^{\text{truth}}\text{ bins}\}}$ points is performed in each η_{det} bin. The response function used is a polynomial of $\log(E^{\text{reco}})$ up to the order $N_{\text{max}} = 8$. The calibration factor is then the inverse of the response function, and is shown *vs.* η_{det} in Fig. 3.20a.

Recent developments (in [89]) have shown that a penalized spline is a better fit function than the polynomial. A spline of degree N is a piecewise polynomial function of degree N where pieces of the spline meet at points called



(a) (b)

Figure 3.19: Distribution of $\langle\rho\rangle$ as a function of μ in the $Z(\rightarrow \mu\mu)$+jet sample (upper panels) and in the dijet sample (middle panels), for data and different MC generators. The difference $\langle\rho\rangle^{\text{dijet}} - \langle\rho\rangle^{Z+\text{jet}}$ is shown in the lower panel either for data or for different choices of MC generators. The ρ topology uncertainty that is propagated to analyses is taken at $\langle\mu\rangle = 33.7$ and is indicated by the vertical lines in the lower panels. **(a)** The previous determination of $\langle\rho\rangle$ is used with charged PFOs being in $|z_0 \sin(\theta)| < 2$ mm, whereas in **(b)** the side-band selection is used with charged PFOs in 2 mm $< |z_0 \sin(\theta)| <$ 4 mm. The gain using the side-band is a factor of 3. The gain using the new SHERPA generator for the dijet events (SHERPA 2.2.5) is a factor of 4. The total gain in the ρ topology uncertainty is a factor of 7. Reproduced from [89].

knots and where the first $N-1$ derivatives are continuous across the knots. The penalization is an extra LAGRANGE term in the least-square minimization that requires the function to be smooth, to avoid over-fitting the points. This method has shown to provide a better closure than the polynomial fit at low energies. The MC JES calibration obtained with the penalized spline is shown in Fig. 3.20c.

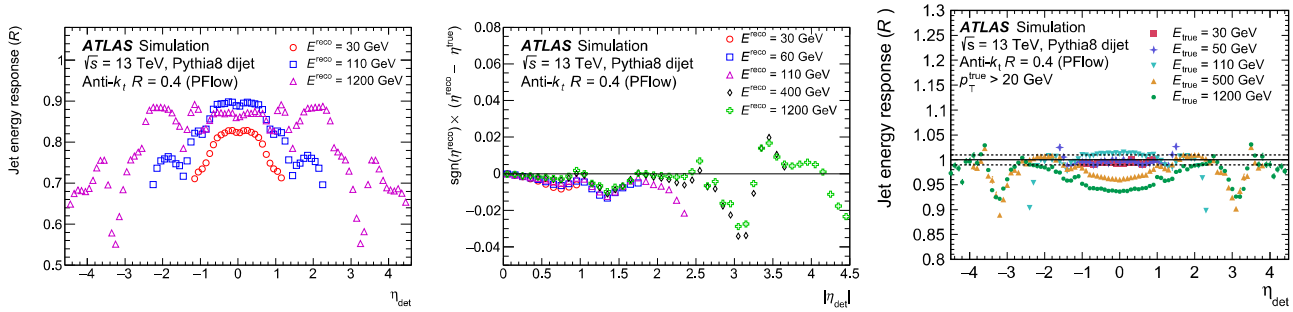
The η direction correction is the average signed difference between the η of the reco jet and the η of the truth jet, where the mean is extracted by a Gaussian fit of the core of the $\eta^{\text{reco}} - \eta^{\text{truth}}$ distribution. It is shown in Fig. 3.20b.

3.3.1.3 Global calibration

The energy response of the calorimeter has been corrected on average, but on a jet-by-jet basis it is still sensitive to the jet particle composition and to the distribution of energy within the jet. Most notably, the type of parton that initiated the jet (a quark or a gluon) has an important impact on the shower shape. A quark-initiated jet often includes hadrons with a higher fraction of the jet p_T that penetrate further into the calorimeter, and the energy is more concentrated in the center of the jet. A gluon-initiated jet has a higher multiplicity of hadrons (about double that of quark-initiated jets), the hadrons carry less energy and are emitted at wider angles from the jet center, thus resulting in a lower calorimeter response and a wider jet profile.

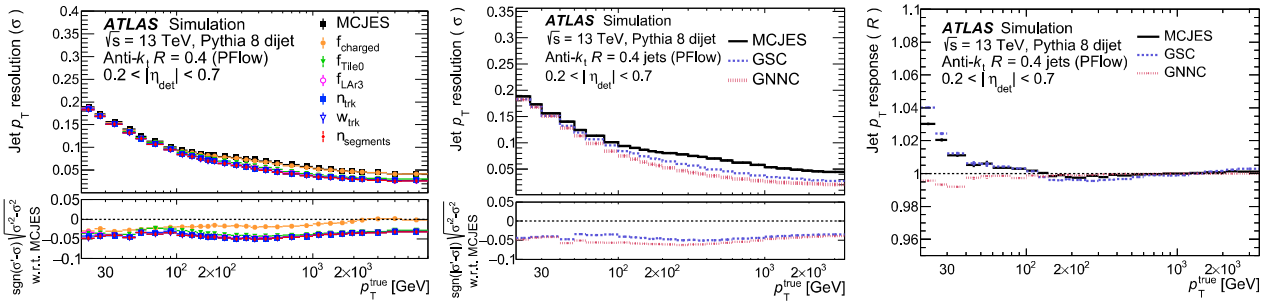
This makes the jet energy sensitive to global jet observables, and a correction can be derived to reduce the sensitivity of the energy scale to each of these variables, successively, to reduce the fluctuations and thus improve the jet energy resolution without affecting the jet energy scale. The six global observables retained for the Global Sequential Calibration (GSC) are:

- f_{ch}: the fraction of the jet p_T carried by charged particles, measured using ghost-associated tracks with $p_T > 500$ MeV ($\eta_{\text{det}} < 2.5$);
- f_{Tile0}: the fraction of jet energy measured in the first layer of the hadronic tile calorimeter ($\eta_{\text{det}} < 1.8$);
- f_{LAr3}: the fraction of jet energy measured in the third layer of the electromagnetic LAr calorimeter ($\eta_{\text{det}} < 3.5$);
- N_{track}: the number of tracks with $p_T > 1$ GeV ghost-associated with the jet ($\eta_{\text{det}} < 2.5$);
- w_{track}: the "track width", *i.e.* the average p_T-weighted transverse distance in the (η, ϕ) plane, between the jet axis and all tracks of $p_T > 1$ GeV ghost-associated with the jet ($\eta_{\text{det}} < 2.5$);
- N_{segments}: the number of muon track segments ghost-associated with the jet ($\eta_{\text{det}} < 2.8$).



(a) (b) (c)

Figure 3.20: **(a,c)** Average energy response as a function of η_{det} and **(b)** signed difference between the reconstructed and truth jet η. **(a)** is obtained with the polynomial fit, whereas **(c)** is obtained with the penalised splines. Note that the two responses cannot be directly compared, as the pile-up corrections applied onto these sets are different, which explains the important scale difference. Reproduced from [72, 89].



(a) (b) (c)

Figure 3.21: **(a)** The jet p_T resolution after the MC JES and after each correction of the GSC. **(b)** The jet p_T resolution for $0.2 < |\eta_{\text{det}}| < 0.7$ after the MC JES only and after the MC JES followed either by the GSC or the GNNC. **(c)** The jet p_T closure for $0.2 < |\eta_{\text{det}}| < 0.7$ for the MC JES only and for the MC JES followed either by the GSC or the GNNC. Reproduced from [89].

f_{Tile0} and f_{LAr3} give information on the penetration of the jet. N_{track} and w_{track} give information on the particle multiplicity of the jet and their transverse profile. N_{segments} is used as input for the "punch-through" correction, that aims at correcting the response of high-p_T jets that are not fully contained in the calorimeter, by using the punch-through signal in the muon spectrometer.

The procedure to obtain each of the six GSC corrections is similar to the MC JES one: the p_T response is fit in bins of $|\eta|$, p_T^{truth} and the GSC observable, normalized to avoid changing the already calibrated p_T scale, "inverted" to move from a p_T^{truth} binning to a p_T^{reco} binning, and smoothed. The improvement in the jet energy resolution after each of these corrections is shown in Fig. 3.21a.

Improvements of the GSC: the Global Neural Network Calibration (GNNC)

The GSC sequentially corrects for each variable, so does not exploit correlations between the variables. The new Global Neural Network Calibration (GNNC) uses a Deep Neural Network (DNN), trained to predict the p_T response. It can use more variables, since it can take the correlations into account. The technical details of the DNN architecture are provided in [89].

The variables taken into account are:

- Calorimeter-based variables: the fraction of energy measured in each of the four layers of the electromagnetic LAr calorimeter, in each of the 3 layers of the hadronic tile calorimeter, in each of the 4 layers of the hadronic end-cap calorimeter, in each of the 3 layers of the forward calorimeter, and the minimum number of clusters containing 90 % of the jet energy;
- The jet kinematics: the jet p_T after the MC JES calibration and the detector η;
- Tracking-based variables already taken into account in the GSC: w_{track}, N_{track} and f_{ch};
- The number of muon segments ghost-associated with the jet, N_{segments}, already taken into account in the GSC;

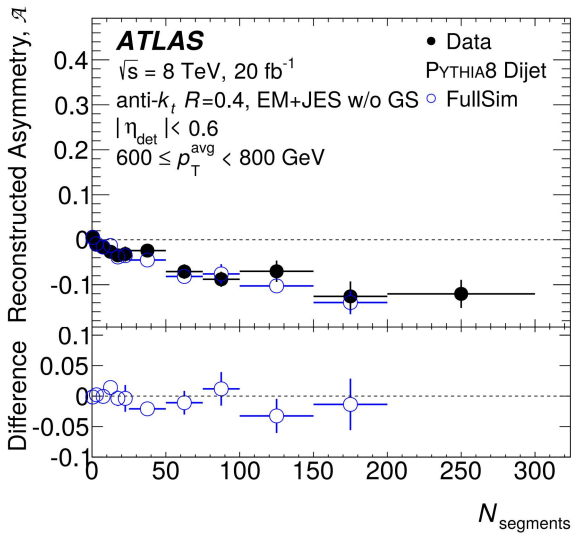


Figure 3.22: *In situ* validation of the modeling of the N_{segments} variable. The asymmetry is converted to a response, and the data/MC ratio is taken as an uncertainty on the JES: the "punch-through" uncertainty. Reproduced from [107].

- Pile-up sensitive variables: the average number of interactions per bunch crossing μ and the number of reconstructed primary vertices N_{PV}.

It reduces even further the differences between jets with different fragmentation (*i.e.* quark-initiated jets *vs.* gluon-initiated jets) compared to the GSC. The JER is improved by 15 to 25 % on average compared to the GSC as shown in Fig. 3.21b. The closure of the calibration is greatly improved compared to the GSC, as shown in Fig. 3.21c.

***In situ* validation of the GSC and the punch-through uncertainty** The global calibration is validated *in situ* with dijet events using a tag-and-probe technique, similarly to the η-intercalibration described hereafter in Section 3.3.2.1 except that both jets of the dijet events are required to be in the same $|\eta|$ region. The choice of the reference jet and probe jet is arbitrary and the events are used twice, alternating the roles of reference and probe: no global calibration is applied to the probe jet, and the asymmetry $\mathcal{A} = \frac{p_T^{\text{probe}} - p_T^{\text{ref}}}{p_T^{\text{avg}}}$ between the two jets is measured in bins of the variables used in the global calibration. Good agreement is found between data and MC for all variables.

Only one effect does not receive any *in situ* calibration in the later stages of the calibration, it is the calibration with respect to N_{segments}, *i.e.* to the number of muon segments ghost-associated with the jet, a correction known as the "punch-through" correction, referencing the fact that jets pass through the calorimeter into the muon system. Therefore, this validation measurement provides an uncertainty on the JES, by taking the difference in jet response between data and MC in bins of N_{segments}.

3.3.1.4 Flavor uncertainties

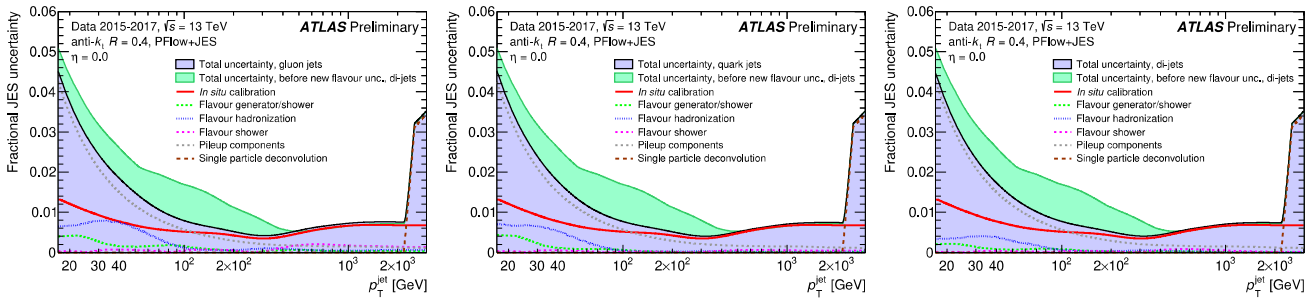
Despite the application of the GSC or GNNC, a residual dependence of the calorimeter response if found on the fragmentation pattern and hadron content, and therefore on the jet flavor as well as on the showering and hadronization models used by the different MC generators.

Two flavor dependence uncertainties are derived from MC simulations to account for it [106]. Jets are labelled by partonic flavor: they are matched to the highest-energy parton found inside the jet, using the leading order label provided by the MC generator. The response of quark-initiated jets $\mathcal{R}_q$ and gluon-initiated jets $\mathcal{R}_g$ is obtained both in PYTHIA8 and in HERWIG7. The fraction of gluon-initiated jets f_g is obtained for each analysis, as well as the associated uncertainty $\sigma(f_g)$.

The flavor composition uncertainty accounts for the fact that the jet response is different for quark-initiated jets and gluon-initiated jets, and for the fact that quark/gluon composition of the sample is not perfectly well known. It is defined as:

$$\sigma_{\text{composition}} = \sigma(f_g)\frac{\mathcal{R}_q - \mathcal{R}_g}{f_g\mathcal{R}_g + (1-f_g)\mathcal{R}_q} \tag{3.10}$$

The flavor response uncertainty accounts for the fact that, unlike the quark-initiated jet response, the gluon-initiated jet response is found to differ significantly between generators. This difference mostly comes from the



(a) (b) (c)

Figure 3.23: Comparison of the new individual flavor uncertainties and of the total JES uncertainties for **(a)** gluon jets; **(b)** quark jets; **(c)** a dijet sample including gluon and quark jets. The light green line shows the generator uncertainty, the pink line shows the shower uncertainty and the blue line shows the hadronization uncertainty. The green band shows the total uncertainties obtained with the "standard" approach, while the purple band shows the total uncertainties obtained with the new flavor uncertainties. Reproduced from [110].

dependence of the response on the baryon and kaon energy fractions and from the difference in baryon and kaon energy fractions between the Lund-string-based and cluster-based hadronization models [109]. This uncertainty is defined as:

$$\sigma_{\text{response}} = f_g \left(\mathcal{R}_{g,\text{Pythia8}} - \mathcal{R}_{q,\text{Herwig}}\right) \tag{3.11}$$

Improvements of the flavor uncertainties

The response difference between the generators has been split into three components, following the factorization theory [110]:

- The "generator/shower" uncertainty evaluates the difference in response caused by a different shower model as well as generator differences (such as Multi-Parton Interaction (MPI) and the Underlying Event (UE)), by comparing Pythia8 and Sherpa, both using the Lund string hadronization model (see Chapter 1, Section 1.3.1 for explanation on the MC generators and the different models used);
- The "shower" uncertainty evaluates the difference in response caused by a different shower model, by comparing the angular-ordered and the dipole shower models of Herwig7;
- The "hadronization" uncertainty evaluates the difference in response caused by a different hadronization model, by comparing the cluster-based hadronization and the Lund string hadronization of Sherpa. The hadronization models that are disfavoured by ATLAS and LEP data are not taken into account [110].

This uncertainty is now computed on a jet-by-jet basis using the `PartonTruthLabelID` provided by the MC generators, that has proven to be useful in this particular situation. This allows to avoid applying a mix of gluon and quark uncertainties onto each jet, but to directly apply the correct uncertainty on each jet. It also saves the analyzers from having to determine the quark-gluon mix of their analyses.

Both these changes lead to an overall improvement that can be observed in Fig. 3.23, where the difference between the purple and the green bands shows the overall improvement. Figure 3.23a and Fig. 3.23b show the flavor uncertainties to be applied respectively to gluon and quark jets based on the MC-provided `PartonTruthLabelID`, and Fig. 3.23c shows an example of a sample where quark and gluon jets are mixed.

Now that the flavor uncertainties are smaller than the response difference between generators, it is essential to calibrate all generators with respect to Pythia8, *i.e.* to bring the JES of those generators to the one of Pythia8, thanks to "MC-to-MC" calibrations. Indeed, Pythia8 is used for almost all the *in situ* calibrations. This prevents the response of a generator from lying outside the flavor uncertainty band centered on Pythia8. This also helps in reducing the "modeling" uncertainties in physics analyses, that is usually evaluated as a difference between multiple MC generators, as only the physics differences will be taken into account and not the difference in detector response; this ensures that the detector effects are not counted twice (once in the flavor uncertainties and once in the modeling uncertainty of the physics analysis).

3.3.1.5 E/p calibration

Historically, the first calibration of the jet energy scale in ATLAS has been preformed thanks to the so-called E/p measurements, and using test-beam measurements [111, 112]. Nowadays, these techniques are only used in phase space region where the *in situ* measurements (described hereafter) do not provide a calibration, *i.e.* in the high energy region above 2 TeV where the multi-jet balance (MJB) does not have enough data due to the low cross-section.[3]

The E/p measurements rely on single charged particles producing isolated tracks, for which both the transverse momentum p_T and the energy E can be reconstructed, respectively thanks to the track in the tracking system and to the energy deposits in the calorimeters (see Chapter 2). The selection criteria ensure that the tracks are isolated and geometrically matched to the topo-cluster(s) with $\Delta R < 0.2$. The measurements focused on special LHC runs with low instantaneous luminosity and identified isolated tracks from minimum-bias events; this allowed to probe the low-momentum phase-space with $p < 30$ GeV, as pure QCD interactions rather produce jets than high-momentum isolated tracks. The calorimeter response to hadrons with energies up to 350 GeV was probed thanks to test-beam measurements using the Super Proton Synchrotron (SPS). For energies above 400 GeV, no direct measurement in a test beam or *in situ* exist, so an additional conservative uncertainty of 10 % is added in quadrature to that of the 350 GeV measurement uncertainty, to cover possible effects.

To obtain the mean calorimeter response to single charged particles in (p_T, η) bins, the mean of the E/p distributions (see Fig. 3.24a) is extracted, from which background noise coming from neutral particles (mostly $\pi^0 \to \gamma\gamma$) has to be subtracted. The background noise is estimated by selecting the energy deposits in the electromagnetic calorimeter in a ring around the charged hadron ($0.1 \leq \Delta R < 0.2$) and by taking the mean of the E/p of the background energy depositions $\langle E/p \rangle_{\mathrm{BG}} = \left\langle \frac{E^{0.2}_{\mathrm{EM}} - E^{0.1}_{\mathrm{EM}}}{p} \right\rangle$. The background-corrected mean E/p measurement is then:

$$\langle E/p \rangle = \langle E/p \rangle_{\mathrm{raw}} - \frac{4}{3} \langle E/p \rangle_{\mathrm{BG}} \tag{3.12}$$

The corrected calorimeter response for single particles $\langle E/p \rangle$ is shown in Fig. 3.24b.

To propagate the calorimeter response and associated uncertainties from isolated tracks to jets, two informations are required: the calorimeter response to identified hadrons, and measurements of the jet substructure. Indeed, a jet includes a variety of hadrons that may differ from the inclusive sample of isolated hadrons, and the hadrons in a jet are not isolated. Single hadrons are identified using decays of K_S (for $\pi^{\pm}$), Λ (for protons) and $\bar{\Lambda}$ (for $\bar{p}$), allowing to separate the effects of hadronic interactions from annihilations. The pions from the K_S decay can be more or less boosted, thus allowing to study the impact of the collimation of a pair of pions (a single pion might not be reconstructed due to thresholds, and be reconstructed when in combination with another pion due to the collimation). The jet substructure is estimated thanks to MC simulations, the important quantity being the fraction of the jet energy carried by the different particles, in the different momentum ranges; it is shown in Fig. 3.24c.

Additional uncertainties are added to take into account a broad range of effects that might affect the propagation from the single particle energy response to the jet energy scale, listed in [112]. The mean response and the associated uncertainties are shown in Fig. 3.24d.

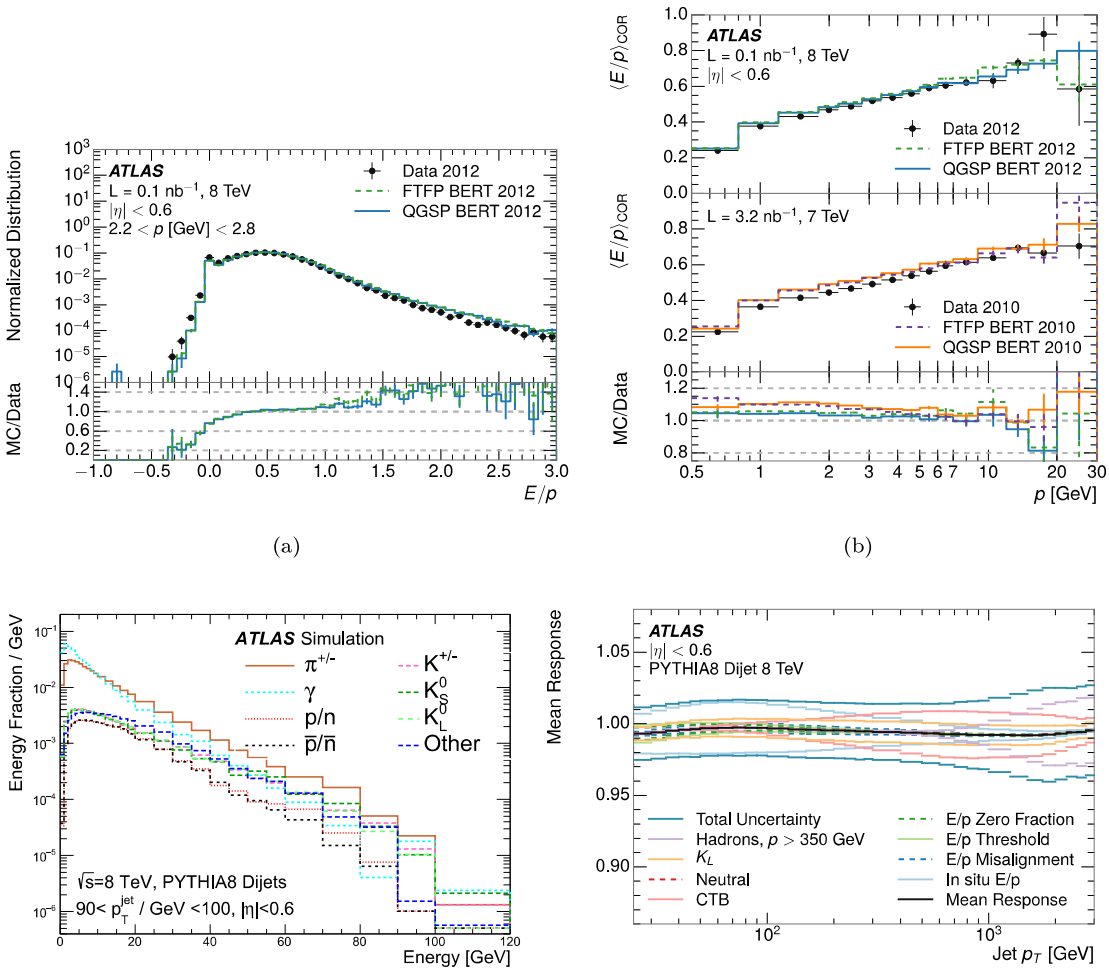
The uncertainty derived through the E/p study is of 2 to 5 % for jets with $|\eta| < 0.6$ and $20 < p_T < 2500$ GeV. The uncertainties are dominated by the non-uniformity of the calorimeter, imperfect knowledge of the effect of the material in front of it, and, for positive beams, the fraction of protons.

Improvements of the E/p calibration

The $W^{\pm} \to \tau^{\pm}(\to \pi^{\pm}\nu_{\tau})\nu_{\tau}$ events provide highly-energetic isolated charged pions over a broad energy range [113]. A complete understanding of the background processes, along with good modelization of those background processes, allow for a very precise measurement of the mean E/p in this dataset. The mean of the E/p distribution is obtained by a Gaussian fit in (p_T, η) bins, also taking the background processes into account with an additional LANDAU distribution, as shown in Fig. 3.25a.

For isolated charged pions, the calorimeter response is observed to be overestimated by 2 % in the central region and underestimated by 4 % in the endcaps in the ATLAS simulation, as shown in Fig. 3.25b. The precision achieved for the response of the calorimeter to isolated charged pions is of ≤ 1 % over the broad phase space delimited by $|\eta| < 2.4$ and $10 < p_T < 3000$ GeV. This results in a dramatic improvement of the high-p_T jet uncertainties compared to the combined test-beam data described above.

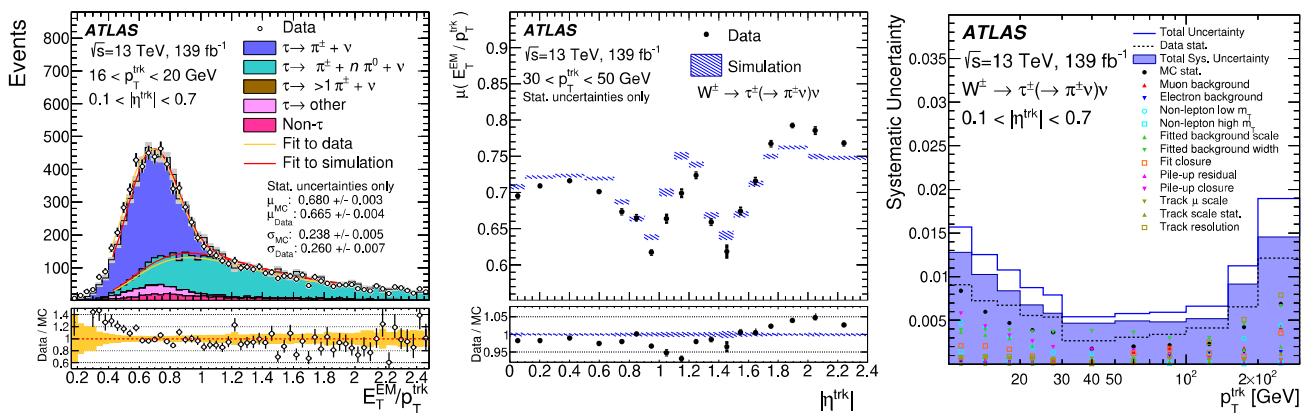
[3] The E/p techniques being data-driven, they are *in situ* in the strict sense of the word; however in ATLAS, "*in situ* " is usually reserved to the tag-and-probe techniques described in Section 3.3.2.



(a) (b)

(c) (d)

Figure 3.24: **(a)** The E/p distribution for isolated tracks with $|\eta| < 0.6$ and $1.2 < p < 1.8$ GeV. **(b)** $\langle E/p \rangle$ for tracks with $|\eta| < 0.6$. **(c)** The fractional contribution to the total jet energy of particles in a certain range of momenta, for particles inside anti-k_t, $R = 0.4$ jets with $90 < p_T < 100$ GeV. **(d)** The jet energy scale uncertainty contributions, as well as the total jet energy scale uncertainty, as a function of jet p_T for $|\eta| < 0.6$. Reproduced from [112].



(a) (b) (c)

Figure 3.25: **(a)** The E/p distribution for isolated tracks with $0.1 < |\eta| < 0.7$ and $16 < p_T < 20$ GeV, showing the different background processes. **(b)** $\langle E/p \rangle$ for tracks with $30 < p_T < 50$ GeV. **(c)** The various systematic uncertainty for $0.1 < |\eta| < 0.7$. Reproduced from [112].

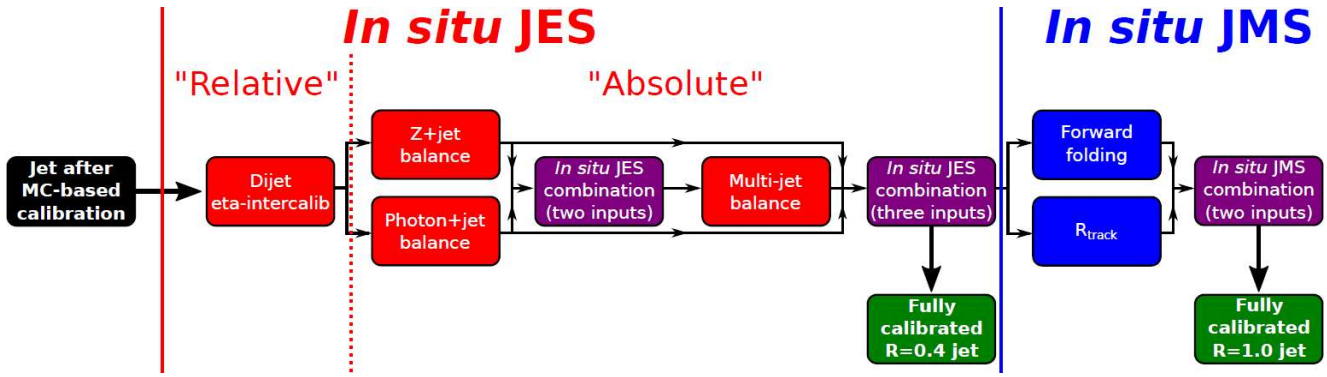
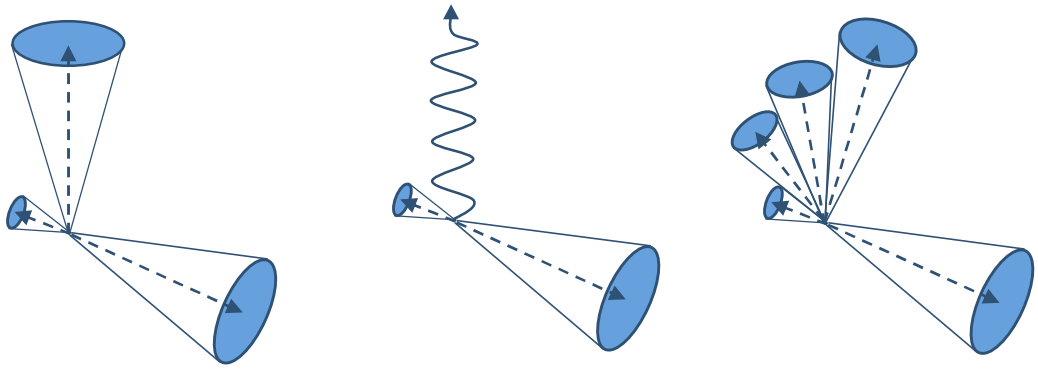


Figure 3.26: The *in situ* calibration chain of small-R and large-R jets. The calibration of small-R jets stops at the Jet Energy Scale (JES), while the large-R jets receive the additional Jet Mass Scale calibration. Reproduced from [115].



(a) (b) (c)

Figure 3.27: Topologies of events of the tag and probe techniques of the *in situ* calibration step of the JES: **(a)** dijet events for the η-intercalibration **(b)** Z+jet or γ+jet events, and **(c)** multijet systems for the multijet balance. In all these topologies, additional jets (or radiations in general) can exist, affecting the balance between the probe jet and the reference object.

3.3.2 Residual *in situ* calibration

The previous calibration steps (pile-up corrections, MC JES, GSC) mostly rely on the theoretical knowledge of the physics processes and of the detector in simulations: matrix element, showering, hadronization, underlying event, pile-up, particle showers in the active or dead material, the detector's geometry and materials, *etc.* Those calibrations only use experimental data events for the evaluation of some pile-up related uncertainties. This theoretical knowledge is imperfect, and the simulations are truncated to some degree of precision: simulations do not perfectly represent data. The modeling of the response to hadrons is complicated by the large variety of physics processes in hadronic interactions and by the non-compensating nature of the ATLAS calorimeters; as a comparison, the calorimeter response to electromagnetic particles such as electrons and photons is very well known due to their showers being easier to simulate accurately and to precise *in situ* measurements using $Z \to ee$ events: the energy scale of electrons is known with an uncertainty of less than 0.1 % for $25 < p_T < 70$ GeV in the central region [114]. As a final argument, regardless of how good the simulations are, experimental results should rely as little as possible on simulations.

The *in situ* calibrations aim at having a data-driven correction of the residual detector effects, that are different in data and in MC. After the MC-based calibrations described previously, $\left\langle \frac{p_T^{\text{MC reco}}}{p_T^{\text{MC truth}}} \right\rangle = 1$ is achieved. If the correction was applied to MC, this desired property would be broken, therefore the data/MC differences are applied to data to match MC.

The *in situ* calibrations are based on tag-and-probe methods, where the p_T of a probe jet is compared to that of a precisely known reference object, by exploiting transverse momentum conservation. This is done both in data and MC.

The *in situ* calibration step is a chain of multiple calibrations, that calibrate jets in one phase space region

thanks to jets in another phase space region or thanks to other objects, thus transferring the uncertainties from the reference object to the jet under study. The chain is depicted in Fig. 3.26, and comprises:

- A "relative" calibration, the η-intercalibration, using dijet events (see Fig. 3.27a), that aims at making the calibration homogeneous across the whole η range of the detector, by calibrating the forward $|\eta| > 0.8$ region of the detector relatively to the central region of the detector. The next *in situ* calibration can therefore focus on the proper calibration of the central region, as the η-intercalibration ensures that a correction derived centrally translates directly to forward jets as well.
- Multiple "absolute" calibrations, the $Z(\rightarrow ee)$+jet, $Z(\rightarrow \mu\mu)$+jet, γ+jet and multijet balances, that calibrate jets to the absolute scale by balancing them against a precisely know reference object: either other objects (Z bosons, photons, see Fig. 3.27b), or multiple other jets in the case of the multijet balance (see Fig. 3.27c).
- The *in situ* Jet Mass Scale (JMS), currently only used in the case of large-R jets, that corrects the mass of the jet using a well-known reference (considering that the focus of this thesis is small-R jets, these steps will not be described).
- The combination steps, that statistically combine separate *in situ* measurements.

This picture does not include the *in situ* measurements of the Jet Energy Resolution (JER), that are detailed later in Section 3.3.3.

The corrections are binned in p_T^{ref} (p_T^{avg} and η in the case of the η-intercalibration) to avoid the impact of the large p_T resolution of the probe jet. The correction is expressed as a function of the probe jet p_T with a change of variables.

The p_T response is defined as:

$$\mathcal{R} = \left\langle \frac{p_T^{\text{probe}}}{p_T^{\text{ref}}} \right\rangle \tag{3.13}$$

and the intercalibration coefficient as:

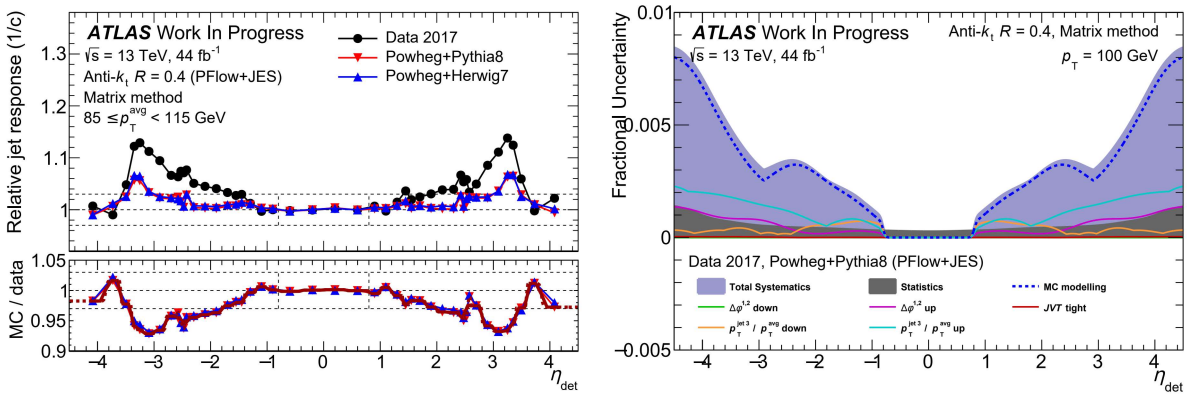
$$c = \frac{1}{\mathcal{R}} \tag{3.14}$$

The intercalibration coefficient c cannot be directly used as a calibration coefficient, since it is affected both by detector effects (that are the target) and by physics effects. Indeed, the transverse momentum balance is affected by possible additional radiations, like an additional jet (depicted in the topologies presented in Fig. 3.27) or out-of-cone radiations. The following ratio is then used as the calibration, as it is more resilient to such physics effects as long as they are well modeled in MC:

$$\mathcal{C} = \frac{c_{\text{data}}}{c_{\text{MC}}} \tag{3.15}$$

This aspect of convolution of physics and detector effects will be detailed more thoroughly in the discussion about the absolute implementation of the *in situ* calibration methods in Chapter 4, Section 4.1.

To mitigate these physics effects as much as possible, event selection criteria ensure the quality of the p_T balance: for instance, the probe and reference objects should be fairly back-to-back in the transverse plane, the additional jet should have a negligible p_T compared to the p_T average of the two leading jets, the JVT criterion is used to reject jets that are probably of pile-up origin, *etc.* The variations of these selection criteria provide systematic uncertainties on the measurement. Since the *in situ* calibrations are performed sequentially, the systematic uncertainties are propagated from each calibration to the next one. The uncertainty from the measurement of the reference object is also propagated. Two other sources of uncertainties are considered: the statistical uncertainty in each bin, and the modeling uncertainty. Indeed, the different MC generators model the physics differently, thus affecting the balance; the modeling uncertainty is taken to be the difference between two MC generators. The systematic uncertainties that quantify the physics effects are expected to be smooth, and are therefore rebinned until they become statistically significant, and smoothed so that they are not affected by statistical fluctuations. Finally, the "closure" of the method is tested as a cross-check: the calibration obtained by one method is applied to the jets and the calibration is obtained a second time using the exact same method, as it is expected to be 1 in case there is no methodological bias. In phase space regions where the closure is significantly different from 1 and not already covered by any of the other uncertainties, then an additional "non-closure" uncertainty is taken into account.



(a) (b)

Figure 3.28: **(a)** η-intercalibration in the $85 \leq \mathrm{p_T} < 115$ GeV bin and **(b)** the associated uncertainties at $\mathrm{p_T} = 100$ GeV, for one set of data-taking conditions (in this case, for the 2017 period).

3.3.2.1 η-intercalibration

The η-intercalibration being one of the main studies I performed, it is described thoroughly in Chapter 4. Here, I only summarize the main features of the η-intercalibration that make it differ from the other *in situ* calibrations, as well as the associated uncertainties.

The η-intercalibration exploits the transverse momentum balance between a reference jet in the central region of the detector, $|\eta| < 0.8$, and a probe jet in the forward region of the detector, $0.8 < |\eta| < 4.5$. This calibration flattens the JES across the detector η range. Its average calibration factor is fixed to 1 in the central region.

Due to the p_T of the reference jet having a large resolution, the mean of the $\frac{p_T^{\mathrm{ref}}}{p_T^{\mathrm{probe}}}$ distribution cannot be directly measured with precision, and the study cannot be binned in p_T^{ref}. Instead, $p_T^{\mathrm{avg}} = \frac{p_T^{\mathrm{probe}} + p_T^{\mathrm{ref}}}{2}$ is used for the binning, and the mean asymmetry $\mathcal{A} = \frac{p_T^{\mathrm{probe}} - p_T^{\mathrm{ref}}}{p_T^{\mathrm{avg}}}$ is measured with a Gaussian fit and then inverted to get c.

Two techniques can be used to determine the intercalibration coefficients:

- The "central-reference" method, that only uses dijet events that have one of the two leading jets in the central region;
- The "matrix" method, that uses all dijet events, regardless of the η region in which the two leading jets arrive. In this case, a system of constraints is minimized to obtain the intercalibration coefficients. This method achieves a much better statistical precision and is therefore used as the default one to produce the calibration.

The calibration is performed separately for the three following periods of LHC's run 2: 2015+2016, 2017, 2018. Indeed, the detector and operation conditions have changed enough in between the three periods to justify the determination of three calibrations. As of [72], the 2017 calibration is also used for 2018, with an extra non-closure uncertainty.[4]

The data and MC relative responses are shown in the upper part of Fig. 3.28a, and the calibration in the lower part. The associated uncertainties are shown in Fig. 3.28b. The η-intercalibration has one of the leading sources of uncertainties for jets in the forward region, dominated either by the modeling uncertainty or by the systematic variation of the selection on the third jet. The η-intercalibration comes with the following uncertainties:

- 532 statistical uncertainties for each of the periods, as there are 532 $(p_T^{\mathrm{avg}}, \eta)$ bins in the analysis.
- A "total systematic" uncertainty, that is the sum in quadrature of the systematic uncertainties designed to cover for physics effects: the systematic variations of the selections on $\frac{p_T^3}{p_T^{\mathrm{avg}}}$, $\Delta\phi^{1,2}$ and JVT, as well as the modeling uncertainty.
- 3 non-closure uncertainties, for three regions of phase space were the closure was found to be significantly different than 1 and not covered by the other uncertainties. Those are possibly due to a smoothing that was too strong, that smoothed out features of the calibration, and to a p_T freezing that occurred too low (I will comment those points in Chapter 4, Section 4.2.7).
- A non-closure uncertainty to use the calibration of 2017 onto 2018 data.

[4] I have computed the η-intercalibration independently for each year, as shown in Chapter 4; however my calibration has not been used in the cross-section measurement presented in Chapter 5 due to time constraints, because not all the remaining *in situ* analyses have been run.

3.3.2.2 V+jet balance

The three methods that exploit the balance of a probe jet against a well-measured Z boson (either decaying to two electrons or to two muons) or a photon are derived using the same possible two techniques: the Direct Balance (DB) and the Missing-p_T Projection Fraction (MPF). They allow to cover the transverse momentum range $17 < p_T < 1200$ GeV.

The Direct Balance method computes the response as:

$$\mathcal{R}_{\mathrm{DB}} = \left\langle \frac{p_T^{\mathrm{probe}}}{p_T^{Z/\gamma} |\cos \Delta\phi|} \right\rangle \tag{3.16}$$

where p_T^{probe} is the p_T of the probe jet (the leading jet of the event), $p_T^{Z/\gamma}$ is the p_T of the reference object, and $\Delta\phi$ is the azimuthal separation between the probe and reference object. Indeed, the DB method is sensitive to out-of-cone (OOC) radiations: while the Z boson or photon is well measured, there are always additional QCD radiations lying outside of the jet's cone, mostly in the same hemisphere as the jet, affecting the p_T balance. $p_T^{Z/\gamma}|\cos\Delta\phi|$ is therefore the projection of the boson's p_T onto the jet axis in the transverse plane, to partially reduce this effect.

The Missing-p_T Projection Fraction balances the reference object against the whole hadronic recoil measured at the constituent level (with Particle Flow Objects). This method is more stable against out-of-cone radiations, pile-up and the underlying event. Assuming that any missing p_T in the event is due to the low response to the hadronic recoil, the following equality holds true:

$$\vec{p}_T^{\,\mathrm{ref}} + \vec{p}_T^{\,\mathrm{recoil}} + \vec{p}_T^{\,\mathrm{miss}} = \vec{0} \tag{3.17}$$

At the event level, the response is defined as:

$$r_{\mathrm{MPF}} = -\frac{\vec{p}_T^{\,\mathrm{recoil}} \cdot \hat{n}^{\mathrm{ref}}}{p_T^{\mathrm{ref}}} \tag{3.18}$$

with $\hat{n}^{\mathrm{ref}}$ the unit vector in the direction of the reference object. Projecting Eq. (3.17) onto $\hat{n}^{\mathrm{ref}}$, and defining the response as the mean of the event-level responses $\mathcal{R}_{\mathrm{MPF}} = \langle r_{\mathrm{MPF}} \rangle$ yields:

$$\mathcal{R}_{\mathrm{MPF}} = \left\langle 1 + \frac{\vec{p}^{\,\mathrm{miss}} \cdot \hat{n}^{\mathrm{ref}}}{p_T^{\mathrm{ref}}} \right\rangle \tag{3.19}$$

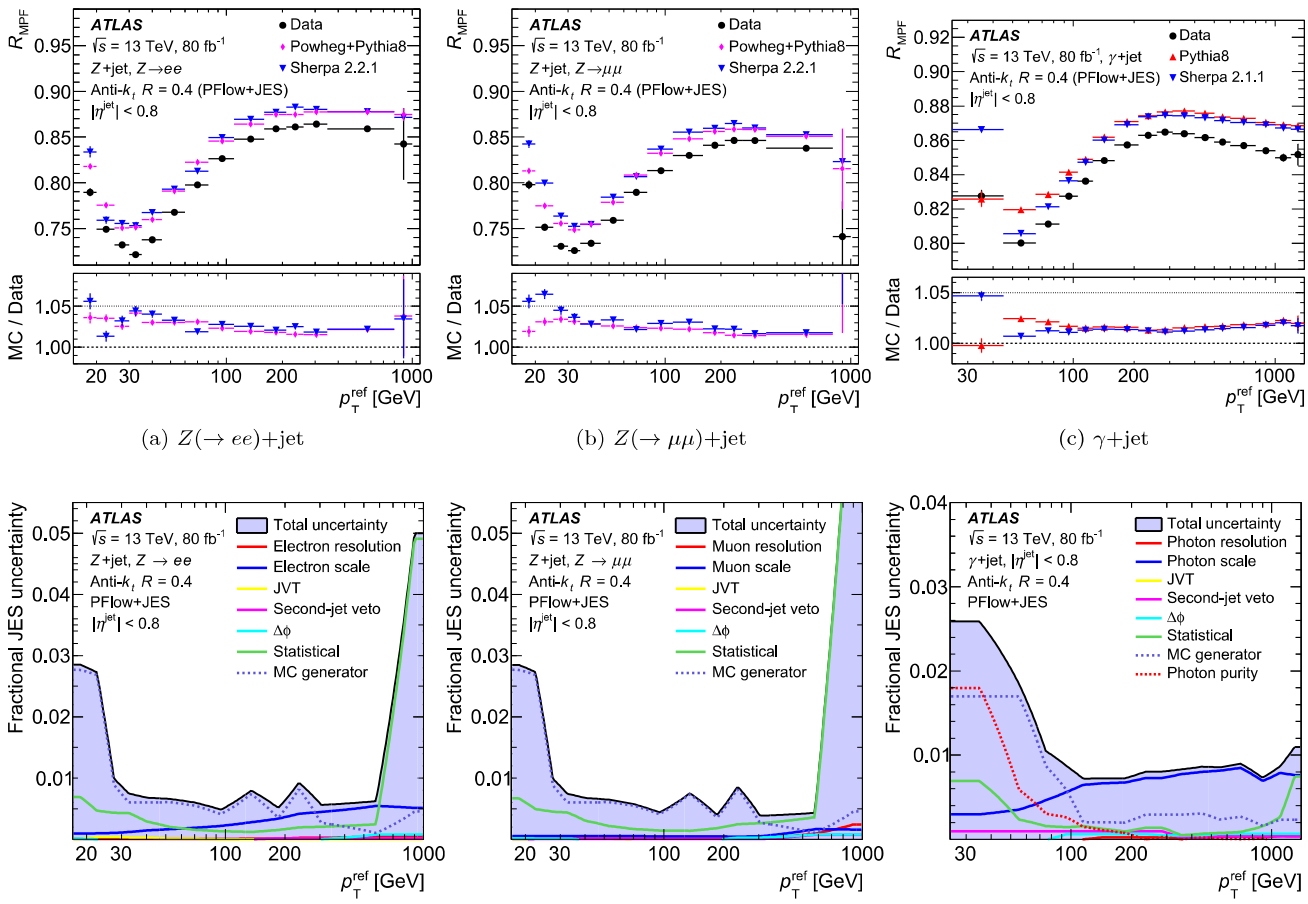
In both approaches, the response is obtained as the mean of a distribution thanks to the fit of a Modified Poisson distribution. The mean response is obtained in bins of p_T^{ref}, and mapped to the uncalibrated probe jet by finding the average p_T^{probe} in each bin. The two approaches yield compatible results, but the MPF method has proven to be more stable and is therefore used as the baseline.

Both methods suffer from a residual dependency to out-of-cone radiations that has to be corrected for:

- In the DB method, the jet's p_T will always be measured slightly lower than the Z boson or photon's p_T. The average p_T density of tracks is measured as a function of the angular distance ΔR between the track direction and the jet axis, allowing to determine the fraction of the radiation outside the jet cone, both in data and in MC. An out-of-cone uncertainty is evaluated on the basis of the simulation's ability to model the measured value of this quantity.
- In the MPF method, two effects are considered and estimated thanks to MC. The "showering correction" quantifies the migration of energy across the jet boundary of the calorimeter jet relative to the truth-particle jet. The "topology correction" accounts for the differences in calorimeter response between energy deposits in the periphery of the jet and energy deposits in the dense core of the jet. The uncertainties of both corrections are found to be significantly smaller than the DB's out-of-cone uncertainty, therefore the DB's out-of-cone uncertainty is kept as a conservative approach.

The data and MC relative responses are shown in the upper part of Figs. 3.29a to 3.29c, and the calibration in the lower part. The associated uncertainties are shown in Figs. 3.29d to 3.29f. The V+jet analyses come with the following uncertainties:

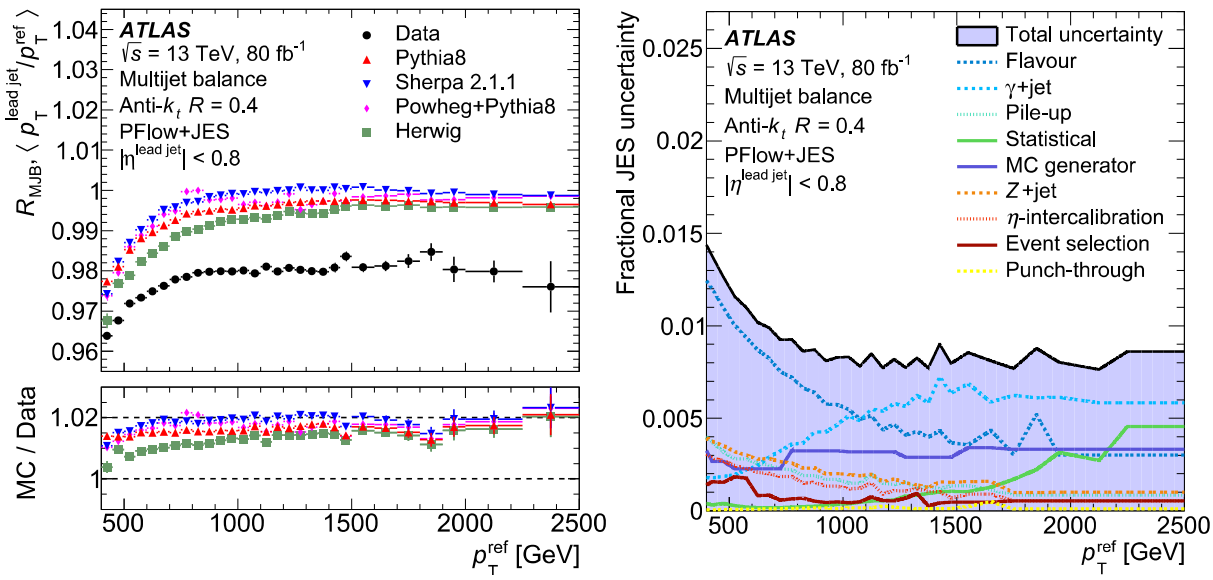
- Each bin comes with a statistical uncertainty: 14 for $Z(\to ee)$+jet, 14 for $Z(\to \mu\mu)$+jet, 16 for γ+jet.
- The uncertainties of the electron, muon and photon energy scales are propagated through the analysis: 2 uncertainties for the electron energy scale, 5 uncertainties for the muon energy scale, 2 uncertainties for the photon energy scale.
- The variations of the selection criteria on $\Delta\phi$, on the additional jet and on JVT also yield systematic uncertainties.
- A modeling uncertainty is evaluated by comparing two MC generators.
- The "shower topology" uncertainty evaluates the sensitivity to the out-of-cone radiations, as described above.



(a) $Z(\rightarrow ee)$+jet (b) $Z(\rightarrow \mu\mu)$+jet (c) γ+jet

(d) $Z(\rightarrow ee)$+jet (e) $Z(\rightarrow \mu\mu)$+jet (f) γ+jet

Figure 3.29: Jet Energy Scale **(a,b,c)** calibration and **(d,e,f)** uncertainties, for $Z(\rightarrow ee)$+jet, $Z(\rightarrow \mu\mu)$+jet and γ+jet. Reproduced from [72].



(a) (b)

Figure 3.30: Jet Energy Scale **(a)** calibration and **(b)** uncertainties for the multijet balance. Reproduced from [72].

3.3.2.3 Multijet balance

The multijet balance technique consists in using well-calibrated jets at low p_T (up to 1.2 TeV) to calibrate jets at higher p_T, by exploiting the balance of a high-p_T probe jet against the p_T of a recoil system of multiple lower-p_T jets. The leading jet therefore only receives the η-intercalibration, while the jets of the recoil system receive both the η-intercalibration and the calibration resulting from the combination of the V+jet methods, as described below in Section 3.3.2.4. This allows to calibrate jets up to 1.8 TeV, then a second iteration of the MJB balance method is performed to extend the range of the calibration up to 2.4 TeV, where jets in the $1.2 < p_T < 1.8$ TeV range are calibrated thanks to the MJB calibration obtained at the first step.

Selection criteria ensure the good quality of the events as for the other *in situ* techniques, and also prevent using dijet events to only have true multijet events. The response is defined as:

$$\mathcal{R}_{\mathrm{MJB}} = \left\langle \frac{p_T^{\mathrm{probe}}}{p_T^{\mathrm{recoil}}} \right\rangle \tag{3.20}$$

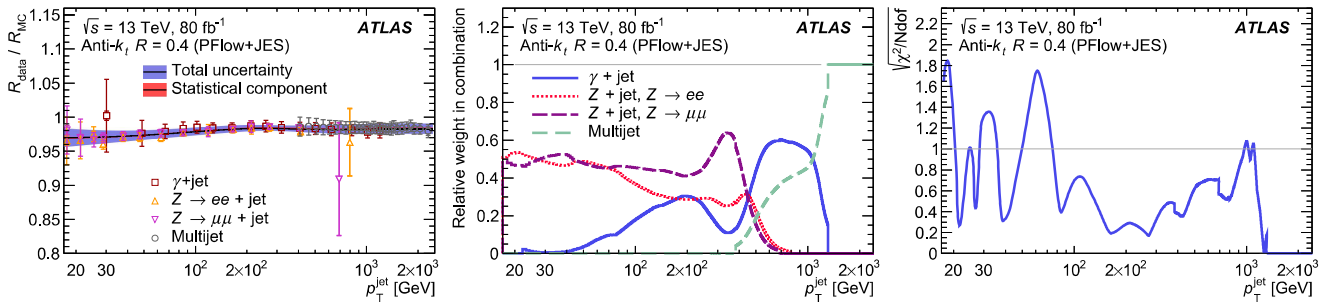
where p_T^{probe} is the p_T of the leading jet and p_T^{recoil} is obtained through the vector sum of the recoiling jets. The mean response is obtained in bins of p_T^{recoil} and then mapped to bins of p_T^{probe} with a similar method as for the V+jet methods.

The data and MC relative responses are shown in the upper part of Fig. 3.30a, and the calibration in the lower part. The associated uncertainties are shown in Fig. 3.30b. The multijet analysis comes with the following uncertainties:

- 28 statistical uncertainties, one for each bin.
- All the uncertainties affecting the jets in the recoil system are propagated, namely the uncertainties coming from pile-up, flavor, punch-through, η-intercalibration, Z+jet, γ+jet.
- The variations of the selection criteria provide 5 systematic uncertainties (p_T threshold for a jet to enter the recoiling system, angle α between the leading jet and the vector of the recoiling system, angle β between the leading jet and the nearest jet of the recoiling system, minimum asymmetry $\frac{p_T^2}{p_T^{\mathrm{recoil}}}$ between the p_T of the leading jet of the recoiling system and the p_T of the recoiling system, and JVT), grouped under "event selection" in Fig. 3.30b.
- A modeling uncertainty is evaluated by the comparison between two MC generators (Sherpa and Pythia8).

3.3.2.4 Combining the corrections and uncertainties

The correction factors and the uncertainties of the four "absolute" calibration methods ($Z(\to ee)$+jet, $Z(\to \mu\mu)$+jet, γ+jet, multijet) have to be combined to produce a single calibration covering the full range of jet p_T, from 17 GeV



(a) (b) (c)

Figure 3.31: Combination of the corrections of the *in situ* methods. **(a)** Combination of the corrections of the *in situ* methods. **(b)** Weights of the four methods as obtained by the χ^2 minimization. **(c)** Tension factor $\sqrt{\frac{\chi^2}{N_{\text{dof}}}}$; the uncertainties are multiplied by the tension factor in bins where the tension factor is larger than 1. Reproduced from [72].

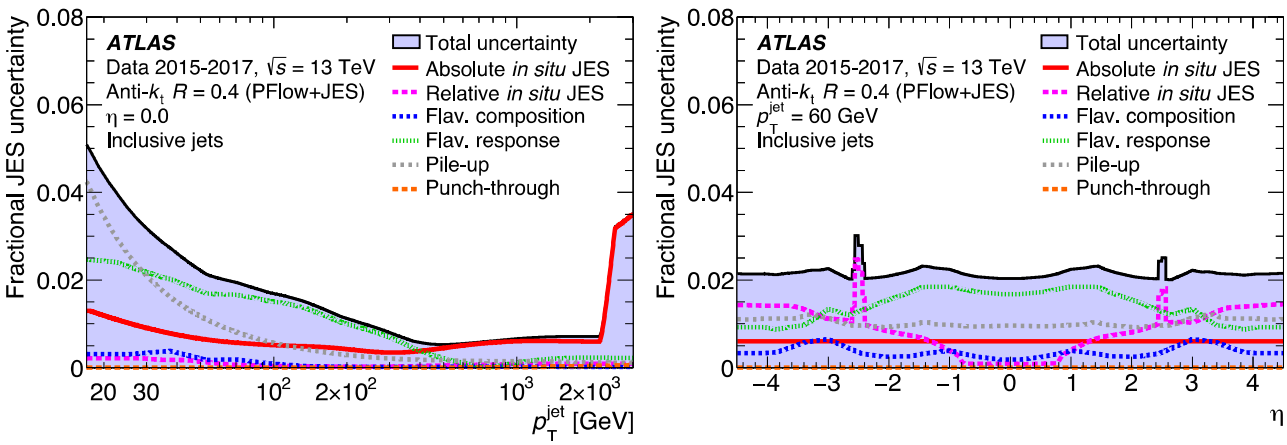
to 2.4 TeV. The different methods have overlapping p_T ranges; the combination procedure must account for their relative uncertainties and possible tensions between the measurements.

Each calibration is first smoothed thanks to an interpolation by a second-order spline and obtained in 1 GeV bins. Then, the calibration curve is obtained as a combination of the four methods, weighted by their uncertainties, thanks to a χ^2 minimization. The four absolute *in situ* calibrations are shown in Fig. 3.31a along with the combined calibration. Figure 3.31b shows the relative weights of the four methods.

Each uncertainty of the *in situ* measurements is linearly propagated through the combination with pseudo-experiments: 1 σ variations of the correction are obtained and propagated with the same procedure, except that the weights of the nominal calibration are used in the combination. The uncertainties are taken to be fully correlated over p_T, and independent of each other in most cases. There are two sets of V+jet uncertainties, the second set being those propagated through the MJB; the two sets are taken to be fully correlated. The tension between the measurements is quantified by the $\sqrt{\frac{\chi^2}{N_{\text{dof}}}}$, shown in Fig. 3.31c: in bins where a tension is observed, *i.e.* $\sqrt{\frac{\chi^2}{N_{\text{dof}}}} > 1$, the uncertainties of the combination result are scaled by $\sqrt{\frac{\chi^2}{N_{\text{dof}}}}$ to ensure overall agreement between methods within uncertainties. The sum in quadrature of the uncertainties is shown in Fig. 3.31a as an uncertainty band around the calibration.

Above 2.4 TeV, the calibration is frozen, and the uncertainties from the E/p analysis are used, conservatively.

All the uncertainties enumerated throughout this section are shown in Fig. 3.32, combined by category. In the central region, below 30 GeV, the uncertainties are dominated by the pile-up uncertainties and reach 5 % at 20 GeV. In the region between 30 GeV and 400 GeV, the flavor uncertainties dominate, with the total uncertainty being as low as 0.7 % around 400 GeV. Above 400 GeV, the total uncertainty is dominated by the absolute *in situ* methods, namely γ+jet below 1 TeV, then the multijet balance below 2.5 TeV, and by the E/p uncertainties above 2.5 TeV. The uncertainties *vs.* η at 60 GeV are dominated by the flavor uncertainties over the whole range, except in the most forward region and in the $2.3 < |\eta| < 2.6$ region where the η-intercalibration uncertainties are dominant. The spike at $|\eta| \approx 2.5$ is due to the end of the tracker coverage and to the transition between the precision and extended regions of the endcap electromagnetic calorimeter, as observed in Chapter 2, Fig. 2.22; this will be observed and discussed in Chapter 4 too.



(a) (b)

Figure 3.32: Combination of all the uncertainties. **(a)** Fractional uncertainties *vs.* p_T. **(b)** Fractional uncertainties *vs.* η. Reproduced from [72].

3.3.3 Jet Energy Resolution

The detector has a certain precision and a certain accuracy in the measurement of the observables, which in this case is the jet p_T. After the JES corrections, the accuracy is good, but there is spread in the measurements, an intrinsic imprecision called the Jet Energy Resolution (JER). The spread is assumed to have a Gaussian shape of width $\sigma(p_T)$. The measurement of the resolution in data and MC has multiple uses:

- It validates the simulation of the detector.
- It allows to correct for the data/MC difference by "smearing" the distributions of MC to match those of data, when the resolution is better in MC than in data, by convolving the MC with a Gaussian of width $\sigma_{\text{smear}} = \sigma_{\text{data}} \ominus \sigma_{\text{MC}}$ (where in this context $\ominus$ is the difference in quadrature: $a \ominus b = \sqrt{a^2 - b^2}$). In case the resolution is better in data, it is not possible to "anti-smear" data on an event-by-event basis, but an uncertainty can be derived. In both cases, systematic uncertainties on the measurement of the JER have to be propagated to the physics analysis.
- It allows to connect the truth and reco levels through the "folding": events from one truth bin spread to the adjacent reco bins due to the resolution (the bins "fold onto each other"). The unfolding step in physics analyses is used to correct for this effect on average.

Unlike for the JES, there is a functional form expectation for the relative JER:

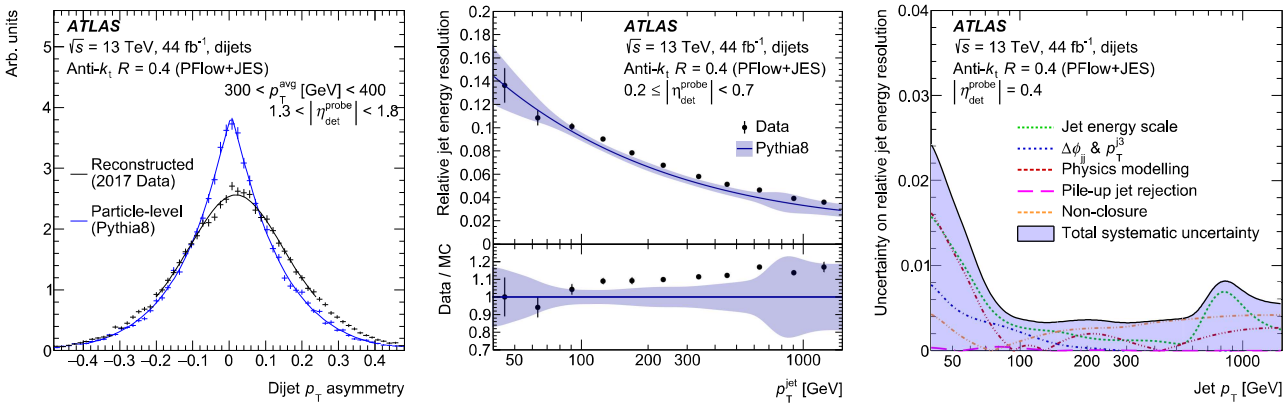
$$\frac{\sigma(p_T)}{p_T} = \frac{N}{p_T} \oplus \frac{S}{\sqrt{p_T}} \oplus C \tag{3.21}$$

where in this context $\oplus$ is the sum in quadrature ($a \oplus b = \sqrt{a^2 + b^2}$), and the different terms are:

- N, the Noise term, is mostly relevant at low p_T and expresses the p_T-independent part of the resolution $\sigma(p_T)$. It is decomposed into two components: the pile-up noise and the electronic noise, $N = N_{\text{pile-up}} \oplus N_{\text{electronic},\mu=0}$.
- S, the Stochastic term, is related to the physical development of the hadronic shower in the calorimeters: the number of particles in the hadronic shower, hence the number of energy deposits, go like $\propto \sqrt{p_T}$. In an ideal calorimeter with ideal operation conditions, this is the only term in the resolution.
- C, the constant term, mostly relevant at high p_T, is a constant fraction of the jet p_T. It has multiple sources: passive material in front of the calorimeters, starting point of the hadron shower, non-uniformities of response across the calorimeter (for instance due to detector geometry, readout system, detector aging, temperature gradient), *etc.*

3.3.3.1 Jet Energy Resolution measurement with dijet events

Measuring the JER implies to compare the jet p_T to that of a well-measured object. In MC, a truth jet can be compared to the associated reco jet (through geometrical matching). In data and MC, tag-and-probe methods can be used, just like for the *in situ* measurements of the JES, where the p_T of a well-calibrated probe object is balanced against that of a probe jet. Multiple methods are used in ATLAS: Z+jet, γ+jet and dijet; though the most recent analysis (at 13 TeV, [72]) only used the dijet topology, which is presented hereafter.



(a) (b) (c)

Figure 3.33: **(a)** Asymmetry distribution in MC truth and 2017 data, in the bin $300 < p_T^{\text{avg}} < 400$ GeV and $1.3 < |\eta^{\text{probe}}| < 1.8$, fitted by $\mathcal{A}^{\text{truth}}$ for MC truth and by $\mathcal{A}^{\text{truth}} \otimes \mathcal{G}(\mu_{\mathcal{A}}^{\text{detector}}, \sigma_{\mathcal{A}}^{\text{reco}})$ for data. **(b)** Relative Jet Energy Resolution (JER) and **(c)** the associated uncertainties, in the central part of the detector. Reproduced from [72].

In perfect dijet events, the two jets should have the same p_T. This is never the case, because of the detector effects and the physics effects:

- The detector effects are the JES and the JER. The full JES calibration is applied to the events, so only the effects of the JER are left.
- The physics effects are the additional radiations (like a third jet or the out-of-cone radiations). These effects make the asymmetry distribution have an intrinsic width, that is not caused by detector effects, and that has to be subtracted from the total asymmetry distribution.

Therefore, the physics effects, present at truth level, have to be subtracted from the the width measured at reco level:

$$\sigma_{\mathcal{A}}^{\text{detector}} = \sigma_{\mathcal{A}}^{\text{reco}} \ominus \sigma_{\mathcal{A}}^{\text{truth}} \tag{3.22}$$

Technically, this is done by assuming a functional form for the asymmetry distributions: first, the truth asymmetry distribution, only affected by physics effects, is fitted by an empirical function $\mathcal{A}^{\text{truth}}$ based on exponential curves; then the reco asymmetry distribution is fitted by a convolution of $\mathcal{A}^{\text{truth}}$ and a Gaussian, $\mathcal{A}^{\text{truth}} \otimes \mathcal{G}(\mu_{\mathcal{A}}^{\text{detector}}, \sigma_{\mathcal{A}}^{\text{reco}})$. The procedure to extract the distribution width from the asymmetry distributions is shown in one bin in Fig. 3.33a. Note that this subtraction of the physics effects makes the measurement "absolute" in the sense of Chapter 4, Section 4.1.

To go from the asymmetry resolution to the probe jet resolution, the following first order formula is used:

$$\sigma_{\mathcal{A}} = \frac{\sigma_{p_T}^{\text{probe}} \oplus \sigma_{p_T}^{\text{ref}}}{\langle p_T^{\text{avg}} \rangle} = \left\langle \frac{\sigma_{p_T}}{p_T} \right\rangle^{\text{probe}} \oplus \left\langle \frac{\sigma_{p_T}}{p_T} \right\rangle^{\text{ref}} \tag{3.23}$$

The jet p_T resolution is first determined in the central region, with both jets in $0.2 \leq |\eta| < 0.7$ (the seam at $\eta = 0$ is excluded to ensure that the reference jet energy is as cleanly measured as possible):

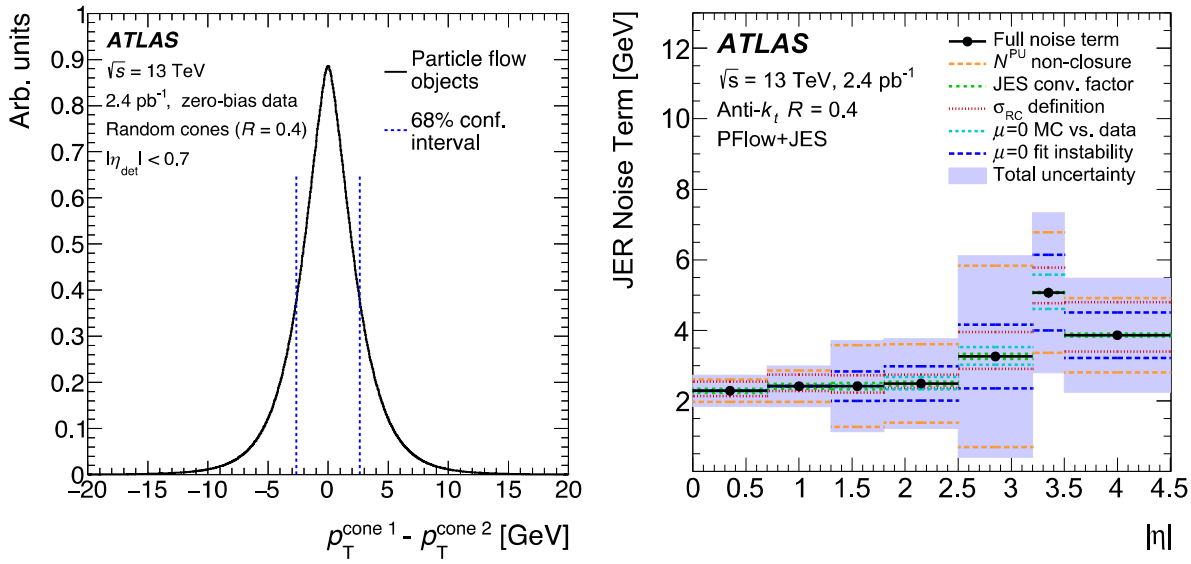
$$\left\langle \frac{\sigma_{p_T}}{p_T} \right\rangle^{\text{ref}} = \frac{1}{\sqrt{2}} \sigma_{\mathcal{A}} \tag{3.24}$$

Then the jet p_T resolution is determined in the forward regions, with:

$$\left\langle \frac{\sigma_{p_T}}{p_T} \right\rangle^{\text{probe}} = \sigma_{\mathcal{A}} \ominus \left\langle \frac{\sigma_{p_T}}{p_T} \right\rangle^{\text{ref}} \tag{3.25}$$

The data and MC relative resolutions are shown in the upper part of Fig. 3.33b. The associated uncertainties are shown in Fig. 3.33c. The dijet analysis for the JER comes with the following uncertainties:

- Statistical uncertainties, one per bin.
- One combined uncertainty on the JES.
- The variations of the dijet selection criteria provide 3 systematic uncertainties ($\Delta\phi^{1,2}$, $\frac{p_T^3}{p_T^{\text{avg}}}$, JVT).



(a) (b)

Figure 3.34: **(a)** Δp_T^{RC} distribution in the central region with 68 % confidence interval drawn. **(b)** pile-up noise term $N_{\mathrm{pile-up}}$ estimated through random cones, along with the systematic uncertainties. Reproduced from [72].

- A modeling uncertainty is evaluated by the comparison between two MC generators.
- A non-closure uncertainty, evaluated as the difference between the resolution using the dijet method and the resolution measured as $\frac{\sigma(\mathcal{R})}{\mathcal{R}}$ with $\mathcal{R} = \frac{p_T^{\mathrm{reco}}}{p_T^{\mathrm{truth}}}$.

3.3.3.2 The noise term of the Jet Energy Resolution

The direct fit of the N, S and C parameters on the resolution obtained by the tag-and-probe technique(s) results in a good control of the S and C parameters but a rather weak measurement of N, because of the low statistics and the degeneracy of N and S at low p_T. The two components of the noise term $N = N_{\mathrm{pile-up}} \oplus N_{\mathrm{electronic},\mu=0}$ can be measured directly. $N_{\mathrm{pile-up}}$ is measured through the random cones method, while $N_{\mathrm{electronic},\mu=0}$ is measured in MC simulations with no pile-up.

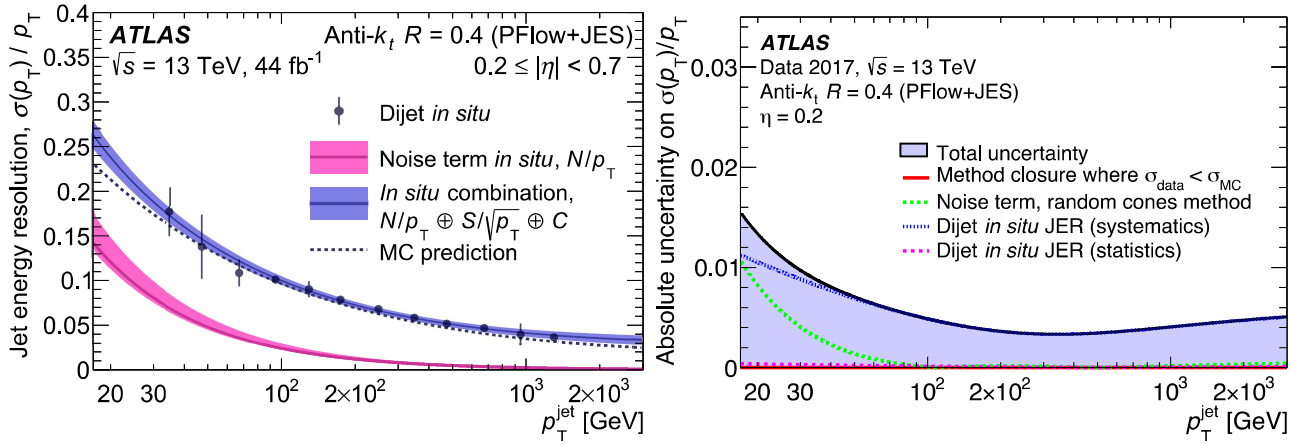
Using randomly triggered data (in the "zero bias" data stream), the energy deposits at the constituent level are summed in randomly drawn cones that have an area similar to that of anti-k_t jets with $R = 0.4$: on average, there is no hard activity in those cones. The cones are selected to be back-to-back in the transverse plane ($\Delta\phi = \pi$) and should therefore balance in p_T on average. The imbalances $\Delta p_T^{\mathrm{RC}} = p_T^{c1} - p_T^{c2}$ quantify the amount of ambient noise present in the detector. The width σ_{RC} of the distribution is taken to be the 68 % confidence interval of the distribution of Δp_T^{RC}, as shown in Fig. 3.34a, and the pile-up noise term is then $N_{\mathrm{pile-up}} = \frac{\sigma_{\mathrm{RC}}}{2\sqrt{2}}$ (2 to get the half-width and $\sqrt{2}$ because there are two cones). The pile-up noise term is estimated in multiple bins of $|\eta|$. The p_T in the cones is estimated at the constituent level, therefore the value of $N_{\mathrm{pile-up}}$ has to be corrected by the application of the JES.

A data/MC comparison, referred to as a "closure test", is performed. The noise term is estimated in MC simulations in two scenarios, with and without pile-up, and the quadrature difference between them gives the MC estimate of the pile-up-only noise term. The difference between the data estimate coming from the random cones method in the zero bias stream and this MC estimate is taken as an uncertainty.

The electric noise cannot be measured *in situ* by the random cones method, due to the topological clustering process. It is therefore estimated by an NSC fit to the JER measured in a dedicated MC simulation sample with no pile-up, where the electric noise term is the N term.

The total noise term is shown in Fig. 3.34b along with the systematic uncertainties. The estimation of the noise term through the random cones method and the no-pile-up sample comes with the following uncertainties:

- Statistical uncertainties, one per bin.
- An uncertainty on the JES.
- The "non-closure" uncertainty, *i.e.* the data/MC differences for the pile-up noise term, which is dominant.
- The variations of the quantile of the confidence interval for the measurement of σ_{RC}.



(a) (b)

Figure 3.35: **(a)** Jet Energy Resolution (JER) in the central region and **(b)** the associated uncertainties. Reproduced from [72].

- The data/MC differences for the electronic noise term are conservatively estimated to be of 20 %.
- An uncertainty due to the fit parameterization and stability of the electronic noise term.

3.3.3.3 Jet Energy Resolution combination and results

The combination of the different methods to obtain the JER is performed in a similar manner as the combination for the *in situ* corrections of the JES (Section 3.3.2.4). The functional form of Eq. (3.21) is fitted to the JER determined through the dijet method, with the noise parameter fixed to the value determined by the random cones method and the $\mu = 0$ simulations. The systematic uncertainties of the dijet method are taken fully correlated over the η range, and the uncertainties of the noise term are taken independent between the $|\eta| < 2.4$ and $|\eta| > 2.4$ ranges. The JER is shown in Fig. 3.35a and the uncertainties grouped by method are shown in Fig. 3.35b.

The number of statistical uncertainties is reduced using an eigenvalue decomposition of the correlation matrix of the JER: the leading eigenvalues are kept while the remaining ones are summed to form one single residual uncertainty. The number of eigenvalues to keep is determined such that the loss in correlation information is below a threshold: the difference between the correlation matrix obtained with the full set of uncertainties and the correlation matrix obtained with the reduced set should not exceed 0.05 in any bin. This allows to reduce the number of uncertainties to be propagated through the physics analysis while keeping as much information on the correlations as possible. This procedure is also used for the JES uncertainties for many physics analyses; for the cross-section analysis presented in Chapter 5, only the reduced set of JER uncertainties is used to reduce the number of modified transfer matrices, while the full set of JES uncertainties is propagated.

Chapter 4

Towards an Absolute η-intercalibration

Mais que diable allait-il faire dans cette chébèque ?

Alain Ayroles et Jean-Luc Masbou, *De Cape et de Crocs*, tome 1, 1995

Jets are collimated sprays of hadrons resulting from the hadronization of a quark or a gluon, as described in Chapter 1, Section 1.3. Those quarks and gluons are produced by high-energy proton-proton collisions thanks to the LHC, see Chapter 2, Section 2.1, at a center-of-mass energy of $\sqrt{s} = 13$ TeV during the Run 2 period of operation, *i.e.* from 2015 to 2018, and with a high pile-up environment ($\langle\mu\rangle = 33.7$ collisions per bunch crossing on average). When using simulations, the resulting hadrons can already be studied at particle-level: they are clustered together by a jet algorithm, resulting in particle-level jets, also called "truth" jets hereafter. The hadrons enter the ATLAS detector, see Chapter 2, Section 2.2, produce energy deposits in the calorimeters and tracks in the tracking systems, which are clustered together to obtain "reconstructed" jets, as described in Chapter 3, Section 3.1, also called "reco" jets hereafter.

The Jet Energy Scale (JES) has to be calibrated for reconstructed jets to be used in physics analyses. This is done in multiple steps described in Chapter 3, Section 3.3. The first steps aim at mitigating the effects of pile-up. The next steps are based on Monte Carlo simulations and aim at correcting the dominant effects in the JES and at improving the Jet Energy Resolution (JER). The last step of the JES calibration is composed of *in situ* techniques, *i.e.* data-driven techniques, in order for the final calibration to rely as little as possible on theory and simulations, see Chapter 3, Section 3.3.2. In addition, the *in situ* techniques provide estimates of the uncertainties on the calibration. Those techniques exploit the balance in transverse momentum between two objects for different topologies of events, to calibrate jets in a certain phase space region with respect to other objects (photons, Z bosons, or other jets).

During the course of this Ph.D., I had the opportunity to work on various aspects of the η-intercalibration, ranging from the understanding and improvement of the method, to obtaining the calibration in multiple different scenarios.

The η-intercalibration correction aims at making the detector response homogeneous across the whole η range of the ATLAS detector, by applying a two-dimensional p_T correction to jets, in bins of (p_T, η). I have obtained this calibration, see Section 4.3.1, making careful cross-checks at all the steps described in Section 4.2. In particular, the η-intercalibration has been used along with other methods to validate the recent improvements in the previous steps of the calibration chain, described in Section 3.3.

I also worked on new ideas to improve *in situ* calibrations, called the "absolute implementations" of the *in situ* techniques, that I studied in the context of the η-intercalibration. The absolute implementation of the *in situ* calibration techniques aims at a better deconvolution in the calibration coefficient of physics effects (additional radiations like extra jets or out-of-cone radiations) and detector effects, by using the MC simulations at particle level, which was not done before in the *in situ* techniques. The absolute implementation is described in Section 4.1, and the comparison between the results of the "relative implementation" (which is the current standard) and of the "absolute implementation" is shown in Section 4.3.2. This led me to study the physics effects independently of the detector effects, on Monte Carlo samples at particle level, which allowed me to improve the MC modeling uncertainty in the low-p_T region, as shown in Section 4.3.1.1. It also provided me with a method to check that the systematic uncertainties indeed cover the physics effects that they are designed to cover.

Unless otherwise noted, the discussions and results shown in this chapter are for small-R jets, *i.e.* jets reconstructed with the anti-k_t algorithm using $R = 0.4$. During this Ph.D. project, I have also ran parts of the η-intercalibration package for large-R jets, *i.e.* jets reconstructed with $R = 1.0$, as part of a teamwork effort to also

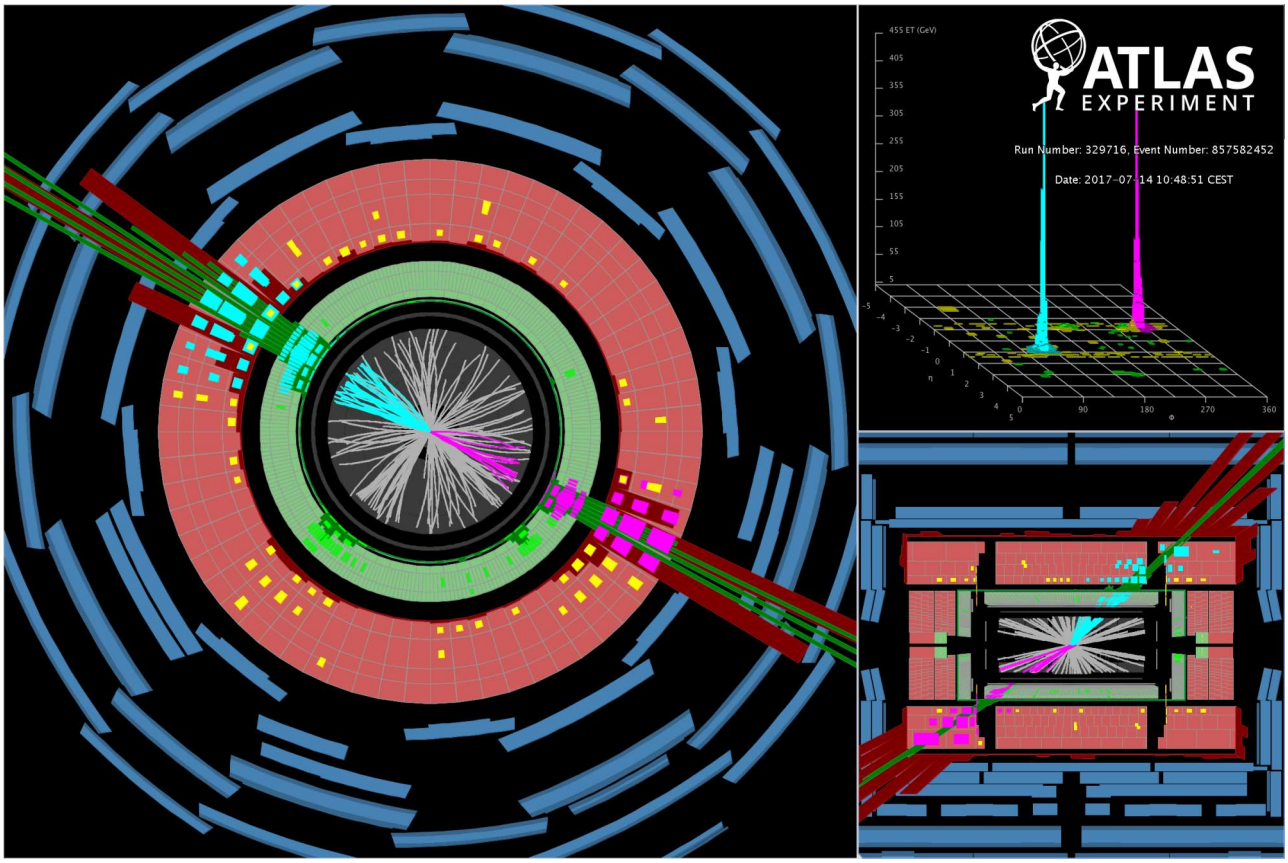


Figure 4.1: Display of a dijet event with $m_{jj} = 9.5$ TeV, produced in pp collisions at $\sqrt{s} = 13$ TeV data collected in 2017. The two jets with highest p_T have $p_T = 3.0$ TeV and 2.9 TeV, one is at $\eta = -1.2$ (magenta) and the other at $\eta = 0.9$ (cyan). The view of the event in the plane transverse to the beam direction is shown on the left side. The top-right figure represents the calorimeter clusters transverse energies in the (η, ϕ) plane. The bottom-right figure presents the event in the longitudinal view, *i.e.* (Z, Y) plane. Reproduced from the auxiliary figures of [116].

obtain a precision calibration for these types of jets. The η-intercalibration for large-R jets is therefore also shown in Appendix A.7 and briefly discussed.

Finally, I spent a significant amount of time refactoring the code of the legacy η-intercalibration package, a tedious but necessary task, without which these studies would not have been possible.

This chapter is organized as follows, after a brief introduction to the η-intercalibration:

- The changes that the absolute implementation implies for all the *in situ* calibration techniques and for the physics analyses are discussed in Section 4.1.
- The η-intercalibration is described in details in Section 4.2, whether it is for the relative or for the absolute implementations: its multiple steps and the recent improvements.
- Results are shown and analysed for both implementations of the η-intercalibration, and improvements with respect to previous studies are outlined in Section 4.3.

Introduction to the η-intercalibration

The *in situ* calibrations aim at comparing data and MC with a data-driven approach, and to correct data with the ratio of both, see Section 3.3.2. The detector has a structure that makes the response of jets different in the multiple regions of the detector, see Fig. 2.22, due to changes in the technologies used for calorimetry in the different regions, to the lack of tracking in the most forward region, to gaps between the instrumented regions, *etc.* The first of those *in situ* calibrations is therefore the η-intercalibration, that corrects the transverse momentum of forward jets with respect to central jets, to make the response of jets homogeneous across the whole η range of the detector.

To do so, the η-intercalibration exploits the transverse momentum balance between the two leading jets of multi-jets events, *i.e.* the two jets with the highest transverse momentum p_T, hereafter referred to as a the "dijet system". An example of a dijet event is shown in Fig. 4.1. The intercalibration factor c measures the average imbalance

between a probe jet in the forward region and a reference jet in the central region: $c = \left\langle \frac{p_T^{\text{ref}}}{p_T^{\text{probe}}} \right\rangle$. The ratio between data and MC is taken as the calibration factor to be applied onto data:

$$\mathcal{C}^{\text{relative}} = \frac{c_{\text{data}}}{c_{\text{MC}}^{\text{reco}}} = \frac{\left\langle \frac{p_T^{\text{ref}}}{p_T^{\text{probe}}} \right\rangle_{\text{data}}}{\left\langle \frac{p_T^{\text{ref}}}{p_T^{\text{probe}}} \right\rangle_{\text{MC reco}}} \tag{4.1}$$

This correction factor is derived in bins of (p_T, η), following the detector structure. Jets in the central region of the detector ($|\eta| < 0.8$) receive a correction of 1 on average as they are used as reference objects; they are corrected along with the forward jets in the next *in situ* calibrations.

Some selection criteria are needed for an event to be used in the η-intercalibration. Indeed, the transverse momentum balance between the two leading jets might be affected by additional radiations: most multi-jets events contain more than two jets, and some out-of-cone radiations might not be part of the jet due to its finite radius of $R = 0.4$. To mitigate those physics effects, the two following selection criteria are applied: the two leading jets of the dijet system must be fairly back-to-back in the transverse plane, $\Delta\phi^{1,2} > 2.5$ rad, and the additional jets must have a small enough transverse momentum, $p_T^{\text{jet }3}/p_T^{\text{avg}} < 0.25$. The calibration is also computed with varied values of these selection criteria, and compared to the calibration with the nominal values, to obtain systematic uncertainties. Those physics effects being modeled differently by the different MC generators (different jet substructures), a comparison between the calibration factors obtained with two different MC generators also provides a systematic uncertainty.

An important subtlety of the η-intercalibration measurement is that the intercalibration factor c cannot be determined accurately by simply looking at the mean of the $\frac{p_T^{\text{ref}}}{p_T^{\text{probe}}}$ distribution. Indeed, the jet energy resolution (JER, see Section 3.3.3) is rather large for jets, especially at low-p_T, making the transverse momentum of the reference jet fluctuate by large amounts. Therefore, as an intermediate step, the mean asymmetry $\langle \mathcal{A} \rangle = \left\langle \frac{p_T^{\text{probe}} - p_T^{\text{ref}}}{p_T^{avg}} \right\rangle$ is derived. The asymmetry is a lot more stable to the resolution fluctuations, as discussed in Section 4.2.4, and has the side advantages of being more symmetrical and of having rapidly falling tails. A simple Gaussian fit in a restricted range around the mean asymmetry allows to accurately determine the mean asymmetry $\langle \mathcal{A} \rangle$, which is used to obtain the intercalibration factor c as $c = \frac{2 - \langle \mathcal{A} \rangle}{2 + \langle \mathcal{A} \rangle}$.

Two strategies exist to derive the correction. The simplest strategy is the "central-reference method", that uses dijet systems in which the reference jet is in the central region of the detector and the probed jet is in the forward region. A second strategy is the "matrix method", that uses all dijet systems to study the relative response of all combinations of η regions. Those relative responses make an over-constrained system and are combined according to their statistical weight, providing the central-forward intercalibration coefficient c as in the central-reference method. Due to its much better statistical uncertainty and to its stability, the matrix method is used for the nominal results; the central-reference method only being used for cross-checks.

An asymmetry between the two leading jets can arise either from physics effects (additional radiations like a third jet, out-of-cone radiations), or from detector effects. Both types of effects are entangled in the intercalibration coefficients c at reconstructed level. That is the reason why the intercalibration factor c_{data} alone cannot be used as a calibration applied onto data, and why the double ratio of Eq. (4.1) is used: for the physics effects to cancel out between the data and the MC intercalibration coefficients.

The goal of the "absolute implementation" is to better deconvolve physics effects and detector effects in the η-intercalibration, and in all "tag and probe" *in situ* techniques in general.[1] To that extent, the MC simulations at particle level are used, since they are only affected by physics effects.

[1] In the ATLAS collaboration, the word "absolute" is often used with two different meanings in the context of *in situ* calibrations, which should not be confused.

The Z+jet and γ+jet techniques are often called "absolute" methods of calibration because they do not rely on the calibration of jets for their reference objects, while the η-intercalibration is often called a "relative" method of calibration because it calibrates one region of the (p_T, η) phase space with respect to another region of the phase space. One notable difference is that the relative techniques have to deal with the fact that their object(s) of reference, jets, have a large resolution, hence the use of the asymmetry for the η-intercalibration. The multijet balance is either classified as "absolute" or "relative", depending on the context: if the emphasis is that it calibrates high-energy jets with respect to already calibrated jets, it is called "absolute", and when the emphasis is on the fact that the reference object is a jet, with a large resolution, it is called a "relative" technique.

This is not to be confused with the "absolute implementation" of those techniques, which is the subject of Section 4.1. The "absolute implementation" aims at disentangling the detector effects and the physics effects in the tag-and-probe *in situ* techniques, to make these techniques only correct for detector effects present in data without any dependence on detector effects present in MC nor on physics effects, hence making them "absolute".

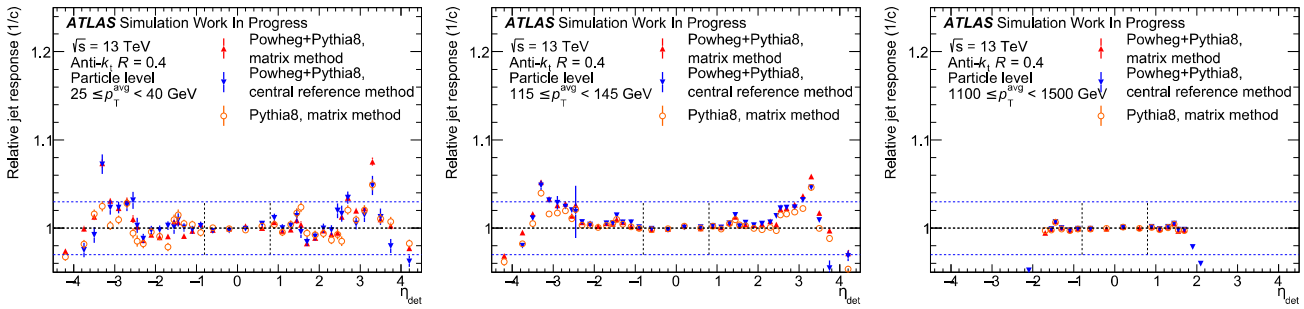
It is therefore possible to have an "absolute implementation" of an "absolute calibration method": by using the MC at truth level in the γ+jet or Z+jet methods.

The η-intercalibration in CMS

In the CMS collaboration, the other experiment at LHC studying similar physics as ATLAS (see Chapter 2), an η-intercalibration is also performed, with a similar goal [117–119]. There are differences in the calibration methods used before and after the η-intercalibration in the calibration chain, but the main calibration steps are fairly similar, at least in their goals (see Chapter 3, Section 3.3.2), and the differences do not significantly affect the way the η-intercalibration is performed: this calibration could be derived in the same way in the two collaborations.

For the η-intercalibration specifically, there are a few notable differences that are worth mentioning:

- The main difference is that CMS uses a Missing-p_T Projection Fraction (MPF) method (similar to that used for the γ+jet, Z+jet and multijet balances in ATLAS, see Chapter 3, Section 3.3.2.2), instead of the matrix method used in ATLAS. The MPF method is very sensitive to initial state radiations (ISR) and final state radiations (FSR) (see Chapter 1, Section 1.3.1.2), which is why these effects need to be corrected for in CMS, requiring additional systematic uncertainties.
- The response is "extrapolated to zero radiation", meaning that the response is obtained at different values of the selection criterion $\alpha = \frac{p_T^{\text{jet 3}}}{p_T^{\text{avg}}}$, and extrapolated linearly to a value of $\alpha = 0$, corresponding to perfectly dijet event, with no third jet. In ATLAS, this idea has been tested and has shown to have a negligible impact. The effect of the third jet is encapsulated in a systematic uncertainty that is at a reasonable level. Furthermore, not doing this extrapolation makes the interpretation of the systematic uncertainties clearer, and avoids using a model for the extrapolation.
- CMS uses a "resolution correction", that corrects for the bias stemming from the large resolution of the jet p_T. This bias is not corrected for in ATLAS, as the correction factors used by CMS cancel out in the data over MC ratio (and because in ATLAS the JER is only measured after the JES calibration is applied). However, the new equations derived in Section 4.2.4 (particularly Eq. (4.22)) make it possible to apply this correction very easily in ATLAS by measuring the width of the asymmetry distribution thanks to a Gaussian fit. This would not even require an additional uncertainty, as CMS has (the statistical uncertainty on the width of the distribution can be propagated with the bootstrap method).
- The p_T dependency of the calibration is fitted by a log-linear two-parameter function. In ATLAS, this is not performed; indeed, there is no physical justification for this parametrization (especially considering that the JES is not calibrated to the absolute scale yet), it introduces *ad hoc* uncertainties, it will inevitably introduce some non-closure (as I observed when I used a smoothing that was too strong in p_T, see Section 4.2.7.3) and there is no need for this extra step (besides maybe smoothing out statistical fluctuations, which are smoothed out in a different way in ATLAS).
- One single calibration is derived for both the forward and backward directions, meaning that the calibration is derived *vs.* $|\eta|$, not *vs.* η. In ATLAS, the detector is not assumed to behave exactly the same way in both directions: it is known that the calibration can be different in both regions, for instance due to dead tile modules.
- The η binning is not as granular as the one used in ATLAS. The granular binning in ATLAS is required due to the known presence of sharp response transitions between detector regions, and is made possible due to the large statistics available thanks to the matrix method. Besides that, the p_T range of the calibration is not as wide as that in ATLAS: $[60, 1500]$ GeV in CMS compared to $[20, 2000]$ GeV in ATLAS.
- CMS does not assess an "MC modeling" uncertainty, that would account for the difference in modeling by the different MC generators of the physics effects affecting the dijet balance. In ATLAS, this uncertainty is known not to be negligible: it used to be dominant in most phase space regions, and thanks to this work it is of the same order of magnitude as other uncertainties, see Section 4.2.8.7.
- CMS does not assess a "non-closure uncertainty". ATLAS does so (see Section 4.3.1.2).



(a) (b) (c)

Figure 4.2: Intercalibration factors $c_{\text{MC truth}}$ obtained in the case of the η-intercalibration. Any deviation from unity can only be caused by physics effects (additional jets, out-of-cone radiations), since these intercalibration factors are computed at truth level. The plots show different ranges in p_T^{avg}: **(a)** $25 \leq \text{p}_\text{T}^{\text{avg}} < 40$ GeV, **(b)** $115 \leq \text{p}_\text{T}^{\text{avg}} < 145$ GeV, **(c)** $1100 \leq \text{p}_\text{T}^{\text{avg}} < 1500$ GeV.

4.1 The absolute implementation for the *in situ* calibration

The "absolute implementation" aims at a better separation of physics effects and detector effects, for all tag and probe *in situ* techniques. It is therefore described here in the broad context of all *in situ* techniques, not just for the η-intercalibration: the concepts and formulas are general.

4.1.1 Deconvolving physics effects and detector effects in the intercalibration coefficient

All *in situ* techniques exploit the conservation in transverse momentum between a probe jet and a reference object to determine the intercalibration factor $c = \left\langle \frac{p_T^{\text{ref}}}{p_T^{\text{probe}}} \right\rangle$ in each phase space region. The calibration $\mathcal{C}$ is then built as the ratio of the data and MC intercalibration factors c_{data} and c_{MC} as defined in Eq. (4.1).

The intercalibration factor, both in data and in MC, can be affected by two essentially independent causes:

- Physics effects, like additional radiations (third jet, out-of-cone radiations) that can create a systematic imbalance between the probe jet and the reference object. For instance, this is observed in the context of the η-intercalibration, see Fig. 4.2.
- Detector effects. For instance, the detector response is different between a gluon-originated jet and a quark-originated jet, as shown in [109]. The detector effects have already been addressed by the previous steps in the calibration chain that are based on MC simulations, which capture the main effects, but residual detector effects remain as the simulations do not perfectly describe data.

As a first order hypothesis, detector effects and physics effects factorize at reco level (both in data and in MC): $c_{\text{reco}} = c^{\text{physics}} \times c^{\text{detector}}$. The MC intercalibration factor at truth level is only affected by physics effects: $c_{\text{MC truth}} = c_{\text{MC}}^{\text{physics}}$.

It is then possible to remove the physics effects from the data intercalibration coefficient thanks to a ratio with the MC intercalibration coefficient at truth level, to obtain the "absolute" calibration $\mathcal{C}_{\text{data-specific}}^{\text{absolute}}$:

$$\mathcal{C}_{\text{data-specific}}^{\text{absolute}} = \frac{c_{\text{data}}}{c_{\text{MC truth}}} = \frac{c_{\text{data}}^{\text{physics}} \times c_{\text{data}}^{\text{detector}}}{c_{\text{MC}}^{\text{physics}}} = c_{\text{data}}^{\text{detector}} \tag{4.2}$$

with the second hypothesis that $c_{\text{data}}^{\text{physics}} = c_{\text{MC}}^{\text{physics}}$: physics should be well described in MC. The ratios of Eq. (4.1) and Eq. (4.2) look very much alike, only the level of MC at the denominator changes (truth level or reco level). A similar correction is derived for MC, called hereafter the "MC-specific" correction. The physics effects can be removed from the MC intercalibration factor by a ratio with the MC at truth level:

$$\mathcal{C}_{\text{MC-specific}}^{\text{absolute}} = \frac{c_{\text{MC reco}}}{c_{\text{MC truth}}} = \frac{c_{\text{MC}}^{\text{physics}} \times c_{\text{MC}}^{\text{detector}}}{c_{\text{MC}}^{\text{physics}}} = c_{\text{MC}}^{\text{detector}} \tag{4.3}$$

This correction coefficient only contains effects of the simulation of the detector.

The relative and absolute implementations are related through

$$\mathcal{C}^{\text{relative}} = \frac{\mathcal{C}^{\text{absolute}}_{\text{data-specific}}}{\mathcal{C}^{\text{absolute}}_{\text{MC-specific}}} = \frac{c_{\text{data}}}{c_{\text{MC truth}}} \times \frac{1}{\mathcal{C}^{\text{absolute}}_{\text{MC-specific}}} \tag{4.4}$$

This shows that the goal of deconvolving the physics effects and the detector effects, which both affect the MC intercalibration factor at reco level, is achieved. The physics effects and the simulated detector effects that are in $c_{\text{MC reco}}$ are respectively isolated in $c_{\text{MC truth}}$ and in $\mathcal{C}^{\text{absolute}}_{\text{MC-specific}}$. The physics effects and the detector effects that are in c_{data} are respectively isolated in $c_{\text{MC truth}}$ (thanks to the hypothesis $c^{\text{physics}}_{\text{data}} = c^{\text{physics}}_{\text{MC}}$) and in $\mathcal{C}^{\text{absolute}}_{\text{data-specific}}$. Therefore, $\mathcal{C}^{\text{absolute}}_{\text{data-specific}}$ only contains data detector effects, and $\mathcal{C}^{\text{absolute}}_{\text{MC-specific}}$ only contains simulated detector effects.

The two hypotheses $c_{\text{reco}} = c^{\text{physics}} \times c^{\text{detector}}$ and $c^{\text{physics}}_{\text{data}} = c^{\text{physics}}_{\text{MC}}$ are also implicitly done in the relative implementation, for the physics effects from the numerator and the denominator to cancel out, so these hypotheses are neither new nor specific to the absolute implementation.

4.1.2 Benefits of the absolute implementation

4.1.2.1 Isolation of the different effects and meaning of the calibration

The absolute implementation allows to study detector effects and physics effects separately. Indeed, as said, the truth-level intercalibration factor $c_{\text{MC truth}}$ contains the physics effects, while $\mathcal{C}^{\text{absolute}}_{\text{data-specific}}$ contains the detector effects of data and $\mathcal{C}^{\text{absolute}}_{\text{MC-specific}}$ contains the detector effects of MC at reco level.

This allows to study the physics effects independently from the detector effects, which has been done in the case of the η-intercalibration: it allowed me to check that the systematic uncertainties of selection are not underestimated because they do cover the physics effects that they are designed to cover, in most phase space regions, see Section 4.2.8.6 for more details.

The difference with the relative calibration appears clearly from the comparison of Eqs. (4.2) and (4.3) and the following equation describing the physics and (residual) detector effects for the relative implementation:

$$\mathcal{C}^{\text{relative}} = \frac{c_{\text{data}}}{c_{\text{MC reco}}} = \frac{c^{\text{physics}}_{\text{data}} \times c^{\text{detector}}_{\text{data}}}{c^{\text{physics}}_{\text{MC}} \times c^{\text{detector}}_{\text{MC}}} = \frac{c^{\text{detector}}_{\text{data}}}{c^{\text{detector}}_{\text{MC}}} \tag{4.5}$$

The physics effects cannot be studied independently of detector effects in the relative implementation, as $c_{\text{MC truth}}$ is never looked at in this approach, and the relative calibration $\mathcal{C}^{\text{relative}}$ convolved the detector effects of both data and MC.

The absolute implementation is also more appropriate for a "data-driven" correction. Indeed, it suppresses the dependency to the MC simulation beyond the truth level, *i.e.* the simulation of the detector with GEANT4 is not used in $\mathcal{C}^{\text{absolute}}_{\text{data-specific}}$ (of course it is still used in $\mathcal{C}^{\text{absolute}}_{\text{MC-specific}}$).

The relationship with the truth level is well-defined when using the absolute implementation. Indeed, by applying the absolute calibrations both to data and to MC with $\mathcal{C}^{\text{absolute}}_{\text{data-specific}}$ and $\mathcal{C}^{\text{absolute}}_{\text{MC-specific}}$ respectively, the following two equalities are achieved for *in situ* topologies (within uncertainties): $\mathcal{R}^{\text{data}} = 1$ and $\mathcal{R}^{\text{MC}} = 1$, where $\mathcal{R}$ is the detector response, *i.e.* $\mathcal{R} = \left\langle \frac{p_T^{\text{reco}}}{p_T^{\text{truth}}} \right\rangle$.

On the contrary, the relative implementation, with $\mathcal{C}^{\text{relative}}$ applied to data only, only achieves $\frac{R^{\text{data}}}{\mathcal{R}^{\text{MC}}} = 1$. In this case, the relationship with the truth level comes from the previous calibration steps of the MC JES, that achieve $\left\langle \frac{E^{\text{MC reco}}}{E^{\text{MC truth}}} \right\rangle = 1$. However, the MC JES calibrates the energy whereas the *in situ* techniques calibrate the transverse momentum; and the topologies used are not the same ones for the MC JES and for the *in situ* calibrations. Though not having any important impact, this makes the relative implementation have a less direct relationship with the truth level.

4.1.2.2 Possible improvements in the uncertainties

Some of the dominant uncertainties of the *in situ* calibration could be better understood and thus possibly reduced. For instance, the MC modeling or the flavor dependency could be better understood; the additional jet uncertainty is also pretty strong in the forward region. See the fractional uncertainties in Fig. 4.3: those uncertainties are dominant in some phase space regions.

As an example, in the case of the η-intercalibration, I have been able to check that the additional jet uncertainty is of the same order of magnitude than the physics effects that it is supposed to cover, as shown in Section 4.2.8.6.

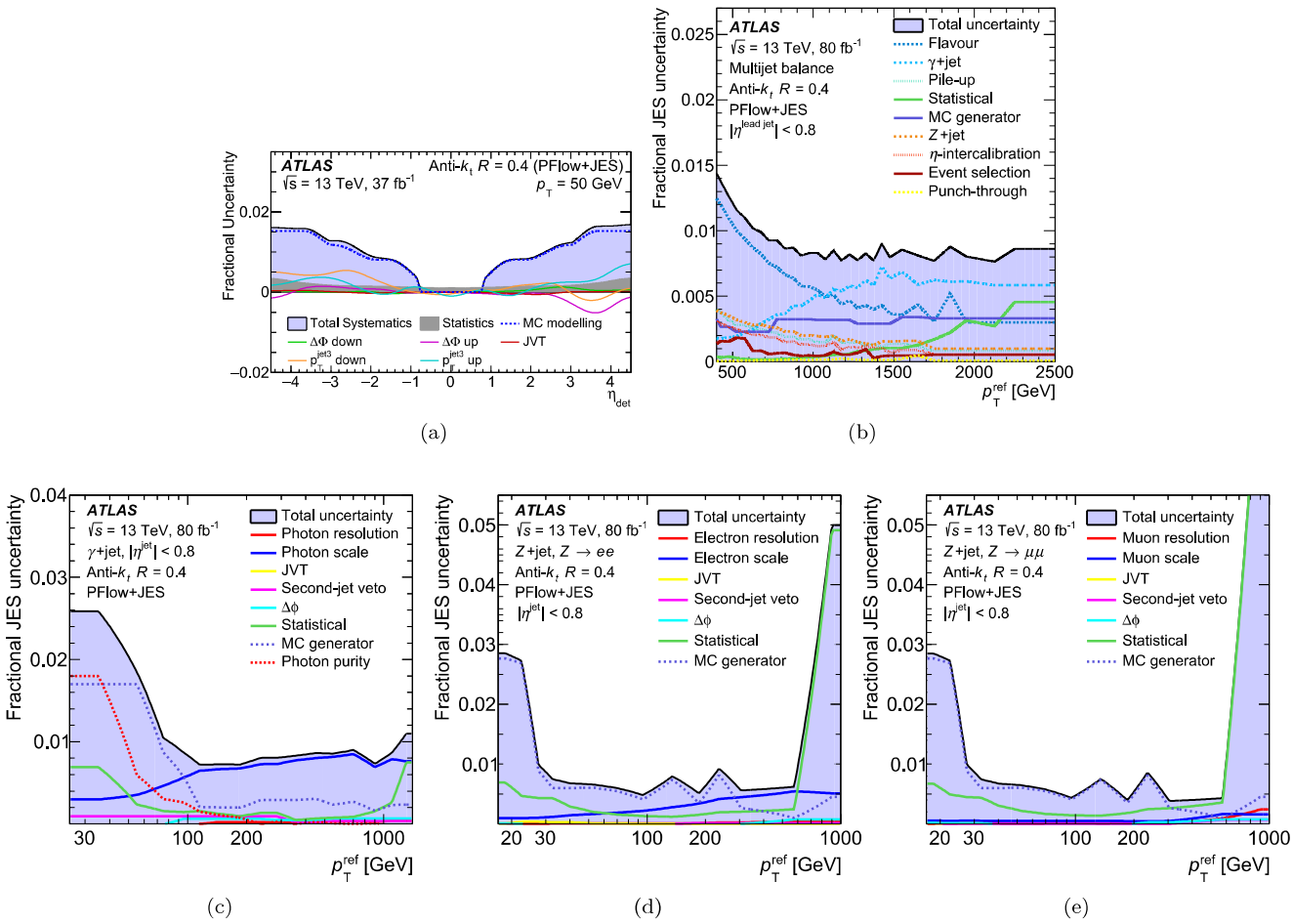


Figure 4.3: Fractional uncertainties for **(a)** the η-intercalibration, **(b)** the multijet balance, **(c)** the γ+jet balance, **(d)** the $Z(\rightarrow ee)$+jet balance and **(e)** the $Z(\rightarrow \mu\mu)$+jet balance. Reproduced from [72].

Therefore, any improvement of those uncertainties by future studies (*i.e.* any improvement in the event selection, since those uncertainties come from variations of the selection) would have a metric to show that they are neither underestimated, nor overestimated.

The absolute implementation makes sure to avoid the double-counting of uncertainties due to detector effects. Indeed, some detector effects are taken into account in separate uncertainties, like the flavor uncertainties, and thus should not be counted again in the uncertainties of the *in situ* method.

Such a reasoning allowed me to reduce the MC modeling uncertainty in the case of the relative implementation of the η-intercalibration, by avoiding to double-count the uncertainties due to the difference in detector response between the different flavors of jets, see 4.2.8.7. The absolute implementation makes such a reasoning easier to come up with and to implement in practice, as all the possible effects are studied separately.

4.1.3 Dealing with the absolute implementation of the calibration in a physics analysis

The absolute implementation produces two *in situ* calibration coefficients: one for data, $\mathcal{C}^{\text{absolute}}_{\text{data-specific}}$, and one for MC, $\mathcal{C}^{\text{absolute}}_{\text{MC-specific}}$. The relative implementation produces only one calibration for data, $\mathcal{C}^{\text{relative}}$.

In practice, applying two *in situ* calibrations should not be any different than applying only one, because in a physics analysis both data and MC are calibrated anyway: the *in situ* calibration is not the only calibration, the previous steps of the calibration chain are applied both to data and to MC. The only thing that would change would be the provider of the calibration, therefore it would be seamless for the user.

The values of selection criteria applied for the event selection in the analysis would have to be studied again (like with any calibration change). Indeed, as already mentioned, the meaning of data and MC calibrated with the absolute implementation is a bit different than with the relative implementation: the absolute calibration aims at

removing independently the (residual) detector effects in data and the (residual simulated) detector effects in MC, while the relative implementation convolves detector and simulated detector effects in the correction applied onto data and does not correct MC, *i.e.* the relative implementation corrects data to MC.

More storage used.
For the study of every systematic uncertainty in a physics analysis, the calibrations for data and MC should be varied at the same time. This does not lead to any higher complexity of the analysis, but to more storage used.

In a typical physics analysis using jets calibrated with the relative implementation (like the cross-section measurements, see Chapter 5), the propagation of a calibration uncertainty is done as follows:

1. Unfolding using the nominal calibration;
2. Varying the calibration of data (or MC acting as pseudo-data) by the amount of the uncertainty on the calibration;
3. Unfolding using the varied data (or the varied pseudo-data) and nominal MC;
4. Comparing the spectra unfolded with the nominal calibration and with the varied calibration.

With the absolute implementation, the second step changes: instead of just varying the calibration applied onto data by the amount of the uncertainty on the data-specific calibration, the MC-specific calibration is also varied by the amount of the same uncertainty, at the same time. For this approach, one would need to store the calibration-varied jets not only for data (or MC acting as pseudo-data) but also for MC, before filling histograms.

This would double the total size of the input files before filling the histograms, and therefore require more CPU to create the input files and to process them into histograms. It would also require one transfer matrix for every single systematic uncertainty, instead of one transfer matrix only for the JER systematic uncertainties, thus drastically increasing the CPU and RAM required to create the histograms, and the storage size required.

While being not conceptually and practically very different from the unfolding after the relative implementation of the calibration, the unfolding after the absolute implementation might require a lot more computational resources.

Solution: using ratios of correction factors.
Another option than the one described above is possible. In a physics analysis, one could only correct data (or MC) with the ratio of the correction factors. This ratio is:

$$\frac{\left(\mathcal{C}^{\text{absolute}}_{\text{data-specific}}\right)^{\text{selection}}}{\left(\mathcal{C}^{\text{absolute}}_{\text{MC-specific}}\right)^{\text{selection}}} = \frac{\frac{(c_{\text{data}})^{\text{selection}}}{(c_{\text{MC truth}})^{\text{selection}}}}{\frac{(c_{\text{MC reco}})^{\text{selection}}}{(c_{\text{MC truth}})^{\text{selection}}}} = \frac{(c_{\text{data}})^{\text{selection}}}{(c_{\text{MC reco}})^{\text{selection}}} = \left(\mathcal{C}^{\text{relative}}\right)^{\text{selection}} \tag{4.6}$$

where "selection" is the nominal calibration or one of its systematic variations. This would be closer to what users are currently doing as they would not need to vary both the calibrations of data and MC to estimate the systematic uncertainties: only varying the calibration applied onto data (or MC) would be enough. However, the two absolute calibrations being correlated, some care should be taken when computing the uncertainties and the correlation matrices.

At first glance, it looks like this ratio of the absolute calibrations in Eq. (4.6) is simply equal to the relative implementation, that does not use the MC at truth level. So, why not using the relative implementation in the first place? In fact, this ratio does not amount to using the correction from the relative implementation, because most of the benefits listed in Section 4.1.2 remain:

- The effects of physics and of the detector are isolated before doing the ratio, allowing for a better understanding of them and maybe a reduction of the uncertainties.
- It makes sure that there is no double-counting of detector effects.

However, it should be checked in a physics analysis that the uncertainties with this approach are not higher. They are not expected to be higher, due to the correlations between the data-specific and the MC-specific corrections.

4.2 Detailed description of the η-intercalibration

This section aims as describing the multiple steps of the η-intercalibration in more details:

- How the sample of events is constituted, both for data and MC, and how events are normalized, in Section 4.2.1;
- What is the binning of the analysis, in Section 4.2.2;
- What is the asymmetry and how the mean asymmetry is determined, in Section 4.2.4;
- How the intercalibration coefficients are determined, either with the central-reference method in Section 4.2.5 or with the matrix method in Section 4.2.6;
- How the calibration is computed from intercalibration coefficients, and how the Gaussian kernel smoothing is applied, in Section 4.2.7;
- How the different uncertainties are dealt with, in Section 4.2.8;

A brief description of the code used for the η-intercalibration can be found in Appendix A.1. The method described hereafter is only slightly different for large-R jets, the differences being listed in Appendix A.2.

Though a lot of the work presented in this section has been done by previous analyzers working on the η-intercalibration, everything is still presented here because a detailed description of all the steps of this calibration is necessary to understand the work I did, and because I made various cross-checks for most of the steps. An emphasis will be put on the work I did myself.

4.2.1 Event samples

All the datasets are provided by the JetETmiss group of ATLAS. The datasets used are the dijet ones.

Experimental data (called "data" hereafter) was recorded during the Run 2 data-taking period of the LHC (from 2015 to 2018) with collisions of protons at an energy of 13 TeV in the center of mass, at high pile-up conditions (high-μ). The data-taking periods are referenced by year, *i.e.* 2015, 2016, 2017, 2018. The events recorded in 2015 and 2016 are analysed together, as the conditions in the measurement (detector, triggers, processing) were similar enough.

Simulated data (called "MC" in this note) uses Monte Carlo generators, see Chapter 1, Section 1.3.1.6, to simulate collisions of protons at an energy of 13 TeV, providing the "truth" level particles. Pile-up is added (as described in Section 2.2), and the detection is fully simulated with Geant4, providing the "reconstructed" ("reco") level.

Jets are reconstructed with the anti-k_T algorithm, see Chapter 3, Section 3.1.2.1, using a radius of $R = 0.4$, either at truth level (for MC) using the truth particles, providing the truth jets, or at reco level (for data and MC), with the PFlow algorithm, see Section 3.1.1.2. Jets at reco level are calibrated, see Section 3.3.1, with the improved calibration described in [89], *i.e.* the following calibration steps are applied:

1. Improved pile-up subtraction based on the jet area and improved residual pile-up subtraction (described in Section 3.3.1.1);
2. Improved MC JES (described in Section 3.3.1.2);
3. Global Neural Network Calibration (GNNC, described in Section 3.3.1.3).

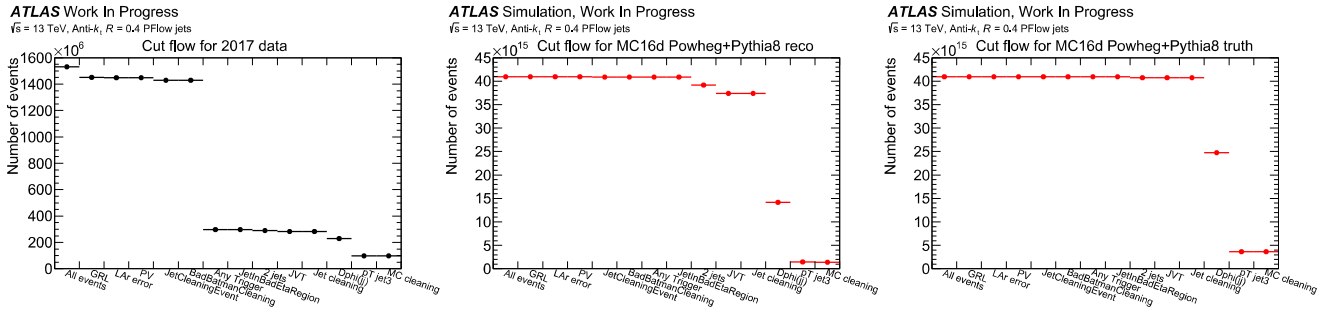
No *in situ* correction is applied prior to the η-intercalibration study, since the η-intercalibration correction is always the first *in situ* calibration to be derived and applied.

During the course of this Ph.D., multiple calibrations have been used, following the recommendations of the JetETmiss group of ATLAS. For instance, the first calibration used was the one described in [72], *i.e.* using the GSC instead of the GNNC and in general not benefiting from the improvements of [89]. The results of the η-intercalibration looking very similar between the different calibrations, only those obtained with the latest calibration sequence are shown.

Monte Carlo event samples

The different MC generators used are listed in Table 4.1. For analyzers of ATLAS, internal details of the tags and identifiers are provided in Appendix A.3. For the work done during this Ph.D., Powheg+Pythia8 is used as the nominal MC generator and the alternative MC generator for the MC modeling uncertainty (see Section 4.2.8.7) is Powheg+Herwig7 (with angular-ordered shower). All the other MC generators presented here have been ran at one point or another, but not kept in the final results, for the following reasons:

- Pythia8 and Sherpa are leading order (LO) generators. The η-intercalibration being very sensitive to the modeling of the third jet, the use of next-to-leading order (NLO) generators yields more accurate calibrations, resulting in much lower modeling uncertainties, as shown in the comparison between my new calibration only using NLO generators and the previous calibration, see Section 4.3.1.1.



(a) 2017 data (b) Powheg+Pythia8 reco level (c) Powheg+Pythia8 truth level

Figure 4.4: Unweighted number of events after each selection step. These plots only give qualitative information. The most stringent selection criteria are the physics ones: the requirements for multijet events to be dijet events.

- Not enough events have been generated with the Sherpa generators. This would have forced an increase of the bin size, but the detector has sharp variations in η that induce sharp fluctuations of the calibration and hence necessitate small η bins.
- The Herwig7 generator had too many events with negative weights below 100 GeV, giving nonsensical asymmetry distributions. It could not be used for the small-R η-intercalibration. The large-R η-intercalibration, starting at a higher p_T (160 GeV) was able to use Herwig7 for cross-checks.

MC generator	Matrix elements	Parton shower	Hadronization	PDFs
Pythia8.230	$2 \to 2$ LO	Dipole-style p_T-ordered	Lund string	NNPDF2.3
Powheg V2 + Pythia8.235	$2 \to 2$ NLO	Dipole-style p_T-ordered	Lund string	NNPDF3.0NLO
Powheg + Herwig7	$2 \to 2$ NLO	Angle-ordered	Cluster	NNPDF3.0NLO
Herwig 7.1.3	$2 \to 2$ NLO	Angle-ordered	Cluster	MMHT2014NLO
Herwig 7.1.3	$2 \to 2$ NLO	Dipole	Cluster	MMHT2014NLO
Sherpa 2.2.5	$2 \to 2$ LO	Dipole	Cluster	CT14NNLO
Sherpa 2.2.5	$2 \to 2$ LO	Dipole	Lund string	CT14NNLO

Table 4.1: Characteristics of the different MC generators used here. See Section 1.3.1 for explanations and references.

Powheg+Pythia8 has been used as the nominal generator for historical reasons: it was the only NLO generator available when I started this Ph.D., Powheg+Herwig7 being available only near the end of my work on this task. Furthermore, Powheg+Pythia8 was used as the nominal generator for the previous η-intercalibration published in [72], following a decision detailed in [103]; thus using it as the nominal in this study simplifies the comparisons of the new results to the previous ones. Finally, having all of the *in situ* calibration steps using Pythia8 as their nominal generator allows to use the reduced flavor uncertainty as described in Section 3.3.1.4, derived using Pythia8. Indeed, the new flavor uncertainties are smaller than the differences between the generators, thus forcing to use the MC-to-MC calibrations described in Section 3.3.1.4 to calibrate the other generators to the energy scale of Pythia8.

4.2.1.1 Selection criteria of events

The quality of the sample of dijet events is ensured by a series of selection criteria (often called "cuts" in ATLAS), summarized in Table 4.2, and successively applied onto data and Powheg+Pythia8 either at reco or truth level in Fig. 4.4. These selection criteria ensure that there is no systematic bias in the sample, and that the physics effects mentioned previously are not dominant (extra radiations like a third jet or out-of-cone radiations).

Detailed description of the event selection.
First, a data event should not be affected by issues in the LHC or in ATLAS. This is ensured by checking that the event has been recorded during a time period (called a "luminosity block") that is listed in a "Good-Runs List" (GRL), and also by removing the events that are corrupted by local problems in the detector (LAr, Tile, Core).

	Data	MC reco	MC truth	Cuts (nominal values)
1	✓			Good-Runs List (GRL) v89 (2015-2016), v99 (2017), v102 (2018)
2	✓			Errors in LAr/Tile/core (corrupted events)
3	✓	✓		Primary vertex with at least two tracks, and at least two track particles
4	✓			Event-level cleaning (BadLoose)
5	✓			Bad batman (for 2015/2016 data only)
6	✓			Triggers (see Section 4.2.1.2)
7	✓	✓	✓	Require at least 2 jets
8	✓	✓		Jet-Vertex Tagger: $JVT > 0.2$ (see Section 3.2.1)
9	✗	✗		Jet-level cleaning (BadLoose)
10	✓	✓	✓	Angular criterion: $\Delta\phi^{1,2} > 2.5$
11	✓	✓	✓	p_T of the third jet: $p_T^{\text{jet }3}/p_T^{\text{avg}} < 0.25$
12		✓		MC cleaning: require at least one truth jet and $p_T^{\text{avg,reco}}/p_T^{\text{jet 1,truth}} \leq 1.4$
(13)	✓	✓	✓	25 GeV $< p_T^{\text{avg}} \leq$ 2000 GeV (see Section 4.2.2)

Table 4.2: Selection. All the selection criteria are shown in order of application for each type of dataset (data, MC reco and MC truth). When a selection is not required for one type of dataset, the associated cell is empty. The check-mark (✓) indicates that the selection is applied, the unchecked-mark (✗) indicates that the selection should be applied but is not. The latter case arises for the jet-level cleaning, that had not been defined for EMPFlow jets at the time when this work was done. The last selection is implicit in the code: the selection is realized when the histograms are filled, because events that do not enter into one p_T^{avg} bin are simply discarded.

Then a few checks are done to remove events that have a high probability not to contain good quality jets: it is required that there is a primary vertex, that it has at least two tracks and at least two track particles. There is also an event-level cleaning using the BadLoose selection, that rejects jets of non-collision origin, described in Section 3.2.2. Finally, the "batman ears" effect is mitigated for 2015/2016 data [120]. It is caused when there is a saturation of the electronics at the end of a bunch train (more precisely, it is caused by the recovery of the electronics from such a saturation), and is identified by a high number of clusters in the $|\eta| \approx 3$ region with $p_T > 500$ MeV and a bad quality flag.

It is required that any data event passes at least one trigger of the chosen "trigger menu", detailed in Section 4.2.1.2, where all the triggers for this analysis are single-jet triggers. This ensures that all data events contain at least one jet at trigger level. Then the event is required to have at least two jets, a basic requirement of the η-intercalibration.

An event for which one of the two main jets has a high probability to be a pile-up jet is rejected with a criterion on the Jet Vertex Tagger (JVT, see Section 3.2.1). Of course this only applies for reco events (either data or MC) because there is no pile-up at truth level, and the values of the selection criteria used are found in Table 4.3a. Note that this selection criterion is only applied on the two leading jets, and only if the jet satisfies the following criteria: 20 GeV $< p_T^{\text{jet}} <$ 60 GeV and $\left|\eta^{\text{jet}}\right| < 2.4$, which is the region where the likelihoods of the JVT have been determined. Pile-up jets outside this region are not rejected.

Arriving at this step, a jet-level cleaning should be applied. However, this was not possible at the time when I conducted this study, since the JetETmiss group of ATLAS had not provided recommendations for jet-level cleaning for EMPFlow jets. This is not a problem, as the event-level cleaning is already applied, and because it would mostly reject out-of-time pile-up, which is not dominant anyways, and which would mostly fall into the tails of the asymmetry distributions. Only the core of the asymmetry distribution is fitted, not the tails, therefore this cleaning is not expected to play an important role for the η-intercalibration in the first place. Such a cross-check should still be made by future analyzers, to be certain that it does not induce any bias.

Then some physics criteria are applied. To illustrate why, let's consider a hypothetical collision that would have only generated two jets for which the energy would have been perfectly measured by ATLAS; those two jets would be perfectly back-to-back in the transverse plane, with equal transverse momentum. However there are usually more than two jets emitted (additional jets, additional particles in general, often at lower energies, or out-of-cone radiations), thus resulting in an imbalance in the dijet system, and finally in a systematic asymmetry. This asymmetry caused by physics effects can be mitigated by selection criteria:[2]

- An angular criterion in the transverse plane is checked, for jets to be fairly back-to-back: $\Delta\phi^{1,2} = |\phi^{\text{jet }1} - \phi^{\text{jet }2}| > \Delta\phi_{\text{min}}$;

[2]The following conventions will be used throughout the document: the jets are ordered by decreasing transverse momentum, 1 being the index of the jet with highest momentum, called the "leading jet", 2 for the "sub-leading jet", 3 for the "sub-sub-leading jet".

- A criterion on the transverse momentum of the third jet (if any) is also applied, relatively to the average transverse momentum of the third jet: $\frac{p_T^{\text{jet 3}}}{p_T^{\text{avg}}} < r^{\text{max}}$, with $p_T^{\text{avg}} = \frac{p_T^{\text{jet 1}} + p_T^{\text{jet 2}}}{2}$.

The values of these selection criteria are found in Table 4.3. These imbalances arise from physics effects, therefore these selections are needed both at reco and truth levels.

One more selection is applied for MC at reco level: the "MC cleaning", that aims at mitigating the mismodeling of pile-up at high p_T. Indeed, the reco MC exhibits high-p_T pile-up jets that are not observed in data. These events come from the way pile-up jets are modeled, as explained in Section 2.2: events following the expected $\langle\mu\rangle$ distribution are overlaid to the hard-scatter objects. Therefore, some high-p_T pile-up jets can be overlaid to lower-p_T events. The reco level is compared to the truth level, and if the two leading jets at reco level are too far from the truth leading jet, then the event is probably a (mismodeled) pile-up event. The criterion is $p_T^{\text{avg,reco}}/p_T^{\text{jet 1,truth}} \leq 1.4$. Around 100 GeV, the Jet Energy Resolution (JER) is of about 10 %, so having both reco leading jets being 4 σ away from the leading truth jet because of a resolution effect is really improbable and indicates a pile-up event. This value of 1.4 for the MC cleaning has been optimized in [121] to have a decent efficiency (93 % in the lowest-p_T slice, see 4.2.1.3, and more than 99 % in the other slices) while removing enough events to have the best agreement between truth and reco levels.[3][4]

A final selection criterion is implicitly applied when creating the histograms: if an event doesn't enter into any bin, *i.e.* if its p_T^{avg} is not within the boundaries of the first and the last p_T^{avg} bins, then the event is simply discarded. The p_T^{avg} binning is described in Section 4.2.2.

Physics selection criteria.
Three of these selection criteria have a strong impact on the calibration: the selections on JVT, on $\Delta\phi^{1,2}$ and on $p_T^{\text{jet 3}}/p_T^{\text{avg}}$. The efficiency of those selection criteria has been studied in [103] to determine the adequate values, as well as ± 1 σ systematic variations of these selection criteria, the values being listed in Table 4.3. Ideally, this study should have been done again, since the calibration used in both cases have changed in the meantime.

The studies I made at truth level, detailed in 4.2.8.6, show that those values of the selection criteria used at truth level allow to correctly take the physics effects into account in most phase space regions. Note that in principle, the values of the selection at truth and reco levels might be different. Here, the same values are used since they are conservative both at truth and reco levels, as the physics effects are expected to be the same at truth and reco levels (which is the condition for physics effects to cancel out in the numerator and in the denominator, as explained in Section 4.1.1).

When using a variation of one of these physics selection criteria, the events selected are different, providing a different calibration in the end. This different calibration is compared to the nominal calibration, and the difference is used as a systematic uncertainty on the nominal calibration (see Section 4.2.8.6).

Cut	Loose	Nominal	Tight
JVT	-	0.2	0.5
$\Delta\phi^{1,2}$	2.3	2.5	2.8
$p_T^{\text{jet 3}}/p_T^{\text{avg}}$	0.35	0.25	0.15

(a) Selections for reco events

Cut	Loose	Nominal	Tight
$\Delta\phi^{1,2}$	2.3	2.5	2.8
$p_T^{\text{jet 3}}/p_T^{\text{avg}}$	0.35	0.25	0.15

(b) Selections for truth events

Table 4.3: Variations of the selection **(a)** for reco events and **(b)** for truth events. The JVT selection is not used for truth events, as there is no pile-up at truth level. The values used for the reco selections and for the truth selections are the same ones, as they are conservative at reco level and shown to adequately capture the physics effects that they are designed to cover at reco level.

[3]There are concerns that the MC cleaning would modify the shape of the pile-up distribution, especially at low-p_T. This should be studied more in depth, especially considering that the MC cleaning value of 1.4 has been optimized in 2012 for older versions of the simulations of ATLAS.

[4]Note that other options to remove those mismodeled pile-up events have been studied, like removing the events for which the reco jets originate from a primary vertex that is too far away from the truth-level primary vertex; this selection criterion yielded worse efficiencies and truth-reco agreements.

4.2.1.2 Triggers

In ATLAS, not all events are recorded: only the events selected by "trigger chains" are recorded, as explained in Section 2.2. The triggering system of ATLAS is composed of two-level trigger chains: the "Level 1" (L1) triggers, hardware-based that only do a partial event reconstruction, and the "High Level Triggers" (HLT), that do a more thorough event reconstruction. Only a fraction of the events passing the physics requirements of the trigger chains are kept. The HLTs are "seeded" by the L1 triggers, meaning that only events passing the physics requirements of the L1 trigger and being randomly chosen by the L1 trigger are considered at HLT level.

4.2.1.2.1 Triggers as event selectors for a given phase space region

Trigger choice. Only "single jet" triggers are used in the η-intercalibration, meaning that the L1 trigger and the HLT only have requirements on the rapidity and transverse momentum of the leading jet of the event, not on the possible other jets of the event. "Dijet triggers" are not used in ATLAS, because the single jet triggers mostly record dijet events anyways. Indeed, thanks to the transverse momentum conservation, the p_T of the jet that was selected by the single jet trigger must be balanced against another object, which is mostly another jet. Therefore, single jet triggers perfectly suit the η-intercalibration analysis.

There are "multijet triggers", having a requirement on the number of jets, "large-R" triggers for large-R jets, and dedicated triggers for objects other than jets; no such triggers are used here, meaning that the events only recorded by those triggers are not considered in this analysis.

Trigger chains: L1, HLT and prescales. For a given single jet trigger chain, composed of an L1 trigger and of an HLT, the decision to keep an event or not is as follows:

1. The L1 trigger partially reconstructs the event, providing "L1 jets", and checks that the leading jet passes the physics requirements: the leading L1 jet must be in the correct pseudorapidity region and have a transverse momentum above a given threshold;
2. If the event passes the physics criteria, the L1 trigger randomly decides to keep it or not, with the probability $\frac{1}{w_{L1}}$ (w_{L1} is called the "prescale" of the L1 trigger);
3. If the event is selected by the L1 trigger (*i.e.* if it passed its physics criteria and was randomly selected), the HLT randomly decides to keep it or not, with the probability $\frac{1}{w_{HLT}}$ (w_{HLT} is called the "prescale" of the HLT). This random selection is done before the next step to save computing power: in case no HLT randomly selects the events, then the event does not move to the next step and is discarded; it is not fully reconstructed, which is a resource-expensive step.
4. If the event is selected by the L1 trigger (physics and prescale) and by the HLT prescale, then the event is reconstructed by the HLT trigger, providing "HLT jets".[5] The HLT then checks that the leading jet passes the physics requirements: as with the L1 trigger, the leading HLT jet must be in the correct pseudorapidity region and have a transverse momentum above a given threshold.

An event selected by a trigger is often said to "have passed the HLT_xxx", in the jargon of ATLAS. This confusingly refers to the whole trigger chain, L1 and HLT, and sometimes just to the HLT of this trigger chain. I tried to keep it clear: I specify "trigger chain" when both the L1 and HLT are concerned.

In the end, only a fraction of the initial events that pass all the physics requirements is kept, which is $\frac{1}{w_{L1}} \times \frac{1}{w_{HLT}}$. The inverse of this fraction, $w_{\text{trigger}} = w_{L1} \times w_{HLT}$, is called the "prescale" of the trigger chain (*i.e.* of the sequence of the L1 trigger and the HLT). It is the weight that should be attributed to an event that passes this trigger chain in order to account for the fact that $w_{L1} \times w_{HLT}$ events have met the physics requirements to pass this trigger but have randomly not been selected. In other words, an event passing this trigger chain "represents" $w_{L1} \times w_{HLT}$ events. For some of these triggers, the prescale is 1, meaning that all the events that pass the selection criteria are kept and that each event kept only accounts for one single event, *i.e.* with a weight of 1. Such a trigger chain is called an "unprescaled" trigger chain.

Physics criteria of the trigger chains and associated prescales. The inclusive cross-section of production of jets falls rapidly with increasing p_T, as shown for instance in Fig. 5.21: for instance, the jet production cross-section decreases by 4 orders of magnitude in the central region and by 6 orders of magnitude in the forward region, when p_T increases from 100 GeV to 1 TeV. The prescales have been chosen accordingly: the triggers keep all the events at very high p_T, but only keep a small fraction of the events at low p_T. Also, the jet production cross-section falls when $|\eta|$ increases, so it is necessary to keep more events in the forward region, thus to have lower prescales in the

[5] Jargon of ATLAS: when an event is randomly selected by the trigger, it is said to be "chosen/selected by the prescale of the trigger"; when it is not selected, it is said to be "rejected by the prescale of the trigger", or to be "prescaled by the trigger".

forward region to have enough statistics. Therefore, there are central trigger chains and forward trigger chains, that have different prescales in order to record more events from the forward region.

The trigger chains in the central region are named "HLT_jxxx" with "xxx" being the p_T criterion at HLT level, and the trigger chains in the forward region are named "HLT_jxxx_320eta490". The physics criteria for an event to be selected by the HLT_jxxx trigger chain (or the HLT_jxxx_320eta490 trigger chain) are as follows:

1. The HLT is seeded by an L1 trigger, meaning that the event must first be selected by an L1 trigger.
   - If xxx $\in \{15, 25, 35\}$, the HLT is seeded by an L1 trigger called L1_RD0_FILLED than has no physics requirement: it selects events randomly, as long as the bunch is filled.
   - If xxx > 35, the HLT is seeded by an L1 trigger that has physics criteria on the L1 jets: there must be one L1 jet in the event that is in the correct η region (central, *i.e.* $\eta^{\text{L1 jet}} < 3.1$, or forward, *i.e.* $\eta^{\text{L1 jet}} > 3.1$), and that passes the corresponding p_T selection ($p_T^{\text{L1 jet}} > p_T^{\text{L1 selection for HLT_jxxx}}$). Values of the p_T criteria of the L1 triggers are listed in the center-right column in Table 4.4.
2. If the event is selected by the L1 trigger (both its physics criteria and its prescale) seeding that HLT, then there must be one HLT jet in the event that passes the physics criteria: it needs to be in the correct η region (central, *i.e.* $\eta^{\text{HLT jet}} < 3.2$, or forward, *i.e.* $\eta^{\text{HLT jet}} > 3.2$) and needs to pass the corresponding p_T selection ($p_T^{\text{HLT jet}} > p_T^{\text{HLT selection for HLT_jxxx}}$). Values of the p_T criteria of the HLTs are listed in the right-most column in Table 4.4.

Central trigger	Forward trigger	L1 p_T criterion	HLT p_T criterion
HLT_j15	HLT_j15_320eta490	0	15
HLT_j25	HLT_j25_320eta490	0	25
HLT_j35	HLT_j35_320eta490	0	35
HLT_j45	HLT_j45_320eta490	15	45
HLT_j60	HLT_j60_320eta490	20	60
HLT_j85	HLT_j85_320eta490	20	85
HLT_j110	HLT_j110_320eta490	30	110
HLT_j175	HLT_j175_320eta490	50	175
HLT_j260	HLT_j260_320eta490	75	260
HLT_j360	HLT_j360_320eta490	100	360
HLT_j380	-	100	380
HLT_j400	-	100	400
HLT_j420	-	100	420

Table 4.4: Values of the p_T selection criteria at L1 level and at HLT level (in GeV), for the central and forward triggers. The trigger HLT_j85 is not used in the η-intercalibration as it has not been studied in [103]. Some other triggers were found to have defects and were not used for the full extent of Run 2: HLT_j55 and HLT_j150 [122]; those triggers are therefore also not used in the η-intercalibration.

The "trigger menu". Not all triggers existed for all years, therefore the "trigger menu" (*i.e.* the choice of the triggers) depends on the year. The triggers used for this study are shown in Table 4.5. The prescales varied during the years (even at the level of the run, to follow the luminosity curve), so equivalent prescales are used for each year. These are different every year for each trigger.

4.2.1.2.2 Determining the turn-on points of combinations of central and forward trigger chains with the emulation and bootstrapping methods

A trigger can be used in a given phase space region if it is unbiased in that region, *i.e.* fully efficient. In a given phase space region, one needs to know which events to use, hence which trigger provides an unbiased decision in this region. The phase space region is expressed using the observables of the η-intercalibration, *i.e.* p_T^{avg} and η.

To know if the events selected by a given trigger chain constitute an unbiased sample in the phase space region, one must determine if this trigger chain takes unbiased decisions in that phase space. This "probed" trigger chain takes unbiased decisions if, on a sample of events that is already known to be unbiased in this region, it selects all of them. If the trigger does not select all the events from this unbiased sample, then using the events selected by this trigger in this phase space region might result in a bias. This is quantified through the "efficiency": the number of events selected by the trigger chain divided by the total number of events in the unbiased sample.

Trigger	2015	2016	2017	2018
HLT_j15	✓	✓	✓	✓
HLT_j25	✓	✓	✓	✓
HLT_j35	✓	✓	✓	✓
HLT_j45	✓	✓	✓	✓
HLT_j60	✓	✓	✓	✓
HLT_j110	✓	✓	✓	✓
HLT_j175	✓	✓	✓	✓
HLT_j260	✓	✓	✓	✓
HLT_j360	✓	✓	✓	✓
HLT_j380	✓	✓		
HLT_j420			✓	✓

(a) Central triggers

Trigger	2015	2016	2017	2018
HLT_j15_320eta490	✓	✓	✓	✓
HLT_j25_320eta490	✓	✓	✓	✓
HLT_j35_320eta490	✓	✓	✓	✓
HLT_j45_320eta490	✓	✓	✓	✓
HLT_j60_320eta490	✓	✓	✓	✓
HLT_j110_320eta490	✓	✓	✓	✓
HLT_j175_320eta490	✓	✓	✓	✓
HLT_j260_320eta490	✓	✓	✓	✓
HLT_j360_320eta490	✓	✓	✓	✓
HLT_j380_320eta490	✓	✓		
HLT_j420_320eta490			✓	✓

(b) Forward triggers

Table 4.5: Triggers used **(a)** in the central region of the detector and **(b)** in the forward region of the detector. A check-mark (✓) indicates that the trigger is used for the given year. A green check-mark indicates that this trigger is the first unprescaled trigger for the given year. A red check-mark indicates that the trigger doesn't exist in the trigger decision tool of ATLAS, which is not a problem since in this case the trigger decision tool always returns `false` (indicating that the event doesn't pass this nonexistent trigger). The j400 trigger is not used in this study, as explained in Section 4.2.2.

The trigger chain might still be usable in the phase space region if its efficiency in that region is above a certain threshold, because then the potential bias is limited. Indeed, no trigger chain reaches an efficiency of 100 % in any phase space region, the threshold is therefore lowered. In the case of the η-intercalibration, the efficiency threshold above which the decision of a trigger chain can be used is set to 99 %. A selection bias of 1 % of the events will (hopefully) not affect the asymmetry distributions, that is determined through a Gaussian fit on the core of the distribution, excluding the tails. With 99.5 % or 99.9 % efficiency thresholds, less events would pass the selection, also possibly resulting in a worse determination of the mean asymmetry: it is a trade-off between systematic effects and statistical uncertainties.

Studying the efficiency of a trigger chain: emulation of the physics decision of the trigger chain. We use full trigger chains, *i.e.* the sequence of an L1 trigger and an HLT. The L1 trigger and the HLT are therefore not studied independently. And since we only want to know if the trigger sequence introduces a bias (or an inefficiency) in its selection based on physics criteria, we do not care about the prescales of this probed trigger chain: we only want to know if the trigger sequence indeed selects 99 % of the events that it is supposed to select. All what the L1 and HLT prescales do here is reducing the statistics[6], not introducing any sort of bias. This has been thoroughly studied in [103] for the η-intercalibration.

In the technical implementation, the "trigbits" associated to the event contain the information "the event passed the physics criterion of the L1 trigger", but do not contain the same information for the HLT trigger (it only contains the information "the event passed the physics criterion of the HLT trigger AND was selected by the prescales"). The physics decision of the HLT trigger therefore has to be emulated. This is doable because the input files (the "dAODs") always contain the HLT jets.[7]

Constituting an unbiased sample for efficiency measurements. To constitute the unbiased sample on which to test the efficiency of the probed trigger in a given phase space region, one can simply use the events selected by another trigger, provided that we already know that this other trigger is unbiased (fully efficient) in this phase space region. This other trigger, already known to be fully efficient in the studied phase space region, is called the "reference trigger". To choose the reference trigger when multiple fully efficient triggers are available for the given phase space region, the one that provides the biggest number of events should be used, to have the best possible statistical uncertainty on the efficiency of the probed trigger.

For the selection of this unbiased sample, one can also add selection criteria of the analysis. Since one will only

[6] "Statistics" is common jargon for "number of events", as it refers to the statistical uncertainty that is inversely proportional to the square root of the number of events.

[7] Note that instead of using the trigbits for the L1 physics decision, one could also emulate the L1 physics decision by using the L1 jets, also available in the dAODs; the trigbits are slightly more reliable because they contain the actual decision of the trigger (there might be any sort of rare technical issue such that a given event might not be marked as passing the L1 trigger physics criteria even though it passes them, and *vice versa*).

use this subset of events in the analysis, it does not add any bias for this analysis (and for this analysis only: a different analysis, using the same observables but different selection criteria, will not be able to use the determined efficiencies, if such criteria are applied, because this would amount to bias the sample for this other analysis). This will reduce the number of events in the sample, but maybe the probed trigger chain has a better efficiency in the given phase space region on that subset of events than on the whole sample. This might therefore make the probed trigger usable in phase space regions where it would not have been usable if its efficiency would have been determined with the whole unbiased sample. In other words, this slightly extends the phase space region where the probed trigger chain can be used, for this analysis only.

The "bootstrapping" method. One could use the minimum bias trigger as the reference trigger, to study the efficiency of all the triggers. This would provide an unbiased sample of events for sure, always independent of all the probed triggers. However, this channel has a low number of events. And because of the jet production cross-sections, most of the jets are produced at low p_T, meaning that the number of high-p_T events selected by the minimum bias trigger is extremely low (if not null).

Therefore, the minimum bias trigger should only be used in phase space regions where no other trigger is available as the reference trigger. Then the triggers for which efficiencies have been determined thanks to the minimum bias trigger can be used as reference triggers in phase space regions where they are fully efficient, because they will have more events than the minimum bias trigger. This is why it is called the "bootstrapping" method[8]: once the efficiency of a trigger is determined, this trigger can be used as the reference to determine the efficiency of another probed trigger. For instance, the trigger HLT_j85 can be used to determine the efficiency of the trigger HLT_j110, as long as HLT_j85 is fully efficient in the phase space region that is looked at. Then HLT_j110 could be used as the reference trigger to study the efficiency of HLT_j175. *Etc.* Triggers are therefore studied one after the other.

Avoiding biases and testing the full trigger chain: the L1 trigger of the reference and probed trigger chains have to be different. To avoid potential biases, the sample of events needs to be obtained by a method that is uncorrelated from the trigger under study. Otherwise, potential biases of the probed trigger might not be detected, and the probed trigger might be deemed fully efficient in regions where it is not. Indeed, the "unbiased" sample is not always fully unbiased. For instance, if the sample comes from another trigger (the reference trigger), then it is likely not fully unbiased, because no trigger is ever fully efficient (except the minimum bias trigger): the reference trigger is used in regions where it is at least 99 % efficient, but not 100 % efficient.

It is therefore important that the trigger chain used to provide events for the unbiased sample, the reference trigger chain, does not contain trigger steps in common with the probed trigger sequence. In particular, the L1 triggers have to be different. If both HLT triggers (the one of the probed trigger and the one of the reference trigger) are seeded by the same L1 trigger, then only the HLT trigger of the probed trigger chain is probed, not the full L1 + HLT sequence.

Formula for the efficiency of a trigger chain in a given phase space region. In a given (p_T^{avg}, η) phase space region where the reference trigger is fully efficient, the efficiency of the probed trigger is given by:

$$\varepsilon_{\text{L1}\cap\text{HLT}}^{\left(p_T^{\text{avg}},\eta\right)}(\text{probe}|\text{ref}) = \frac{N_{\text{events}}\left(\text{ref} \cap \text{probe}_{\text{L1}}^{\text{physics}} \cap \text{probe}_{\text{HLT}}^{\text{physics}}\right)}{N_{\text{events}}(\text{ref})} \tag{4.7}$$

It is even possible to gain in statistics here, by using the events that pass the emulated physics requirements of the reference trigger, not just the events that passed the reference trigger (*i.e.* not just the events that were selected by the prescales of the reference trigger). Indeed, in the given phase space region, we know that the reference trigger is unbiased, therefore an event that has been recorded by another fully efficient trigger and that passes the physics requirements of the reference trigger is not marked as having passed the reference trigger only because it did not pass the prescale of the reference trigger. So adding all those extra events can only increase the statistics, without adding any sort of bias. The formula becomes:

$$\begin{aligned}&\varepsilon_{\text{L1}\cap\text{HLT}}^{\left(p_T^{\text{avg}},\eta\right)}(\text{probe}|\text{ref})\\ &= \frac{N_{\text{events}}\left(\text{ref}_{\text{L1}}^{\text{physics}} \cap \text{ref}_{\text{HLT}}^{\text{physics}} \cap \text{passed any fully efficient trigger in } (p_T^{\text{avg}},\eta) \cap \text{probe}_{\text{L1}}^{\text{physics}} \cap \text{probe}_{\text{HLT}}^{\text{physics}}\right)}{N_{\text{events}}\left(\text{ref}_{\text{L1}}^{\text{physics}} \cap \text{ref}_{\text{HLT}}^{\text{physics}} \cap \text{passed any fully efficient trigger in } (p_T^{\text{avg}},\eta)\right)}\end{aligned} \tag{4.8}$$

[8]Not to be confused with the "bootstrap" method of propagation of statistical uncertainties, that will be used in the cross-section analysis in Section 5.2.3.2; the origin of the terminology is the same though: the expression "to pull oneself up by one's own bootstrap", referring to the self-centered and iterative aspects of the methods.

The caveat of this method is that is uses the emulation of the physics criteria of the reference trigger, and not the real decision of the whole trigger chain, which might be slightly different for instance due to possible bugs (an event would be marked as selected by one of those triggers while actually not passing the physics criteria). Also note that with this method, events that passed trigger chains with very different prescales are used. The normalization to use in this method is not obvious; the inclusion combination method described later in Section 4.2.1.4 cannot be used, as some events might be selected by multiple triggers.

One could improve it even further by having a "maximal" unbiased sample. The improvement is simply to use all the events selected by all the triggers that are fully efficient in the given phase space region; all the fully efficient trigger chains in this region, that are not seeded by the same L1 trigger, can serve as a reference trigger. The formula becomes:

$$\varepsilon_{\mathrm{L1}\cap\mathrm{HLT}}^{\left(p_T^{\mathrm{avg}},\eta\right)}\left(\mathrm{probe}|\left\{\mathrm{ref}\right\}\right)=\frac{N_{\mathrm{events}}\left(\left\{\mathrm{ref}\right\}\cap\mathrm{probe}_{\mathrm{L1}}^{\mathrm{physics}}\cap\mathrm{probe}_{\mathrm{HLT}}^{\mathrm{physics}}\right)}{N_{\mathrm{events}}\left(\left\{\mathrm{ref}\right\}\right)} \tag{4.9}$$

This method also needs proper weighting of the events.

These improvements have not been used yet, because the gain in statistics is not expected to be important and because the weighting is not obvious.

Combinations of central and forward trigger chains. Only combinations of central and forward triggers are fully efficient over the whole η range, as shown in [103]. Also, to study the $|\eta| \approx 3.2$ region, central and forward triggers have to be combined (fluctuations in the angular resolution may move one jet from one region to another one, even though the angular resolution is very good, and even though there is a small overlap between the L1 and HLT triggers of the central and forward trigger chains). Therefore, central and forward triggers have to be studied together.

However, central and forward triggers do not have the same prescales. Different triggers are used in the same phase space region, and therefore correct weights have to be attributed, according to the prescales of the triggers, for the trigger efficiency to be a ratio of actual numbers of events, and so that the asymmetry distributions faithfully represent the asymmetry distribution of the events in that phase space region. Since some events are selected both by the central and the forward triggers of the trigger combination, the weight attributed to dijet events is not straightforward to avoid biases and to have the optimal statistical weight of the events. To that extent, the inclusion method described in Section 4.2.1.4 is used to weight the events of the reference sample, since the reference "trigger" is not any more a single trigger chain but a combination of a central trigger chain and a forward trigger chain. This method requires the emulation of the physics decisions of the reference trigger chains, both of the L1 trigger and of the HLT (or the use of the trigbits for the L1 trigger). Based on Eq. (4.7), the efficiency of the probed central&forward trigger chain combination then becomes:

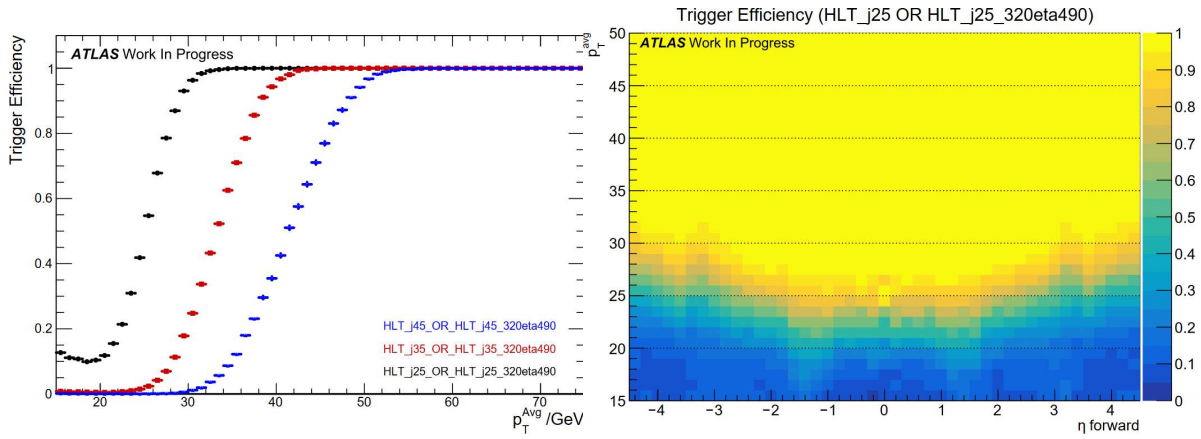
$$\begin{aligned}&\varepsilon_{\mathrm{L1}\cap\mathrm{HLT}}^{\left(p_T^{\mathrm{avg}},\eta\right)}\left(\mathrm{central}\cup\mathrm{forward}|\mathrm{ref}\right)\\&\qquad=\frac{N_{\mathrm{events}}\left(\mathrm{ref}\cap\left(\left(\mathrm{central}_{\mathrm{L1}}^{\mathrm{physics}}\cap\mathrm{central}_{\mathrm{HLT}}^{\mathrm{physics}}\right)\cup\left(\mathrm{forward}_{\mathrm{L1}}^{\mathrm{physics}}\cap\mathrm{forward}_{\mathrm{HLT}}^{\mathrm{physics}}\right)\right)\right)}{N_{\mathrm{events}}\left(\mathrm{ref}\right)}\end{aligned} \tag{4.10}$$

This formula is the one used in [103].

From no on, we are interested in the efficiency of combinations of one central trigger chain and one forward trigger chain, therefore "trigger" will refer to the trigger chain combination (but the reasoning is general).

Resolution effects and trigger turn-on points. The trigger efficiency study only cares about the selection based on physics criteria (the prescales do not matter). The L1 jets are different from the HLT jets, themselves different from the PFlow fully calibrated jets (despite the reconstruction and calibration of HLT jets being close to that of off-line jets nowadays). In particular, their p_T scale and resolution are different. Jets passing the L1 physics selection of a trigger will be reconstructed as HLT jets with a somewhat different p_T: some of them will be reconstructed at HLT level with a lower p_T, some of them with a higher p_T. Then, the HLT selection is applied, removing some of those events. Then the PFlow reconstruction occurs and the full calibration is applied, thus having again a smearing. For instance, some of the events that should have been above the HLT physics selection criteria are missed because they were below the L1 selection criteria. The trigger efficiency study essentially probes the resolution mismatch between L1, HLT and PFlow calibrated jets.

It is not necessary to store the efficiency of a trigger in all phase space regions (unless we want to perform efficiency corrections, which is not the case here): for each η bin, only the p_T^{avg} value above which the trigger is 99 % efficient is necessary. Indeed, in a given η bin and in a given p_T^{avg} bin, if the trigger selects 99 % of the events, then it will select 99 % or more events for all the p_T^{avg} bins above, because the events lost due to the resolution difference



(a) (b)

Figure 4.5: Efficiency of central and forward trigger chain combinations, **(a)** for the trigger combinations j_25, j_35 and j_45, inclusively in η and **(b)** for j_25, binned in η. The trigger efficiency study should be binned in η as the efficiency depends slightly on η, so the turn-on point determined inclusively in η cannot be used, only the maximum per-η turn-on point can be used. Reproduced from [103].

between L1 and HLT are even less numerous. The trigger might be more than 99 % efficient above the threshold, which is good but does not matter. These 99 % efficiency thresholds, obtained per η bin, are called the "turn-on" points: the trigger can be used above, in higher p_T regions, and cannot below. The efficiency curve per η bin looks like the one in Fig. 4.5a from [103], even though this one is inclusive in η.

From these resolution effects and turn-on curves, one can understand how the use of a trigger below its turn-on point would bias the calibration. At HLT level, with HLT jets, the selection is sharp: below the trigger's threshold, there is no jet; the efficiency curve with HLT jets is basically a step. Considering that PFlow calibrated jets and HLT jets are different, there is a resolution smearing when going from HLT jets to PFlow calibrated jets, which is responsible for the shape of the efficiency curve.

When using a trigger in a region where it is not fully efficient, some jets that would have fluctuated up when going from HLT level to PFlow calibrated and that should be present at PFlow calibrated level are absent. Therefore, less low-p_T events are present, thus moving the barycenter of the bin towards higher p_T values: the calibration obtained in this bin is less representative of the low-p_T part of the bin. It also biases the η-intercalibration in a manner that is hard to predict because of the dijet topology; overall, there are less low-p_T events than there should be, thus making the tails of the asymmetry distribution smaller, but one tail might be reduced more than the other one, thus displacing the mean asymmetry.

Single cross-η turn-on point. It was shown in [103] that, for a given trigger, the turn-on point does not depend too much on η, as can be observed in Fig. 4.5b. It is therefore possible to simplify the analysis by only keeping the most conservative turn-on point for the probed trigger, which simply is the highest turn-on point of all the η bins. This makes the analysis much simpler and is conservative, but it looses a bit of statistics because the trigger is not used in some (p_T^{avg}, η) bins where it could be used.

Note: it is not possible to study the turn-on point inclusively in η, contrarily to what Fig. 4.5a shows. The turn-on point is known to be pretty similar across η, but it is not exactly equal in all η bins. Since the inclusive cross-section of jet production falls when $|\eta|$ increases, there are more events in the central region. Hence, if the study was done inclusively in η, the efficiency curve would be dominated by the central region. Then the 99 % efficiency threshold would only be representative of the central region. But the 99 % efficiency threshold of some forward bins might be higher, as is the case shown in Fig. 4.5b. Therefore, the trigger would be used in a forward (p_T^{avg}, η) bin where it is less than 99 % efficient, thus inducing a bias.

Caveats in the use of the turn-on points. The values of the turn-on points are provided without statistical uncertainties, meaning that a safety margin of a few percents of the determined turn-on point has to be taken into account ($5 \sim 10$ GeV were used in [103]).

Triggers could be used below their 99 % efficiency threshold, but the inefficiency should be accounted for and this would represent a lot of work considering that it wouldn't give much more statistics, because the triggers are already well-spaced in p_T^{avg} and satisfactorily provide statistics in all the p_T^{avg} bins, and it might result in an uncontrolled bias in the asymmetry distributions.

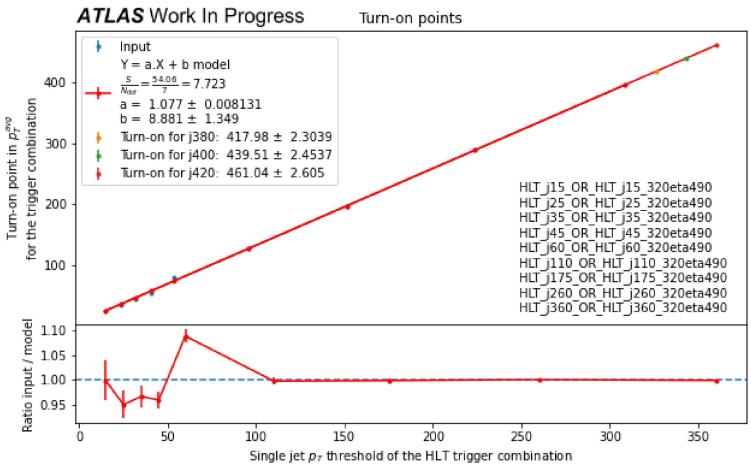


Figure 4.6: 99 % efficiency thresholds in p_T^{avg} for the trigger combinations (noted Y) *vs.* the single-jet p_T cut of the HLT part of the trigger combination (noted X). The values of the thresholds for the three last trigger combinations are extrapolated from the affine fit ($Y = a \cdot X + b$ with a and b the free parameters) of the values for the other trigger combinations. The uncertainties on the input points are not provided in [103], so the minimization is a least squares minimization. This corresponds to a χ^2 minimization by assuming that the uncertainty on the p_T of the threshold of the triggers is null ($\sigma_X = 0$) and $\sigma_Y = 1$ for the p_T^{avg} values. Since this is not a χ^2 minimization but rather a least squares minimization, the sum is called S instead of χ^2 to avoid confusion. The ratio of the input points Y and the affine model Y_{model} is shown in the lower part of the plot; the uncertainties shown here are $\frac{\sigma_Y}{Y_{\text{model}}}$ and give a qualitative perception of the inputs for the fit.

Discussion on the prescales of the low-p_T trigger chains. Anticipating on the discussions of the rest of the chapter, if the prescales for the first triggers (namely HLT_j15) could be lowered for future data-taking periods, it would improve the η-intercalibration. Having at least twice as much statistics recorded in the lowest-p_T^{avg} bins would drastically reduce the statistical uncertainties at low-p_T, as will be discussed in Section 4.2.2, as many $(\eta_{\text{left}}, \eta_{\text{right}})$ bins are lost due to a number of events that is just slightly below the threshold. This would also remove the potential bias at low-p_T that is caused by this lack of events, as discussed later: when the Gaussian fit is too bad, the arithmetic mean of the asymmetry distribution is used, thus maybe causing an uncontrolled bias (see Section 4.2.4.1 for details); this situation occurs very frequently in the first p_T^{avg} bin, and is mostly due to the non-Gaussian shape of the distribution due to the statistical fluctuations. This would overall make the matrix method more robust in the lowest-p_T^{avg} bin.

Summary To sum it up, trigger chains are used to provide events in a given phase space region, when they are known to select events without bias in that region. Only combinations of central and forward trigger chains are fully efficient, therefore only trigger combinations are studied for the η-intercalibration. Therefore, trigger combinations have to be used in p_T^{avg} regions where they are "fully" efficient (99 % efficient in this case): above the determined "turn-on" point (also called "99 % efficiency threshold"). An event selected by a given trigger combination should only be used in a given p_T^{avg} bin if the turn-on point of the trigger is below the lower bound of the p_T^{avg} bin.

4.2.1.2.3 Determining the turn-on points missing from the previous analysis with a fit

The turn-on points in p_T^{avg} above which the triggers are 99 % efficient have been studied in [103], except for three trigger pairs: j_380, j_400, j_420. The values are given in Table 4.6. Ideally, a full trigger efficiency study should be conducted again to determine the turn-on points of the three missing trigger combinations j_380, j_400, j_420. However, due to strong time constraints, I have extrapolated the 99 % thresholds from the values of the turn-on points of the other trigger combinations.

The underlying hypothesis is that the trigger combinations have similar efficiency curves. By taking a quick look at the p_T^{avg} threshold *vs.* the p_T threshold of the triggers, the curve seems affine. An affine fit is performed to extrapolate the values of the thresholds of the remaining triggers, see Fig. 4.6. The values of the 99 % efficiency thresholds, including the extrapolated ones, are given for all the trigger combinations in Table 4.6.

Trigger combination	99 % efficiency threshold in GeV
HLT_j15 or HLT_j15_320eta490	25
HLT_j25 or HLT_j25_320eta490	34
HLT_j35 or HLT_j35_320eta490	45
HLT_j45 or HLT_j45_320eta490	55
HLT_j60 or HLT_j60_320eta490	80
HLT_j110 or HLT_j110_320eta490	127
HLT_j175 or HLT_j175_320eta490	197
HLT_j260 or HLT_j260_320eta490	289
HLT_j360 or HLT_j360_320eta490	396
HLT_j380 or HLT_j380_320eta490	418
HLT_j400 or HLT_j400_320eta490	440
HLT_j420 or HLT_j420_320eta490	461

Table 4.6: Trigger turn-on points, *i.e.* 99 % efficiency thresholds. The values for the trigger combinations j15 to j360 are determined in [103]. The values for the trigger combinations j380, j400 and j420 are extrapolated from the other ones by an affine fit, see Fig. 4.6.

4.2.1.3 MC normalization

MC events have to be weighted so that asymmetry distributions of the MC events are faithful to the real asymmetry distributions.

The production cross-section of jets falls rapidly when p_T^{truth} increases. In order to have enough statistics at high p_T^{truth}, a similar number of events is created in each p_T^{truth} region. Those p_T^{truth} regions are called "JZ slices" in ATLAS [123, 124]. They are defined with a criterion on the truth level leading jet, using anti-k_t with a radius parameter of $R = 0.6$. The slice edges are 20, 60, 160, 400, 800, 1300, 1800, 2500, 3200, 3900, 4600, 5300 GeV. Events are generated either using minimum bias settings of the MC generator (if such settings are provided), or above a certain threshold. Since the minimum bias settings cannot be used at high p_T due to the steeply falling shape of the jet production cross-section, this setting is only used to populate the JZ0 and JZ1 slices, while the production above a threshold is used for all the other slices.

In each phase space region of the η-intercalibration, events from multiple JZ slices are used, and therefore have to be given appropriate weights. For a given JZ slice X, a certain number of events are generated above a threshold (that is chosen to be low enough compared to the lower edge of the JZX slice, to avoid any bias), corresponding to an expected number of events of $\sigma_{JZX} \cdot \mathcal{L}$. Only part of these generated events are kept, because of the $p_T^{\text{leading,truth,JZX}}$ criterion for an event to be part of the JZX slice, thus resulting in an efficiency ε_{JZX}. Therefore, the expected number of events in the JZX slice after selection is $\sigma_{JZX} \cdot \varepsilon_{JZX} \cdot \mathcal{L}$. The MC generator attributes a relative weight w_i to each event. The weight that should be applied to the event i is therefore $W_i \propto w_i$, such that the sum of the weights in the slice corresponds to the expected number of events in the slice, $\sum_{i \in JZX} W_i = \sigma_{JZX} \cdot \varepsilon_{JZX} \cdot \mathcal{L}$, thus resulting in the weight to be applied to the event i:[9]

$$W_i = \sigma_{JZX} \cdot \varepsilon_{JZX} \cdot \mathcal{L} \cdot \frac{w_i}{\sum_{j \in JZX} w_j} \tag{4.11}$$

4.2.1.4 Data normalization

Data events have to be weighted so that asymmetry distributions of the events that are selected by the trigger combinations are faithful to the real asymmetry distributions. This has been shown to be essential in [103].

The η-intercalibration is not a cross-section measurement, therefore the absolute normalization of the events does not matter. In a given phase space region, when events are selected by different triggers that have different prescales, *e.g.* by central or forward triggers, what matters is that they are correctly normalized relatively to each other. The easiest way to achieve this is to normalize the events by the prescale of the trigger that they passed, thus giving them a correct absolute normalization.

But some events are selected both by the central and the forward triggers; in this case, the weight to attribute is not straightforward. Robert Hankache showed in [103] that the correct way to weight the events is by using

[9] Technical notes:
- Only relative weights matter for the η-intercalibration, so $\mathcal{L}$ can be set to 1;
- The i index in the sum runs on all the events generated for the JZX slice, and therefore the normalization has to include all the events *before* any selection criterion is applied.

the "inclusion method", described in [125]. An example of asymmetry distribution using the inclusion weighting method is shown in Fig. 4.9a. In general, it is also possible to attribute a higher weight to events that pass the physics criteria of both the central and the forward triggers, even when the event was not selected by one of those triggers because of the prescale.

The inclusion weighting method When an event is selected by a trigger combination, there is a choice for the decision of the weight that should be attributed to it, depending on which trigger(s) actually selected the event, or which trigger(s) could have selected the event solely based on physics criteria ("raw trigger(s)", also called "emulated trigger(s)").[10]

Let $\mathcal{L}_{\text{trigger}}$ be the luminosity recorded by a trigger and $\mathcal{L}_{\text{total}}$ be the delivered luminosity, *i.e.* the total luminosity that could have been recorded. If a trigger has no prescale (*i.e.* a prescale of 1), the luminosity recorded by that trigger is equal to the total luminosity recorded during that period. Table 4.5 shows the first unprescaled triggers for each year with green check-marks, in the central region for Table 4.5a and in the forward region for Table 4.5b. A prescaled trigger randomly excludes some events that pass the physics criteria and thus that could have been recorded, and therefore records less luminosity[11]. The probability that a trigger records the event is $\frac{1}{w_{\text{trigger}}} = \frac{1}{w_{L1} \times w_{HLT}} = \frac{\mathcal{L}_{\text{trigger}}}{\mathcal{L}_{\text{total}}}$, the prescale of the trigger being w_{trigger}.

Events are classified into the following categories, considering the actual trigger combination that the events passed[12]:

- Central&central if the event only passes the raw central trigger associated to this actual trigger combination. The probability for an event belonging to this category to be recorded is $\frac{1}{w_{\text{central trigger}}}$, so its weight is $w_{\text{central trigger}} = \frac{\mathcal{L}_{\text{tot}}}{\mathcal{L}_{\text{central trigger}}}$.
- Forward&forward if the event only passes the raw forward trigger associated to this actual trigger combination. The probability for an event belonging to this category to be recorded is $\frac{1}{w_{\text{forward trigger}}}$, so its weight is $w_{\text{forward trigger}} = \frac{\mathcal{L}_{\text{tot}}}{\mathcal{L}_{\text{forward trigger}}}$.
- Central&forward if the event passes both the central and forward raw triggers associated to this actual trigger combination. The probability for an event belonging to this category to be recorded is $\frac{1}{w_{\text{central trigger}}} + \frac{1}{w_{\text{forward trigger}}} - \frac{1}{w_{\text{central trigger}}} \times \frac{1}{w_{\text{forward trigger}}}$, so its weight is $\frac{\mathcal{L}_{\text{total}}}{\mathcal{L}_{\text{central trigger}} + \mathcal{L}_{\text{forward trigger}} - \frac{\mathcal{L}_{\text{central trigger}} \times \mathcal{L}_{\text{forward trigger}}}{\mathcal{L}_{\text{total}}}}$.

The $\mathcal{L}_{\text{total}}$ at the numerator is only a normalization and can be omitted, since the absolute normalization of events does not matter for the η-intercalibration; but the one in the right part of the denominator cannot be omitted.

4.2.2 The binning

Depending on the method used to obtain the intercalibration coefficients, either the central-reference method described in Section 4.2.5 or the matrix method described in Section 4.2.6, asymmetry distributions are looked at respectively in bins of $\left(p_T^{\text{avg}}, \eta^{\text{probe}}\right)$ or in bins of $\left(p_T^{\text{avg}}, \eta^{\text{left}}, \eta^{\text{right}}\right)$. Three binnings are therefore necessary:

- A p_T^{avg} binning;
- An η binning, that can be identical for the central-reference and for the matrix method;
- An asymmetry binning.

Note that those binnings cannot be chosen completely independently, as the statistics is the (technical) limiting factor of this analysis: the number of events per bin should be sufficient to not loose too many bins (though there are automatic rebinning strategies; and the purpose of the matrix method is also to have more statistics available).

Those binnings have been determined by previous analyses. The choice of the trigger combinations to use in the different p_T^{avg} bins had however only been performed for the 2015-2016 data-taking period, and I extended it to the 2017 and 2018 data-taking periods, based on the trigger thresholds that I determined with extrapolations in Section 4.2.1.2.3.

[10] Technical note: the information of the triggers that could have been passed by the event, without the prescales, is not in the trigBits, so the Trigger Decision Tool cannot be used and the trigger decision has to be emulated. This is error-prone, and it is observed that some rare events pass an actual trigger combination but do not pass the associated emulated trigger combination; it is not understood why yet. It would be better if the trigger decision tool could provide the information of the triggers passed without taking the prescales into account.

[11] The luminosity recorded by a trigger is determined for each data-taking period using atlas-lumicalc.cern.ch.

[12] Note that here it is not necessarily one of the two leading jets that passes the actual trigger, it could be any other jet in the event. The trigger emulation takes into account all the jets in the event, not only the one that made the event pass the actual trigger, and again not only the two leading jets.

4.2.2.1 The p_T^{avg} binning

The final calibration is produced in $\left(p_T^{\text{probe}}, \eta^{\text{probe}}\right)$ bins. However, the calibration is derived in $\left(p_T^{\text{avg}}, \eta^{\text{probe}}\right)$ bins, and converted at a late stage into the $\left(p_T^{\text{probe}}, \eta^{\text{probe}}\right)$ bins. Indeed, a resolution smearing of the p_T of the probe jet might make the event move from one bin to another one, whereas the p_T^{avg} quantity is a lot more stable to smearing effects, because both jets can fluctuate up or down with almost equal probability.

The p_T^{avg} binning is chosen so that the statistics is satisfying (not too many $(\eta_{\text{left}}, \eta_{\text{right}})$ bins are lost due to lack of statistics, which is mainly happening for the $[25, 40[$ bin both in data and MC, or for the high p_T^{avg} bins in data). It is also chosen to be nearly equally spaced in $\log p_T^{\text{avg}}$. Also, due to the rapidly falling cross-sections, the bins should not be too large, because the barycenters of the bins will be close to the lower edge border anyways. Finally, it is possible to slightly adjust the bin edges to correspond to trigger thresholds, since only one trigger combination is considered per p_T^{avg} bin, as explained later.

The p_T^{avg} binning had been chosen up to 1200 GeV in [106], and bins up to 2000 GeV have been added in [103]. The p_T^{avg} binning is shown in the left column of Table 4.8.

MC

For MC, all the slices are used in all the bins. An event from any slice that belongs to a p_T^{avg} bin will be used in that bin, with the appropriate weight as explained in Section 4.2.1.3.

Data

For most of the p_T^{avg} bins, multiple trigger combinations could be used since they are fully efficient in those bins, provided that they are correctly normalized (which could be done with a generalization of the inclusion method for instance). This is however not necessary, since the lack of statistics is mostly present in the lowest p_T^{avg} bin where there is only one single trigger combination, or in the highest p_T^{avg} bins where the triggers are already unprescaled. It is therefore easier, in each p_T^{avg} bin, to only use events coming from one single trigger combination.

If two triggers have a turn-on point that is lower than the lowest edge of the bin considered, the trigger that has the smallest prescale should be used, to have the best statistics. Indeed, the asymmetry distributions should be similar, but a higher prescale simply means less events. The trigger combinations used for each bin are shown in Table 4.8. The choice of the trigger combinations and the binning can be visualized in figure Fig. 4.7. Note that the p_T^{avg} binning could be re-optimized such that the lower edges of the bins are closer to the turn-on points of the triggers. This would increase the statistics a little bit.

For the $[525, 760[$ GeV bin and all the bins above, all the trigger pairs have their turn-on point below 525 GeV, so all trigger combinations could be used. The trigger combination that has the lowest prescale is, in this case, the lowest unprescaled trigger combination. For 2015 data it is HLT_j360_OR_HLT_j360_320eta490, for 2016 data it is HLT_j380_OR_HLT_j380_320eta490, for 2017 and 2018 data it is HLT_j420_OR_HLT_j420_320eta490. However, since the conditions were very similar in the detector for the data-taking periods 2015 and 2016 (aging, dead modules, *etc*), and since there is not a lot of statistics for the 2015 period, these periods are calibrated together by combining the events of both periods. But the optimal trigger combination to use in the last p_T^{avg} bins is different for the two data-taking periods, and only one of them can be chosen; this should be changed in the code, but has not been done for the sake of time.

For the 2016 period, the HLT_j360_OR_HLT_j360_320eta490 trigger combination has a prescale of 17.8 % (see Table 4.7), so 82.2 % of the highest-p_T^{avg} events would be lost if this trigger combination was used for the 2016 period. Using the HLT_j380_OR_HLT_j380_320eta490 trigger combination for 2015 would result in all its high-p_T^{avg} events being lost. Since there are a lot more events during the 2016 period, it is better to loose the high-p_T^{avg} events of 2015 than 82.2 % of the high-p_T^{avg} events of 2016. So the HLT_j380_OR_HLT_j380_320eta490 trigger combination is chosen for the whole 2015-2016 period.

Note that based on these informations, I could have optimized the choice of the triggers and the p_T^{avg} binning. Due to strong time constraints I did not do it, as it would also potentially have forced me to change the η binning, which would have been a lot of work. The binning chosen in [103] was good enough such that a binning optimization was not necessary, especially considering that the two MC samples used for this study have enough statistics. Indeed:

- The j_45 trigger is never used, but it could be used in the $[60, 85[$ bin, thus increasing the statistics in that bin (and then the j_35 trigger would never be used).
- The trigger j_85, which was not studied in [103], could probably be used in the $[115, 145[$ bin instead of the j_60 trigger.
- The lower bin edges of the p_T^{avg} bins could be optimized to be closer to the trigger turn-on points, to increase the statistics in those bins.

Trigger	2015	2016	2017	2018
HLT_j360	1	0.1783	0.1085	0.0818
HLT_j380		0.9997		
HLT_j400			0.9265	0.6829
HLT_j420			1	1

Table 4.7: Prescales for the central high-p_T^{avg} triggers for the different years. The cell is left empty when the trigger was not used during the data-taking period. Only the central triggers are shown since the corresponding forward triggers either were not used or are not the first unprescaled triggers in the forward region (see Table 4.5a). These values are computed from the luminosities obtained through atlas-lumicalc.cern.ch with $\mathcal{L}_{\text{trigger}}/\mathcal{L}_{\text{total}}$.

p_T^{avg} bin (in GeV)	Trigger combination used
$[25, 40[$	HLT_j15 or HLT_j15_320eta490
$[40, 60[$	HLT_j25 or HLT_j25_320eta490
$[60, 85[$	HLT_j35 or HLT_j35_320eta490
$[85, 115[$	HLT_j60 or HLT_j60_320eta490
$[115, 145[$	
$[145, 175[$	HLT_j110 or HLT_j110_320eta490
$[175, 220[$	
$[220, 270[$	HLT_j175 or HLT_j175_320eta490
$[270, 330[$	
$[330, 400[$	HLT_j260 or HLT_j260_320eta490
$[400, 525[$	HLT_j360 or HLT_j360_320eta490
$[525, 760[$	2015 or 2016 data: HLT_j360 or HLT_j360_320eta490
$[760, 1100[$	
$[1100, 1500[$	2017 or 2018 data: HLT_j420 or HLT_j420_320eta490
$[1500, 2000[$	

Table 4.8: p_T^{avg} binning and trigger combination used in each bin. The triggers used in the last bin vary depending on the year. The HLT_j400_OR_HLT_j400_320eta490 trigger combination is never used because for 2015-2016 it is not the first unprescaled trigger, and for 2017-2018 it is prescaled and both the HLT_j400_OR_HLT_j400_320eta490 and the HLT_j420_OR_HLT_j420_320eta490 (first unprescaled) have their turn-on point in $[400, 525[$ GeV (see Table 4.7).

4.2.2.2 The η binning

Binning for the nominal calibration.
There are η regions in the detector where it is expected for the calibration to have a fine structure: for instance, regions where there is a transition between different parts of the detector that have different materials or granularity, regions where there is dead material (*i.e.* anything that is not a detector, like circuitry or dead calorimeter cells). These regions are shown in Fig. 2.22. Anticipating on the results, the critical regions where sharp variations in the calibration are observed are at $|\eta| \approx 2.5$, which corresponds to the end of the tracker coverage, and at $|\eta| \approx 3.2$, which corresponds to the endcap-forward transition.

It might also be possible to have regions with a fine structure in the calibration without any particular reason; such variations would be particularly important to understand because they might point at problems in the previous steps of the calibration (I did not observe such a situation). Any structure has to be kept, otherwise the non-closure uncertainty will not be negligible; unless there is a reason to smooth the structure away, *i.e.* unless this variation is understood and is nonphysical.

It is more important to have a fine granularity in η than in p_T^{avg}, because the η-intercalibration is the only *in situ* calibration that corrects in η, whereas the other *in situ* calibrations can correct in p_T with a fine granularity. Also, the calibration is known empirically to be smooth in p_T, which therefore requires less granularity in p_T.

The η binning is chosen independently for each p_T^{avg} bin. Indeed, the production cross-section of dijet events falls rapidly when p_T^{avg} increases, and falls when $|\eta|$ increases. There are (p_T^{avg}, η) bins that are simply nonphysical: the $\sqrt{s} = 13$ TeV energy in the center of mass imposes constraints: $p_T < \frac{13 \text{ TeV}}{2 \cosh \eta}$. There are phase space regions where there is not enough statistics. The matrix method uses the η binning both for η_{left} and for η_{right}; in this case, it is necessary for each η bin to be constrained by at least one $(\eta_{\text{left}}, \eta_{\text{right}})$ bin, as will be explained in Section 4.2.6: some $(\eta_{\text{left}}, \eta_{\text{right}})$ bins can be lost due to a lack of statistics, but not too many.

The η binning has been optimized on data from the 2015-2016 period in [103], and then some bins merged

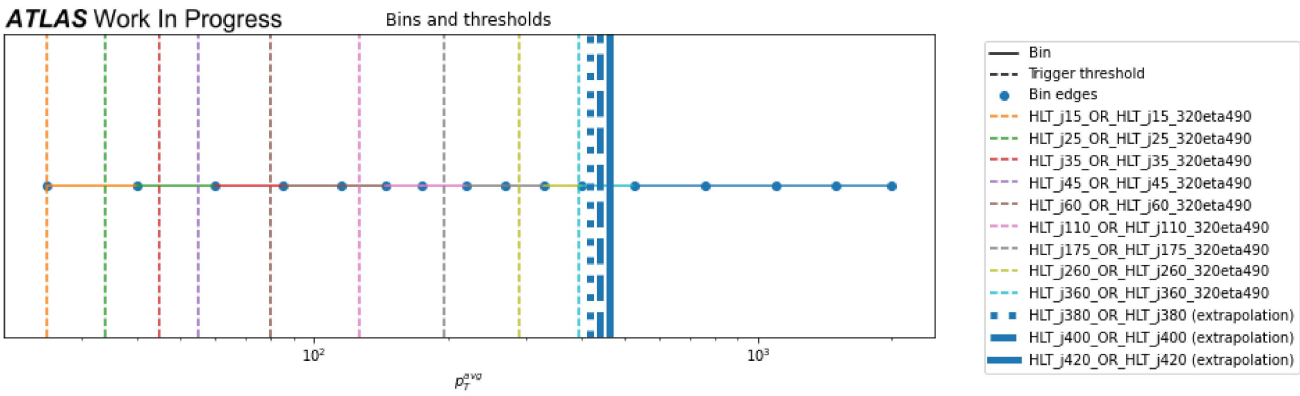


Figure 4.7: The p_T^{avg} bin edges are shown in blue dots, the color of the line between them indicates the choice of the trigger combination, as in Table 4.8. The same color is used for the vertical lines to show the 99 % efficiency thresholds of the trigger combinations, that can be found in Table 4.6. The thick lines on the right indicate the turn-on points that were extrapolated with an affine fit from the values of the other trigger combinations, see Fig. 4.6.

because of a lack of statistics in the MC samples. The (p_T^{avg}, η) binning is shown in Fig. 4.8.

It should be optimized again because there is now more statistics in the MC samples, and to check that it is consistent with the 2017 and 2018 periods. The procedure is as follows: the calibration is derived with a very fine binning (0.05 in η) to catch all the possible structures (even the non anticipated ones). Then, in regions where the fast fluctuations are not statistically significant, bins are clustered together. In the end, in regions where there is no fast variation, bins can be further clustered together, to increase the statistics and to reduce the memory and CPU usage, and to reduce the number of independent statistical uncertainties that need to be propagated in physics analyses in the end.

Binning for the systematic uncertainties.
The η binning for the systematic uncertainties is coarser, as explained in Section 4.2.8.2, because:

- Systematic uncertainties are designed to cover physics effects, and physics is smooth (no fast variations).
- Systematic uncertainties must be statistically significant, and for that reason bigger bins than those of the nominal calibration are necessary.

Therefore, a regular binning of 0.4 in η is used.

4.2.2.3 The asymmetry binning

The asymmetry binning is chosen uniform, with 50 bins in the $[-1.5, 1.5]$ interval, *i.e.* bins of width 0.06. Indeed, as explained in Section 4.2.4, the asymmetry is limited in $[-2, 2]$, is symmetrical around its peak, has rapidly falling tails (especially at high p_T^{avg}), and its peak is centered on 0 if there is no imbalance in the dijet system. The granularity is fine because at high p_T^{avg} the peak is narrow and because there is an automatic rebinning procedure in asymmetry in case the statistics in the asymmetry bins is too low.

4.2.2.4 The final (p_T, η) binning

The calibration points are obtained in a binning that is not regular in p_T and in η, which is the binning described in Section 4.2.2.1 and in Section 4.2.2.2. Also, the nominal calibration and the systematic uncertainties are not produced in the same η binning. Therefore, the final calibration and the uncertainties are produced in another finer binning, regular in p_T and in η (thanks to an interpolation by a Gaussian kernel, see 4.2.7.3):

- The $-4.5 < \eta < 4.5$ range is divided into 450 linearly spaced bins;
- The $20 < p_T < 2000$ GeV range is divided into 100 logarithmically spaced bins.

This final binning is chosen to be at least as granular as the initial (p_T^{avg}, η) binning described in Section 4.2.2.1 and in Section 4.2.2.2.

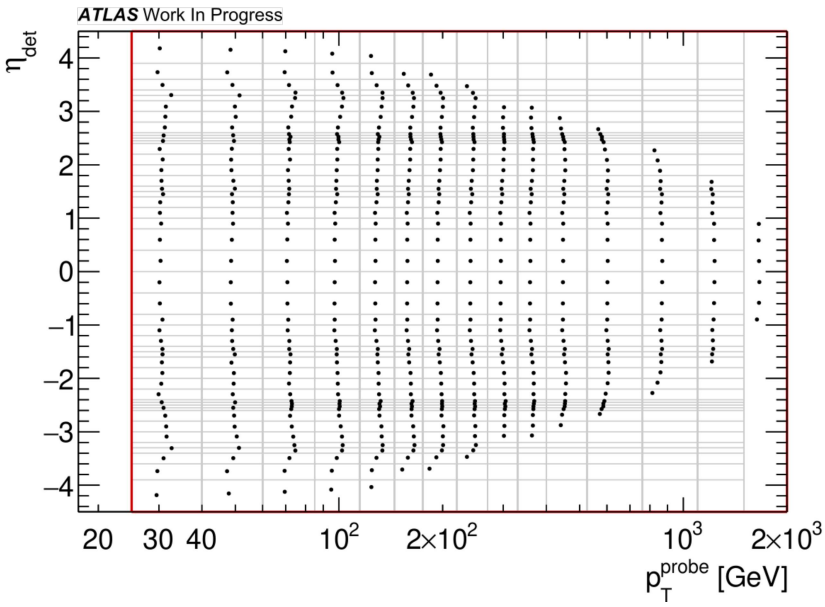


Figure 4.8: (p_T^{avg}, η) binning. The red box indicates the phase space region in which calibration points are obtained. Some areas of this phase space are not covered, either because they are unphysical (the $\sqrt{s} = 13$ TeV energy in the center of mass imposes constraints: $p_T < \frac{13\text{ TeV}}{2\cosh\eta}$) or because there is not enough statistics in those regions. The dots show the barycenter of the bins that are used.

4.2.3 From the asymmetry to the calibration: overview

As a reminder of the introduction, the calibration $\mathcal{C}$ is obtained as the ratio of two intercalibration factors: one of data or MC at reco level at the numerator, and one of MC either at reco or truth level depending on the implementation (relative or absolute), see Section 4.1:

$$\begin{cases} \mathcal{C}^{\text{relative}} = \dfrac{c_{\text{data}}}{c_{\text{MC reco}}} = \dfrac{\left\langle \frac{p_T^{\text{ref}}}{p_T^{\text{probe}}} \right\rangle_{\text{data}}}{\left\langle \frac{p_T^{\text{ref}}}{p_T^{\text{probe}}} \right\rangle_{\text{MC reco}}} \\ \mathcal{C}^{\text{absolute}}_{\text{data-specific}} = \dfrac{c_{\text{data}}}{c_{\text{MC truth}}} = \dfrac{\left\langle \frac{p_T^{\text{ref}}}{p_T^{\text{probe}}} \right\rangle_{\text{data}}}{\left\langle \frac{p_T^{\text{ref}}}{p_T^{\text{probe}}} \right\rangle_{\text{MC truth}}} \\ \mathcal{C}^{\text{absolute}}_{\text{MC-specific}} = \dfrac{c_{\text{MC reco}}}{c_{\text{MC truth}}} = \dfrac{\left\langle \frac{p_T^{\text{ref}}}{p_T^{\text{probe}}} \right\rangle_{\text{MC reco}}}{\left\langle \frac{p_T^{\text{ref}}}{p_T^{\text{probe}}} \right\rangle_{\text{MC truth}}} \end{cases} \tag{4.12}$$

The intercalibration factor c cannot be determined accurately by simply looking at the mean of the $\frac{p_T^{\text{ref}}}{p_T^{\text{probe}}}$ distribution, due to the large transverse momentum resolution of the reference jet. As an intermediate step, the mean asymmetry $\langle \mathcal{A} \rangle = \left\langle \frac{p_T^{\text{probe}} - p_T^{\text{ref}}}{p_T^{avg}} \right\rangle$ is computed in each bin by a Gaussian fit, from which the intercalibration factor is computed as $c = \frac{2-\langle \mathcal{A} \rangle}{2+\langle \mathcal{A} \rangle}$.

Two methods allow to obtain the intercalibration factor: the central-reference method, for which the reference jet is necessarily in the central region of the detector, that compares any region of the detector in which the probe jets falls to the central region; and the matrix method, that uses all dijet systems, regardless of which region the jets fall into. The matrix method is more complicated, as it needs to combine all those constraints to obtain the intercalibration factors, and needs to propagate the statistical uncertainties from the measurement of the mean asymmetry to the intercalibration factors through the constraint combination.

The calibration is then smoothed and interpolated into the final binning, with a Gaussian kernel. Finally, the systematic uncertainties are computed: from datasets either obtained with varied selections or from different MC generators.

4.2.4 The asymmetry

For the central-reference method, the intercalibration factor is computed as

$$c = \frac{2 - \langle \mathcal{A} \rangle}{2 + \langle \mathcal{A} \rangle} \tag{4.13}$$

with the asymmetry being defined as

$$\mathcal{A} = \frac{p_T^{\text{probe}} - p_T^{\text{ref}}}{p_T^{\text{avg}}} \tag{4.14}$$

For the matrix method, the relative response is computed as

$$\mathcal{R} = \frac{2 + \langle \mathcal{A} \rangle}{2 - \langle \mathcal{A} \rangle} \tag{4.15}$$

with the asymmetry being defined as

$$\mathcal{A} = \frac{p_T^{\text{left}} - p_T^{\text{right}}}{p_T^{\text{avg}}} \tag{4.16}$$

The reasoning to explain why the asymmetry is used as an intermediate step is similar for both methods, so only the central-reference method will be described, as it is easier to understand.

Why not directly computing c as the mean of $\frac{p_T^{\text{ref}}}{p_T^{\text{probe}}}$? Why using the mean asymmetry $\langle \mathcal{A} \rangle$?
There are multiple reasons.

The mean of the distribution of $\frac{p_T^{\text{ref}}}{p_T^{\text{probe}}}$ should be defined on a subset of the events in the associated bins: only the core of the distribution should be used, to remove all the events that are not representative of the bin, *i.e.* the tails of the distribution should be removed. There are some rare events that are very asymmetrical, for instance if one of the two jets radiated a lot of its energy into multiple additional jets; those rare events do not respect the transverse momentum conservation at all and are therefore not desired in the determination of the mean asymmetry. The distribution of $\frac{p_T^{\text{ref}}}{p_T^{\text{probe}}}$ is not symmetrical with respect to its peak: in case there is no imbalance on average, the peak is at 1, and the distribution goes from 0 to $+\infty$. This asymmetry makes it difficult to select a criterion on where to cut the tails. Furthermore, the tails of this distribution are falling slowly, and are very different depending on the p_T^{avg} bin, thus making it even harder to have a simple criterion on how to cut the tails. The asymmetry distribution, on the contrary, is very symmetrical around its peak, and has rapidly falling tails since the asymmetry distribution is bounded in the $]-2, 2[$ interval.

The mean of the $\frac{p_T^{\text{ref}}}{p_T^{\text{probe}}}$ distribution is not on the peak of the distribution, due to its asymmetrical shape, thus making its determination difficult, and very dependent on the cuts of the tails. The asymmetry distribution does not have this issue: it is very symmetrical around its peak, due to fluctuations of p_T^{probe} and p_T^{ref} "compensating" on average. This makes the average asymmetry easier to determine, as it can be determined through a Gaussian fit: the mean is simply the position of the peak. The rapidly falling tails also make it easier to use a Gaussian fit. Note that, as mentioned hereafter, the non-Gaussian tails are excluded from the distributions when performing the Gaussian fit or when taking the arithmetic mean. Compared to simply using the arithmetic mean to obtain $\langle \mathcal{A} \rangle$, the Gaussian fit has the important advantage of using shape information of the distribution. This is useful for multiple reasons:

- It provides a χ^2, *i.e.* it quantifies the agreement of the distribution with an expected Gaussian shape.
- It interpolates between the bins and therefore "retrieves" shape information that is lost due to the binning (roughly speaking, it makes a smooth curve out of a staircase). The precision on the determination of the mean asymmetry is at the percent or even permille level, without having bins that are that fine. It also makes the determination of the mean less sensitive to the binning choice.
- It makes the determination of the mean asymmetry less sensitive to rare statistical fluctuations, *e.g.* one asymmetry bin with a strong displacement but a small statistical uncertainty. This has been tested by adding a small bias in one asymmetry bin (not too big or not too statistically significant): the Gaussian fit is less affected and correctly finds the position of the peak.

Finally, the mean of $\frac{p_T^{\text{ref}}}{p_T^{\text{probe}}}$ is sensitive to resolution effects, due to quadratic terms appearing in the TAYLOR expansion with respect to $\left\langle p_T^{\text{probe}} \right\rangle$ and $\left\langle p_T^{\text{ref}} \right\rangle$, see Eq. (4.18). The larger the resolution effects, the farther away the mean of the distribution is from the peak. The resolution is larger at lower p_T values, thus making the

determination of $\left\langle \frac{p_T^{\text{ref}}}{p_T^{\text{probe}}} \right\rangle$ more difficult at lower p_T. On the other hand, the quadratic terms cancel out in the TAYLOR expansion of the asymmetry, see Eq. (4.20), making $\langle \mathcal{A} \rangle$ very robust to resolution fluctuations. The use of p_T^{avg} in the denominator makes the asymmetry more robust to resolution fluctuations.

The second order TAYLOR expansion of $\frac{p_T^{\text{ref}}}{p_T^{\text{probe}}}$ with respect to $\left\langle p_T^{\text{probe}} \right\rangle$ and $\langle p_T^{\text{ref}} \rangle$ is:

$$\frac{p_T^{\text{ref}}}{p_T^{\text{probe}}} \approx \frac{\langle p_T^{\text{ref}} \rangle}{\left\langle p_T^{\text{probe}} \right\rangle} + \frac{1}{\left\langle p_T^{\text{probe}} \right\rangle} \left(p_T^{\text{ref}} - \langle p_T^{\text{ref}} \rangle \right) - \frac{\langle p_T^{\text{ref}} \rangle}{\left\langle p_T^{\text{probe}} \right\rangle^2} \left(p_T^{\text{probe}} - \left\langle p_T^{\text{probe}} \right\rangle \right)$$
$$- \frac{1}{\left\langle p_T^{\text{probe}} \right\rangle^2} \left(p_T^{\text{ref}} - \langle p_T^{\text{ref}} \rangle \right) \left(p_T^{\text{probe}} - \left\langle p_T^{\text{probe}} \right\rangle \right) + \frac{\langle p_T^{\text{ref}} \rangle}{\left\langle p_T^{\text{probe}} \right\rangle^3} \left(p_T^{\text{probe}} - \left\langle p_T^{\text{probe}} \right\rangle \right)^2 \quad (4.17)$$

Hence, taking the mean, we see that the quadratic terms remain:

$$\left\langle \frac{p_T^{\text{ref}}}{p_T^{\text{probe}}} \right\rangle \approx \frac{\langle p_T^{\text{ref}} \rangle}{\left\langle p_T^{\text{probe}} \right\rangle} - \frac{\left\langle p_T^{\text{ref}} p_T^{\text{probe}} \right\rangle}{\left\langle p_T^{\text{probe}} \right\rangle^2} + \frac{\langle p_T^{\text{ref}} \rangle \left\langle {p_T^{\text{probe}}}^2 \right\rangle}{\left\langle p_T^{\text{probe}} \right\rangle^3} \quad (4.18)$$

The second order TAYLOR expansion of $\mathcal{A}$ with respect to $\left\langle p_T^{\text{probe}} \right\rangle$ and $\langle p_T^{\text{ref}} \rangle$ is:

$$\mathcal{A} \approx \frac{\left\langle p_T^{\text{probe}} \right\rangle - \langle p_T^{\text{ref}} \rangle}{\frac{\langle p_T^{\text{ref}} \rangle + \left\langle p_T^{\text{probe}} \right\rangle}{2}} + p_T^{\text{probe}} \left(p_T^{\text{ref}} - \langle p_T^{\text{ref}} \rangle \right) - p_T^{\text{ref}} \left(p_T^{\text{probe}} - \left\langle p_T^{\text{probe}} \right\rangle \right) \quad (4.19)$$

Hence, taking the mean:

$$\langle \mathcal{A} \rangle \approx \frac{\left\langle p_T^{\text{probe}} \right\rangle - \langle p_T^{\text{ref}} \rangle}{\frac{\langle p_T^{\text{ref}} \rangle + \left\langle p_T^{\text{probe}} \right\rangle}{2}} \quad (4.20)$$

$\langle \mathcal{A} \rangle$ is not affected by quadratic terms, *i.e.* by the resolution of the jets.

Relationship between the mean asymmetry $\langle \mathcal{A} \rangle$ and the intercalibration factor c.
The intercalibration factor c is defined as the mean of the $\frac{p_T^{\text{ref}}}{p_T^{\text{probe}}}$ distribution. The relationship with $\langle \mathcal{A} \rangle$ becomes apparent with a TAYLOR expansion of $\frac{p_T^{\text{ref}}}{p_T^{\text{probe}}}$ around $\langle \mathcal{A} \rangle$ at second order[13]:

$$\frac{p_T^{\text{ref}}}{p_T^{\text{probe}}} \approx \frac{2 - \langle \mathcal{A} \rangle}{2 + \langle \mathcal{A} \rangle} - 4 \frac{\mathcal{A} - \langle \mathcal{A} \rangle}{\left(2 + \langle \mathcal{A} \rangle \right)^2} - 4 \frac{\left(\mathcal{A} - \langle \mathcal{A} \rangle \right)^2}{\left(2 + \langle \mathcal{A} \rangle \right)^3} \quad (4.21)$$

and by taking the mean:

$$\left\langle \frac{p_T^{\text{ref}}}{p_T^{\text{probe}}} \right\rangle \approx \frac{2 - \langle \mathcal{A} \rangle}{2 + \langle \mathcal{A} \rangle} + 4 \frac{\langle \mathcal{A}^2 \rangle - \langle \mathcal{A} \rangle^2}{\left(2 + \langle \mathcal{A} \rangle \right)^3} \quad (4.22)$$

with $\langle \mathcal{A}^2 \rangle - \langle \mathcal{A} \rangle^2$ being the variance of the asymmetry distribution (or more precisely: the variance of the part of the asymmetry distribution on which the Gaussian fit is performed). Neglecting the second order term of Eq. (4.22), one computes c as:

$$c := \frac{2 - \langle \mathcal{A} \rangle}{2 + \langle \mathcal{A} \rangle} \quad (4.23)$$

One could measure both the mean and the standard deviation of the asymmetry distribution with the Gaussian fit, not just the mean, and use Eq. (4.22) to compute $\left\langle \frac{p_T^{\text{ref}}}{p_T^{\text{probe}}} \right\rangle$. In principle, this formula is more precise, even though it is subject to the caveats mentioned hereafter. Therefore, not using it results in a bias in the determination

[13]Note that the convergence radius of the TAYLOR expansion of $\frac{p_T^{\text{ref}}}{p_T^{\text{probe}}}$ around $\langle \mathcal{A} \rangle$ is $2 + \langle \mathcal{A} \rangle$, the formula in Eq. (4.21) can be used when $|\mathcal{A} - \langle \mathcal{A} \rangle| < 2 + \langle \mathcal{A} \rangle$, *i.e.* when $\mathcal{A} < 2\langle \mathcal{A} \rangle + 2$. This only restricts its use in a small region; when $\langle \mathcal{A} \rangle$ is close to 0, which is the case for the η-intercalibration, the formula is valid.

of the intercalibration factor c. This bias is non negligible especially in the low p_T^{avg} bins, where the asymmetry distribution is large (of the order of 1 in the lowest p_T^{avg} bin, making the quadratic term of the same order of magnitude as the $\frac{2-\langle\mathcal{A}\rangle}{2+\langle\mathcal{A}\rangle}$ term).

However, this bias is similar in data and in Monte Carlo: the width of the asymmetry distribution is comparable, thus making the bias cancel out (to first order) in the data over MC ratio, when the calibration is computed. Furthermore, if such a bias was not fully canceling in the data over MC ratio, the closure test (detailed in 4.3.1.2) would hopefully detect it, and the bias would then be taken into account in the non-closure uncertainty. Currently, the closure is satisfying.

If we were defining c as $\frac{2-\langle\mathcal{A}\rangle}{2+\langle\mathcal{A}\rangle} + 4\frac{\langle\mathcal{A}^2\rangle-\langle\mathcal{A}\rangle^2}{(2+\langle\mathcal{A}\rangle)^3}$, the drawback would be that the second term is non linear. This would make the propagation of statistical uncertainties non linear, and hence would impose the use of the bootstrap method (described in Section 5.2.3.2) to all physics analyses, thus making them a lot heavier (whereas the statistical uncertainty propagation is linear with $\frac{2-\langle\mathcal{A}\rangle}{2+\langle\mathcal{A}\rangle}$, that is approximately equal to $1-\langle\mathcal{A}\rangle$ at first order). Using this formula would hence need to be strongly justified, for instance if there is a strong methodological bias that is not visible in the closure test (*i.e.* if taking this correction into account changes the calibration but not the closure); or if it allows to drastically reduce the non-closure uncertainty. There are very few precision analyses that use the calibration in the lowest p_T bins and that would therefore benefit from such improvements anyways.

For all these reasons, the quadratic term is not taken into account when defining c.

The approximation of Eq. (4.23) is explicitly justified first in [107] thanks to the following approximation, which proceeds from Eq. (4.20):

$$\frac{2-\langle\mathcal{A}\rangle}{2+\langle\mathcal{A}\rangle} \approx \frac{\langle p_T^{\text{ref}}\rangle}{\left\langle p_T^{\text{probe}}\right\rangle} \tag{4.24}$$

This justification expresses a relationship between $\frac{2-\langle\mathcal{A}\rangle}{2+\langle\mathcal{A}\rangle}$ and $\frac{\langle p_T^{\text{ref}}\rangle}{\left\langle p_T^{\text{probe}}\right\rangle}$, not with $\left\langle\frac{p_T^{\text{ref}}}{p_T^{\text{probe}}}\right\rangle$ (the mean of a ratio is not the ratio of the means). Though valid when neglecting the quadratic terms in Eq. (4.18) or in Eq. (4.22), it is incomplete. Indeed, the quadratic terms, as expressed above, play an important role at low p_T and cannot be neglected; they only cancel out thanks to the data/MC ratio.

$\frac{\langle p_T^{\text{ref}}\rangle}{\left\langle p_T^{\text{probe}}\right\rangle}$ is not an interesting quantity anyways, as the bins in which the mean asymmetry is computed have a finite size. This implies that these quantities, $\langle p_T^{\text{ref}}\rangle$ and $\left\langle p_T^{\text{probe}}\right\rangle$, are not stable to the shape of the PDFs (to the shape of the cross-sections), which fall with increasing p_T and $|\eta|$, contrarily to $\left\langle\frac{p_T^{\text{ref}}}{p_T^{\text{probe}}}\right\rangle$. Furthermore, these quantities are not stable to the change of order at which the matrix element is computed. Therefore, this ratio of means is not stable with respect to the size of the bins.

These new computations presented above, of the relationships between the different quantities at stake, provide a more solid ground for the understanding of the η-intercalibration method.

Statistical uncertainty propagation

For historical reasons, the inverse intercalibration coefficient $\frac{1}{c} = \left\langle\frac{p_T^{\text{probe}}}{p_T^{\text{ref}}}\right\rangle = \frac{2+\langle\mathcal{A}\rangle}{2-\langle\mathcal{A}\rangle}$ is computed instead of the intercalibration coefficient c. Indeed, its meaning would be analogous to a response if the reference jet was calibrated to the truth level; since the reference region is not calibrated (it is calibrated by other *in situ* methods later), and since the goal of the η-intercalibration is to calibrate detector regions with respect to each other and not to the truth level, $\frac{1}{c}$ is often called the "relative response" (and similarly for the matrix method, where $\mathcal{R}$ stands for "relative response"). Using c or $\frac{1}{c}$ is equivalent, the calibration is simply computed as $\mathcal{C} = \frac{\left(\frac{1}{c}\right)_{\text{MC}}}{\left(\frac{1}{c}\right)_{\text{data}}}$ instead of $\frac{c_{\text{data}}}{c_{\text{MC}}}$.

For the propagation of statistical uncertainties from the mean asymmetry to the relative response, one has $\frac{\mathrm{d}\frac{1}{c}}{\mathrm{d}\langle\mathcal{A}\rangle} = \frac{4}{(2-\langle\mathcal{A}\rangle)^2}$ and therefore $\Delta\frac{1}{c} = \frac{4\Delta\langle\mathcal{A}\rangle}{(2-\langle\mathcal{A}\rangle)^2}$.

4.2.4.1 Determining the mean asymmetry in each bin

Two methods are used and compared to extract the mean of the asymmetry distribution: the simple arithmetic mean of the distribution, and a Gaussian fit on the distribution. An example is shown in Fig. 4.9a. Following some

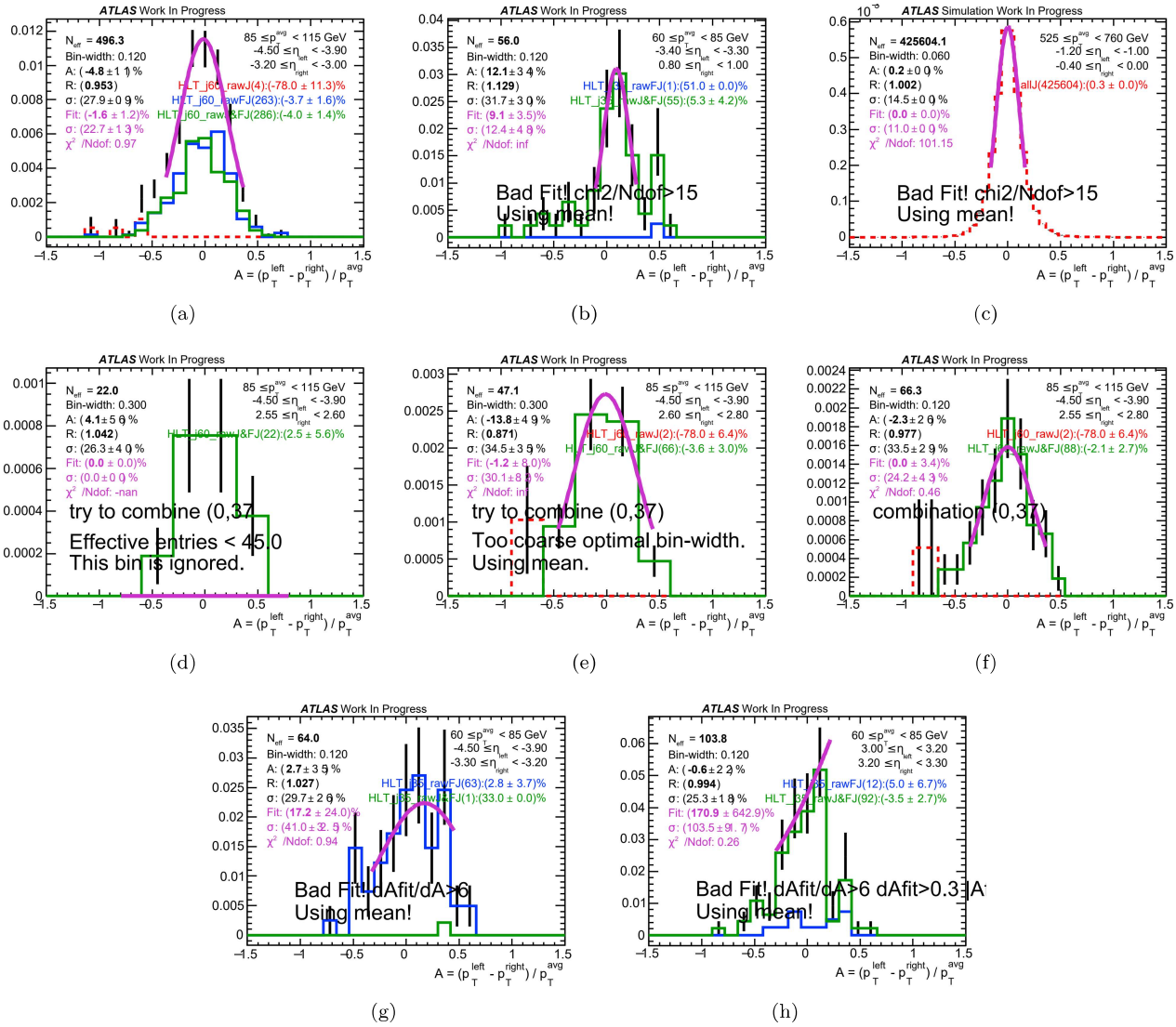


Figure 4.9: Determination of the mean asymmetry on a restricted range, by a Gaussian fit or by the arithmetic mean, in **(a,b,d-h)** 2017 data or **(c)** POWHEG+PYTHIA8 reco (mc16d), in different $(p_T^{\text{avg}}, \eta_{\text{left}}, \eta_{\text{right}})$ bins. N_{eff} is the effective statistics over the full asymmetry range, A is the asymmetry through the arithmetic mean on a restricted range, R is the associated response, "Fit" is the mean asymmetry determined with the Gaussian fit on a restricted range, χ^2/N_{dof} is the χ^2 per degree of freedom of that fit. **(a)** Example of a fit that worked well, where the distribution is the combination of the three categories of data events: central-central (in red, "rawJ"), forward-forward (in blue, "rawFJ") and central-forward (in green, "rawJ&FJ"). **(b)** Example of a fit that did not work because of a bad χ^2/N_{dof} at low p_T^{avg}. **(c)** Example of a fit that did not work because of a bad χ^2/N_{dof} at high p_T^{avg} in MC due to the high effective statistics that make the non-Gaussian tails important, and due to the asymmetry binning that is too coarse. **(d)** Example of a bin that is discarded due to the low effective statistics: $N_{\text{eff}} < 45$. **(e)** Example of a bin that is discarded due to an optimal bin width that is too coarse (more than 0.3). **(f)** Example of a successful bin combination of the distributions shown in **(d)** and **(e)**, where the individual bins would have been lost but the combined bin allows to retrieve a constraint. **(g,h)** Examples of bins for which the Gaussian fit provides a result that is too different from the result of the arithmetic mean, for various reasons, in which case the arithmetic mean is used.

criteria detailed hereafter, the result of one of these two methods is kept, or the bin is discarded. To prevent rare events from affecting the determination of the mean asymmetry, the tails of the distribution are discarded.

Both methods use iterative determination for more robustness: the mean is found in an interval centered around 0, then it is computed again in an interval centered around the previously found mean (since asymmetry distributions are very symmetrical around their mean).

A lot of care is put into making sure that the asymmetry is correctly determined, with a bias as small as possible, which is why the following section is rather technical (I describe everything thoroughly here mostly for documentation purposes). Most of this work has been performed by other people, but I had to ensure the quality of the calibration, meaning that I cross-checked all these individual steps.

Effective sample size.
A first criterion to decide whether to keep or to discard the (p_T^{avg}, η) bin (or the $p_T^{\text{avg}}\eta_{\text{left}}, \eta_{\text{right}}$ bin in the case of the matrix method) is on the effective statistics N_{eff} of the whole histogram. Since the events are weighted, this value is computed with Kish's effective sample size formula [126]:

$$N_{\text{eff}} = \frac{\left(\sum_{i=1}^{n} w_i\right)^2}{\sum_{i=1}^{n} w_i^2} \tag{4.25}$$

where w_i is the weight of the i-th event entering into the histogram. In case of an unweighted histogram, this quantity is equal to the number of entries. For a weighted histogram, this corresponds to the hypothetical number of unweighted entries that the histogram would need to have the same statistical power as this weighted histogram. Note that this is only a qualitative indicator, as it does not account for the fact that the mean asymmetry is only determined onto a reduced interval of this histogram.

If $N_{\text{eff}} < 45$, the bin is discarded. This qualitatively means that there are not enough events for the bin to be statistically significant. This also means that a Gaussian fit would be meaningless: there are not enough events for the distribution to have a clearly defined shape, the result would be very sensitive to statistical fluctuations. This is a somewhat arbitrary criterion, which means that 5 bins with 9 entries per bin is a minimum to perform a Gaussian fit, qualitatively (a Poisson distribution with 9 entries starts looking like a Gaussian).

This criterion mostly rejects low p_T bins or high $|\eta|$ bins (especially bins for which both the left and the right jets are in the forward region), as shown in the response matrix of Fig. 4.10a. An example of a bin that is discarded due to a low effective statistics is shown in Fig. 4.9d.

The iterative procedures to determine the mean asymmetry.
More in details, the iterative process to obtain the mean asymmetry through the arithmetic mean in the bin is as follows:

1. The arithmetic mean in the $[-0.7, 0.7]$ interval is computed and gives $\langle\mathcal{A}\rangle_1$;
2. The arithmetic mean in the $[\langle\mathcal{A}\rangle_1 - 0.5, \langle\mathcal{A}\rangle_1 + 0.5]$ interval is computed and gives $\langle\mathcal{A}\rangle$;
3. The uncertainty on that value $\Delta\langle\mathcal{A}\rangle$ is the root mean square divided by the square root of the effective statistics N_{eff}.[14]

The values 0.7 and 0.5 are empirical.

The iterative process to obtain the mean asymmetry through the Gaussian fit is as follows:

1. The histogram of asymmetry is first rebinned to make the asymmetry bins large enough such that they are statistically significant. The procedure to find the optimal bin width v uses Scott's choice (see [127]), given by the following formula: $v = \frac{3.49\sigma}{\sqrt{N_{\text{eff}}^3}}$, with σ the root-mean-square of the distribution.
2. A Gaussian fit is performed in the $[-0.8, 0.8]$ interval, providing $\langle\mathcal{A}\rangle_1$ and $\sigma_1(\mathcal{A})$
3. A Gaussian fit is performed in the $[\langle\mathcal{A}\rangle_1 - N \times \sigma_1(\mathcal{A}), \langle\mathcal{A}\rangle_1 + N \times \sigma_1(\mathcal{A})]$ interval, providing $\langle\mathcal{A}\rangle_2$ and $\sigma_2(\mathcal{A})$
4. A Gaussian fit is performed in the $[\langle\mathcal{A}\rangle_2 - N \times \sigma_2(\mathcal{A}), \langle\mathcal{A}\rangle_2 + N \times \sigma_2(\mathcal{A})]$ interval, providing $\langle\mathcal{A}\rangle$ and $\sigma(\mathcal{A})$

The values 0.8 and N are arbitrary and may be changed. The enlargement factor N is chosen to be 2 for the two lowest p_T^{avg} bins (where the number of events is the lowest) and 1.5 for the other bins (to better cut the non-Gaussian tails that have more impact on the Gaussian fit since they are more statistically significant, otherwise the χ^2 would be bad and the bin would be lost). I have tested different values of N in $[1.3, 2]$ for different p_T^{avg} bins, the ones that provide the less sparse χ^2 matrices are retained. Examples of χ^2 matrices are shown in Figs. 4.10c and 4.10d.

Note that fitting on a portion of the asymmetry distribution is equivalent to rejecting some rare events, with the selection criterion $\alpha_1 < \mathcal{A} < \alpha_2 \iff \alpha_1 p_T^{\text{avg}} < p_T^{\text{probe}} - p_T^{\text{ref}} < \alpha_2 p_T^{\text{avg}}$, with $(\alpha_1, \alpha_2) \in]-2, 2[^2$ (which in the case

[14] Note that it should only be computed on the restricted interval on which the arithmetic mean is computed, but for now it is computed on all the bins in the histograms (no restriction on the interval).

of the iterative Gaussian fit is $\alpha_1 = \langle\mathcal{A}\rangle_2 - N \times \sigma_2(\mathcal{A})$ and $\alpha_2 = \langle\mathcal{A}\rangle_2 + N \times \sigma_2(\mathcal{A})$). This shows that this selection criterion rejects the events that are very asymmetrical, with respect to the systematic asymmetry in the bin.

The criteria to decide which method to use and if the bin should be kept.
There are multiple criteria taken into account to choose between the arithmetic mean or the Gaussian fit.[15]

If the optimal bin width is larger than 0.3, the arithmetic mean is used. This mostly occurs when the statistics is low, *i.e.* at low p_T and high $|\eta|$. This value is arbitrary, but consistent with the fact that most mean asymmetries are smaller than 0.3, and that the asymmetry interval is $]-2, 2[$, *i.e.* there would only be 13 asymmetry bins, and in this case either the distribution is thin and the Gaussian fit is impossible because it would be performed on a very small number of bins, or the distribution is as wide as the asymmetry interval and the uncertainty on the position of the mean will be as large as the mean because the statistics is low. An example of a bin that is discarded due to an optimal bin width that is too coarse is shown in Fig. 4.9e.

When the fit is considered bad, the arithmetic mean is used. This mostly occurs when the $\frac{\chi^2}{N_{\text{dof}}}$ is big. The arbitrary criterion is if $\frac{\chi^2}{N_{\text{dof}}} > 15$, the fit is considered bad. This arises at high p_T^{avg} in MC where there is a high statistics, either because the non-Gaussian tails enter into the fit and are statistically significant, or because the asymmetry binning is too coarse and the Gaussian fit is performed onto a small number of bins.[16] Examples of bins where the fit is not used due to a bad $\frac{\chi^2}{N_{\text{dof}}}$ are shown in Figs. 4.9b and 4.9c, respectively at low and high p_T^{avg}. This is the cause of the holes in the χ^2 matrix Fig. 4.10d.

The fit is also considered bad in the following cases:

- If $\frac{\Delta\langle\mathcal{A}\rangle_{\text{fit}}}{\Delta\langle\mathcal{A}\rangle_{\text{arithmetic mean}}} > 6$
- If $|\Delta\langle\mathcal{A}\rangle_{\text{fit}}| > 0.3$
- If $|\langle\mathcal{A}\rangle_{\text{fit}}| > 0.3$ and $|\langle\mathcal{A}\rangle_{\text{arithmetic mean}}| < 0.2$

Examples of bins where the fit is not used because of some of these criteria are shown in Figs. 4.9g and 4.9h.

4.2.5 The central reference method to obtain the intercalibration coefficient

The simplest way to make the calibration homogeneous across the whole η range is, as said previously, to use the transverse momentum balance of a dijet system in order to compare the average response of the detector in the central region, chosen as a reference region, to the average response in the other parts of the detector.[17] The central region $|\eta| < 0.8$ is chosen to be the reference region because the response of the detector in this region is known to be close to 1 and because the other *in situ* calibration techniques calibrate jets in the central region only. This means that the η-intercalibration calibrates the forward regions of the detector relatively to the central region. Only dijet events with one jet in the central region are kept for the central reference method (regardless of where the second jet is). The jet in the central region is called the "reference jet", whereas the one in a forward region is called the "probe" jet. The dijet events are separated into the $\left(p_T^{\text{avg}}, \eta^{\text{probe}}\right)$ bins described in Section 4.2.2.

The asymmetry is defined as $\mathcal{A} = \frac{p_T^{\text{probe}} - p_T^{\text{ref}}}{p_T^{\text{avg}}}$. The mean of the asymmetry distribution $\langle\mathcal{A}\rangle_{ki}$ is measured, as described in Section 4.2.4, in every p_T^{avg} bin $k \in [\![1, N]\!]$ (N being the number of p_T^{avg} bins) and in every η^{probe} bin $i \in [\![1, N_k]\!]$ (N_k being the number of η^{probe} bins for the k-th p_T^{avg} bin). The uncertainty on that mean asymmetry is also determined: $\Delta\langle\mathcal{A}\rangle_{ki}$. The inverse intercalibration coefficient (or "relative response") is then computed as follows:

$$\left.\frac{1}{c}\right|_{ki} = \frac{2 + \langle\mathcal{A}\rangle}{2 - \langle\mathcal{A}\rangle} \tag{4.26}$$

and the associated statistical uncertainty is:

$$\Delta\left.\frac{1}{c}\right|_{ki} = \frac{4\Delta\langle\mathcal{A}\rangle_{ki}}{\left(2 - \langle\mathcal{A}\rangle_{ki}\right)^2} \tag{4.27}$$

The central reference method suffers from a lack of statistics, due to the requirement of one of the two leading jets being in the central region. This method is therefore only used as a cross-check for the matrix method.

[15] Note that there are concerns that switching between both methods to determine the mean asymmetry could induce a bias, especially at truth level.

[16] This last issue could be solved by having a more granular asymmetry binning in the highest p_T^{avg} bins for MC. This is technically not easy to do, which is why I have not changed it.

[17] The "central reference method" is also often referred to as the "standard method", the "simple method" or just the "SM".

4.2.6 The matrix method to obtain the intercalibration coefficient

4.2.6.1 Description

The matrix method uses all the dijet events, not only those with one of the two leading jets in the central region. In this method, all regions are calibrated at once with respect to all other regions. Each of the two leading jets of each event serves as a reference jet for the other jet of the event. In other words, the response in one η region i is compared to the response in all the other η regions j such that the response in i is not only obtained with one single measurement as in the central-reference method, but comes from the combination of the comparison to all the regions of the detector. The method is described in [72, 103].

If one of the measurements fails, is inaccurate or imprecise, multiple other measurements are used in the constraint combination to obtain the intercalibration coefficients. This makes the matrix method very robust, while at the same time allowing to have a much better statistical precision than the central-reference method since more events are used.

The relative response.
In a dijet event, each of the two leading jets of each event receives a label "left" or "right" according to the following arbitrary criterion $\eta^{\text{jet left}} < \eta^{\text{jet right}}$, that allows to avoid the double-counting of the events. The dijet events are separated into the $\left(p_T^{\text{avg}}, \eta^{\text{left}}, \eta^{\text{right}}\right)$ bins described in Section 4.2.2.

The asymmetry is defined as $\mathcal{A} = \frac{p_T^{\text{left}} - p_T^{\text{right}}}{p_T^{\text{avg}}}$. The mean of the asymmetry distribution $\langle \mathcal{A} \rangle$ is measured, as described in Section 4.2.4, in every $\left(p_T^{\text{avg}}, \eta^{\text{left}}, \eta^{\text{right}}\right)$ bin: the mean asymmetry $\langle \mathcal{A} \rangle_{kij}$, is determined in the k-th p_T^{avg} bin with $k \in [\![1, N]\!]$ (N the number of p_T^{avg} bins), the i-th η^{left} bin and the j-th η^{right} bin with $(i, j) \in [\![1, N_k]\!]^2$ (N_k the number of η bins for the k-th p_T^{avg} bin). The uncertainty on the mean asymmetry $\Delta\langle \mathcal{A} \rangle_{kij}$ is also determined.

The "relative response" $\mathcal{R}$, which is a generalization of the inverse intercalibration factor $\frac{1}{c}$ that allows to compare any η region of the detector to any other η region of the detector, is then computed as follows:

$$\mathcal{R}_{kij} = \frac{2 + \langle \mathcal{A} \rangle_{kij}}{2 - \langle \mathcal{A} \rangle_{kij}} \tag{4.28}$$

and the associated statistical uncertainty is:

$$\Delta\mathcal{R}_{kij} = \frac{4\Delta\langle \mathcal{A} \rangle_{kij}}{\left(2 - \langle \mathcal{A} \rangle_{kij}\right)^2} \tag{4.29}$$

Combining the relative responses $\mathcal{R}$ to obtain the intercalibration coefficients c.
The relative responses in one p_T^{avg} bin need to be combined to obtain the intercalibration coefficients c in that p_T^{avg} bin. The relative responses act like a set of constraints.

The requirement $\eta^{\text{left}} < \eta^{\text{right}}$ imposes that $i < j$. For the k-th p_T^{avg} bin, the response matrix is therefore upper triangular with an empty diagonal:

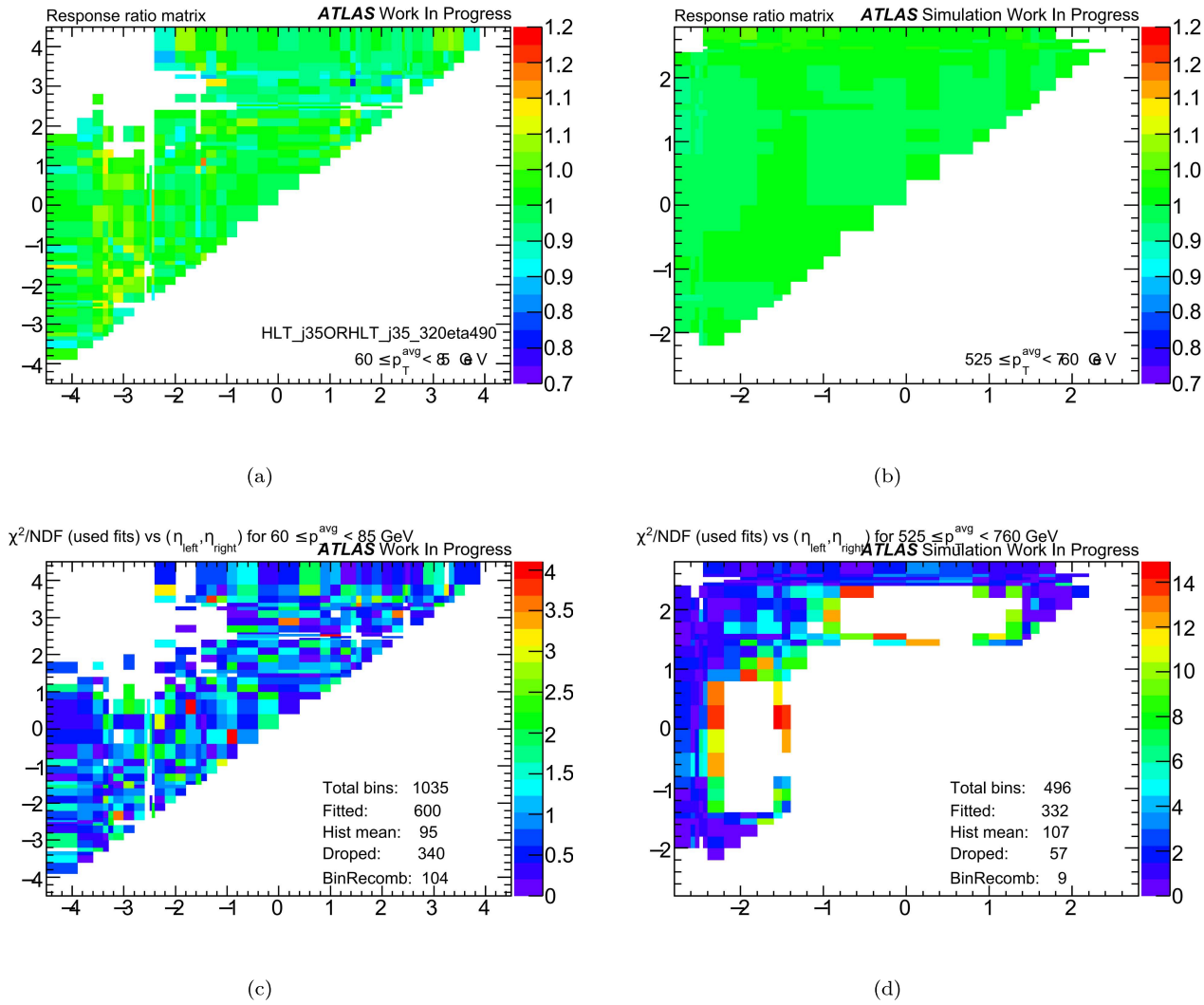
$$\begin{pmatrix} 0 & \mathcal{R}_{k,1,2} & \cdots & \cdots & \mathcal{R}_{k,1,N_k} \\ \vdots & \ddots & \ddots & \mathcal{R}_{k,i,j} & \vdots \\ \vdots & & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & \mathcal{R}_{k,N_k-1,N_k} \\ 0 & \cdots & \cdots & \cdots & 0 \end{pmatrix} \tag{4.30}$$

This response matrix gives its name to the "matrix method". Let's drop the k index for readability, and only consider one p_T^{avg} bin. Examples of response matrices are shown in Fig. 4.10.

To calibrate jets falling into (p_T^{avg}, η) bins, the intercalibration coefficients c are computed by combining the constraints of the relative responses $\mathcal{R}_{ij}$. To first order (*i.e.* by neglecting the quadratic terms), using Eq. (4.20), the following approximate equalities hold true:

$$\mathcal{R} := \frac{2 + \langle \mathcal{A} \rangle}{2 - \langle \mathcal{A} \rangle} \approx \left\langle \frac{p_T^{\text{left}}}{p_T^{\text{right}}} \right\rangle \approx \frac{\left\langle p_T^{\text{left}} \right\rangle}{\left\langle p_T^{\text{right}} \right\rangle} = \frac{\left\langle p_T^{\text{ref}} \right\rangle}{\left\langle p_T^{\text{right}} \right\rangle} \frac{\left\langle p_T^{\text{left}} \right\rangle}{\left\langle p_T^{\text{ref}} \right\rangle} \approx \frac{c^{\text{right}}}{c^{\text{left}}} \tag{4.31}$$

Therefore, the following relation should be verified: $\mathcal{R}_{ij} = \frac{c_j}{c_i}$, which can be rewritten $\mathcal{R}_{ij} c_i - c_j = 0$. There are $\frac{(N_k-1)(N_k-2)}{2}$ such relations (or less if the determination of the mean asymmetry was impossible in some bins),



(a) (b) (c) (d)

Figure 4.10: Examples of **(a,b)** response matrices and **(c,d)** χ^2/N_{dof} matrices, with η^{left} on the x axis and η^{right} on the y axis, for **(a,c)** the 2017 data-taking period, in the $60 \leq \mathrm{p_T^{avg}} < 85$ GeV bin and **(b,d)** for Powheg+Pythia8 reco (mc16d), in the $525 \leq \mathrm{p_T^{avg}} < 760$ GeV bin. The response matrices show the response $\mathcal{R}_{k,i,j}$ in the p_T^{avg} bin k, computed from the mean asymmetry $\langle \mathcal{A} \rangle$, where the mean asymmetry is computed either from the Gaussian fit or from the arithmetic mean. The χ^2/N_{dof} is that of the Gaussian fit: every bin that has an entry in **(c,d)** is a bin where the mean of the Gaussian fit has been kept, and the bins that are filled in **(a,b)** but not in **(c,d)** are those for which the arithmetic mean has been kept. Bins dropped in **(a)** due to the low effective statistics in the forward-forward region. The empty bins of **(d)** are those where the non-Gaussian tails make the χ^2 drop because of the high effective statistics (an example of such an asymmetry distribution is given in Fig. 4.9c); but the comparison with **(b)** indicates that the arithmetic mean is able to retrieve the constraint.

and only N_k coefficients c_i. The system is over-constrained. Therefore we can afford giving more weight to the constraints for which the uncertainty $\Delta\mathcal{R}_{ij}$ is lower.

One way to combine those constraints into a function that will be minimized is a χ^2-like function:

$$S\left(\{c_i\}_i\right) = \sum_{j=1}^{N}\sum_{i=1}^{j}\left(\frac{\mathcal{R}_{ij}c_i - c_j}{\Delta\mathcal{R}_{ij}}\right)^2 \tag{4.32}$$

However, the trivial solution $\forall i, c_i = 0$ exists. Moreover, we know that in the central region $c_i \approx 1$. One therefore adds a Lagrange multiplier to force the average of the intercalibration coefficients to be equal to 1. The function to minimize is therefore

$$S\left(\{c_i\}_i, \lambda\right) = \sum_{j=1}^{N}\sum_{i=1}^{j-1}\left(\frac{\mathcal{R}_{ij}c_i - c_j}{\Delta\mathcal{R}_{ij}}\right)^2 + \lambda\left(\frac{1}{N}\sum_{i=1}^{N}c_i - 1\right)^2 \tag{4.33}$$

The minimization is therefore performed both on $\{c_i\}_i$ and on λ. However, we know empirically that the result does not depend on λ when $\lambda \approx N$: the computation is numerically stable and the constraint $\frac{\partial S}{\partial \lambda} \approx 0$ is satisfied. This only being a regularization term, which goal is only to remove solutions, having the optimal satisfaction of the $\frac{\partial S}{\partial \lambda} = 0$ constraint is not necessary. One could have chosen other regularizations, for instance one could impose mean of the c_i to be equal to 1 only in the central region, instead of all the c_i, but this choice slightly simplifies the calculations for the analytic solution. Therefore, the minimization of S is only performed with respect to $\{c_i\}_i$, with λ chosen equal to N.

4.2.6.2 The analytic solution

An analytic solution has been implemented by Robert Hankache in [103], which is a lot faster than the use of the Minuit minimizer to find the minimum. This is not a piece of work I performed myself. However, it has never been fully documented, so the calculation is shown here.

Let's introduce notations to simplify the computation:

$$X_{i,j}\left(c_i, c_j\right) := \left(\frac{\mathcal{R}_{ij}c_i - c_j}{\Delta\mathcal{R}_{ij}}\right)^2 \tag{4.34}$$

$$S_c := \sum_{\substack{(i,j)\in[\![1,N]\!]^2 \\ i<j}} X_{i,j}\left(c_i, c_j\right) \tag{4.35}$$

$$S_\lambda := \lambda\left(\frac{1}{N}\sum_{i=1}^{N}c_i - 1\right)^2 = \lambda\left(\frac{1}{N^2}\left(\sum_{i=1}^{N}c_i\right)^2 - \frac{2}{N}\sum_{i=1}^{N}c_i + 1\right) \tag{4.36}$$

The constraint with respect to λ is considered already fulfilled, therefore only the minimization with respect to the intercalibration coefficients needs to be computed:

$$\forall k \in [\![1, N]\!], \frac{\partial S}{\partial c_k} = \frac{\partial S_c}{\partial c_k} + \frac{\partial S_\lambda}{\partial c_k} = 0 \tag{4.37}$$

$$\frac{\partial S_c}{\partial c_k} = \sum_{l=1}^{k-1}\frac{\partial X_{l,k}\left(c_l, c_k\right)}{\partial c_k} + \sum_{l=k+1}^{N}\frac{\partial X_{k,l}\left(c_k, c_l\right)}{\partial c_k} \tag{4.38}$$

$$= \sum_{l=1}^{k-1}2\frac{c_k - \mathcal{R}_{lk}c_l}{\Delta\mathcal{R}_{lk}{}^2} + \sum_{l=k+1}^{N}2\frac{R_{kl}\left(R_{kl}c_k - c_l\right)}{\Delta\mathcal{R}_{kl}{}^2} \tag{4.39}$$

$$\frac{\partial S_\lambda}{\partial c_k} = \lambda\left(\frac{2}{N^2}\sum_{l=1}^{N}c_l - \frac{2}{N}\right) = \frac{2\lambda}{N}\left(\frac{1}{N}\sum_{l=1}^{N}c_l - 1\right) \tag{4.40}$$

Putting pieces together:

$$\frac{\partial S}{\partial c_k} = \sum_{l=1}^{k-1}2\frac{c_k - \mathcal{R}_{lk}c_l}{\Delta\mathcal{R}_{lk}{}^2} + \sum_{l=k+1}^{N}2\frac{R_{kl}\left(R_{kl}c_k - c_l\right)}{\Delta\mathcal{R}_{kl}{}^2} + \frac{2\lambda}{N}\left(\frac{1}{N}\sum_{l=1}^{N}c_l - 1\right) \tag{4.41}$$

and rearranging the terms:

$$\frac{\partial S}{\partial c_k} = 2\sum_{l=1}^{k-1}\left(-\frac{\mathcal{R}_{lk}}{\Delta\mathcal{R}_{lk}{}^2}+\frac{\lambda}{N^2}\right)c_l + 2\left(\sum_{m=1}^{k-1}\frac{1}{\Delta\mathcal{R}_{mk}{}^2}+\sum_{m=k+1}^{N}\frac{\mathcal{R}_{km}^2}{\Delta\mathcal{R}_{km}{}^2}+\frac{\lambda}{N^2}\right)c_k + 2\sum_{l=k+1}^{N}\left(-\frac{\mathcal{R}_{kl}}{\Delta\mathcal{R}_{kl}{}^2}+\frac{\lambda}{N^2}\right)c_l - \frac{2\lambda}{N} \tag{4.42}$$

Satisfying the constraints $\forall k \in [\![1,N]\!], \frac{\partial S}{\partial c_k} = 0$ gives a system of equations equivalent to the following matrix equation:

$$\underbrace{\begin{pmatrix} \sum_{m=2}^{N}\frac{\mathcal{R}_{1m}^2}{\Delta\mathcal{R}_{1m}{}^2}+\frac{\lambda}{N^2} & \cdots & -\frac{\mathcal{R}_{1l}}{\Delta\mathcal{R}_{1l}{}^2}+\frac{\lambda}{N^2} & \cdots & -\frac{\mathcal{R}_{1N}}{\Delta\mathcal{R}_{1N}{}^2}+\frac{\lambda}{N^2} \\ \vdots & & \vdots & & \vdots \\ -\frac{\mathcal{R}_{1k}}{\Delta\mathcal{R}_{1k}{}^2}+\frac{\lambda}{N^2} & \cdots & \sum_{m=1}^{k-1}\frac{1}{\Delta\mathcal{R}_{mk}{}^2}+\sum_{m=k+1}^{N}\frac{\mathcal{R}_{km}^2}{\Delta\mathcal{R}_{km}{}^2}+\frac{\lambda}{N^2} & \cdots & -\frac{\mathcal{R}_{kN}}{\Delta\mathcal{R}_{kN}{}^2}+\frac{\lambda}{N^2} \\ \vdots & & \vdots & & \vdots \\ -\frac{\mathcal{R}_{1N}}{\Delta\mathcal{R}_{1N}{}^2}+\frac{\lambda}{N^2} & \cdots & -\frac{\mathcal{R}_{lN}}{\Delta\mathcal{R}_{lN}{}^2}+\frac{\lambda}{N^2} & \cdots & \sum_{m=1}^{N-1}\frac{1}{\Delta\mathcal{R}_{mN}{}^2}+\frac{\lambda}{N^2} \end{pmatrix}}_{U} \underbrace{\begin{pmatrix} c_1 \\ \vdots \\ c_l \\ \vdots \\ c_N \end{pmatrix}}_{C} = \underbrace{\frac{\lambda}{N}\begin{pmatrix} 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \end{pmatrix}}_{B} \tag{4.43}$$

The general term of the matrix $U = (u_{kl})_{(k,l)\in[\![1,N]\!]^2} \in \mathcal{M}_N(\mathbb{R})$ can be expressed:

$$\forall (k,l) \in [\![1,N]\!]^2, u_{kl} = -\frac{\mathcal{R}_{lk}}{\Delta\mathcal{R}_{lk}{}^2}\mathbb{1}_{l<k} + \left(\sum_{m=1}^{k-1}\frac{1}{\Delta\mathcal{R}_{mk}{}^2}+\sum_{m=k+1}^{N}\frac{\mathcal{R}_{km}^2}{\Delta\mathcal{R}_{km}{}^2}\right)\mathbb{1}_{l=k} - \frac{\mathcal{R}_{kl}}{\Delta\mathcal{R}_{kl}{}^2}\mathbb{1}_{l>k} + \frac{\lambda}{N^2} \tag{4.44}$$

Note that for all k, the coefficients $\{u_{kl}\}_l$ and b_k can be multiplied by any coefficient $\alpha_k \in \mathbb{R}^*$. This can be used for numerical stability, and was studied by Robert Hankache. This expression of the system is the one kept in the end, and currently used.

Then the solution of $UC = B$ is simply $C = U^{-1}B$, that provides all the intercalibration coefficients c_k.

Case where some $(\eta_{\text{left}}, \eta_{\text{right}})$ bins are lost.
For some $(\eta_{\text{left}}, \eta_{\text{right}})$ bins, the determination of the relative response $\mathcal{R}_{ij}$ (and its statistical uncertainty $\Delta\mathcal{R}_{ij}$) is impossible, as explained in Section 4.2.4.1. The response matrix is therefore not completely filled, some entries are missing, as shown in Fig. 4.10a. This amounts to the loss of a few constraints. The bins for which the constraints remain are therefore the subset

$$F \subset \left\{(i,j) \in [\![1,N]\!]^2 \middle| i<j\right\} \tag{4.45}$$

The function to minimize is then:

$$S^F\left(\{c_i\}_i, \lambda\right) = \sum_{(i,j)\in F}\left(\frac{\mathcal{R}_{ij}c_i - c_j}{\Delta\mathcal{R}_{ij}}\right)^2 + \lambda\left(\frac{1}{N}\sum_{i=1}^{N}c_i - 1\right)^2 \tag{4.46}$$

and then the calculation gives:

$$\frac{\partial S^F}{\partial c_k} = \sum_{\substack{l\in[\![1,k-1]\!] \\ (l,k)\in F}}\frac{\partial X_{l,k}(c_l,c_k)}{\partial c_k} + \sum_{\substack{l\in[\![k+1,N]\!] \\ (k,l)\in F}}\frac{\partial X_{k,l}(c_k,c_l)}{\partial c_k} + \frac{\partial S_\lambda}{\partial c_k} \tag{4.47}$$

$$= \sum_{l=1}^{k-1}\frac{\partial X_{l,k}(c_l,c_k)}{\partial c_k}\mathbb{1}_{(l,k)\in F} + \sum_{l=k+1}^{N}\frac{\partial X_{k,l}(c_k,c_l)}{\partial c_k}\mathbb{1}_{(k,l)\in F} + \frac{\partial S_\lambda}{\partial c_k} \tag{4.48}$$

$$= \sum_{l=1}^{k-1}2\frac{c_k - \mathcal{R}_{lk}c_l}{\Delta\mathcal{R}_{lk}{}^2}\mathbb{1}_{(l,k)\in F} + \sum_{l=k+1}^{N}2\frac{R_{kl}\left(R_{kl}c_k - c_l\right)}{\Delta\mathcal{R}_{kl}{}^2}\mathbb{1}_{(k,l)\in F} + \frac{2\lambda}{N}\left(\frac{1}{N}\sum_{l=1}^{N}c_l - 1\right) \tag{4.49}$$

then imposing $\forall k \in [\![1, N]\!], \frac{\partial S^F}{\partial c_k} = 0$ and rearranging the terms:

$$\forall k \in [\![1, N]\!], \sum_{l=1}^{k-1} \left(-\frac{\mathcal{R}_{lk}}{\Delta\mathcal{R}_{lk}{}^2} \mathbb{1}_{(l,k)\in F} + \frac{\lambda}{N^2} \right) c_l + \left(\sum_{m=1}^{k-1} \frac{1}{\Delta\mathcal{R}_{mk}{}^2} \mathbb{1}_{(m,k)\in F} + \sum_{m=k+1}^{N} \frac{\mathcal{R}_{km}^2}{\Delta\mathcal{R}_{km}{}^2} \mathbb{1}_{(k,m)\in F} + \frac{\lambda}{N^2} \right) c_k + \sum_{l=k+1}^{N} \left(-\frac{\mathcal{R}_{kl}}{\Delta\mathcal{R}_{kl}{}^2} \mathbb{1}_{(k,l)\in F} + \frac{\lambda}{N^2} \right) c_l = \frac{\lambda}{N} \quad (4.50)$$

and the general term of the matrix U is:

$$\forall (k,l) \in [\![1, N]\!]^2, u_{kl} = -\frac{\mathcal{R}_{lk}}{\Delta\mathcal{R}_{lk}{}^2} \mathbb{1}_{l<k} \mathbb{1}_{(l,k)\in F} + \left(\sum_{m=1}^{k-1} \frac{1}{\Delta\mathcal{R}_{mk}{}^2} \mathbb{1}_{(m,k)\in F} + \sum_{m=k+1}^{N} \frac{\mathcal{R}_{km}^2}{\Delta\mathcal{R}_{km}{}^2} \mathbb{1}_{(k,m)\in F} \right) \mathbb{1}_{l=k} - \frac{\mathcal{R}_{kl}}{\Delta\mathcal{R}_{kl}{}^2} \mathbb{1}_{l>k} \mathbb{1}_{(k,l)\in F} + \frac{\lambda}{N^2} \quad (4.51)$$

4.2.6.3 Automatic rebinning in η

Sometimes, some relative responses $\mathcal{R}_{ij}$ are lost, as explained in Section 4.2.4.1, mostly due to a lack of statistics. This makes the events entering those bins unused, thus losing useful statistics (N_{eff} too low or an optimal asymmetry bin-width that is too coarse), especially in the low-p_T region, and in the forward-forward region. Since the matrix method relies on combining constraints from different regions of the detector to obtain the intercalibration coefficients, having less constraints also makes the method less robust.

It is possible to still use those events by merging adjacent $(\eta_{\text{left}}, \eta_{\text{right}})$ bins: the increased statistics might make the resulting bin usable, which allows to obtain constraints for the initial bins that would otherwise be lost. One caveat is that those constraints obtained for the initial bins are statistically fully correlated, and thus some care is taken at the minimization step.

This work has been performed by Robert Hankache and never documented anywhere (even inside ATLAS), which is why I document it here based on the information found in the software.

One makes the assumption of linearity: the two initial bins contribute to the final merged bin proportionally to their effective statistics. This assumption means that two adjacent bins should have a similar behavior, *i.e.* that their asymmetry peaks are not too far away, such that the combined bin does not exhibit two peaks in the asymmetry distribution but only one. This can be checked *a posteriori* by looking at the asymmetry distributions in the merged bins.

It also assumes that events in both initial bins have the same statistical power, *i.e.* that $N_{\text{eff}}^{\{i,i+1\}} = N_{\text{eff}}^{i} + N_{\text{eff}}^{i+1}$, whereas in general $N_{\text{eff}}^{\{i,i+1\}} \leq N_{\text{eff}}^{i} + N_{\text{eff}}^{i+1}$ holds true. This assumption is not bad: in MC, the events in one p_T^{avg} bin mostly come from the same JZ slice, and in data the events in two adjacent bins most frequently come from the same trigger, and therefore have similar weights in general. It happens for some rare events to have very big weights, thus destroying the effective statistics of the bin; when combining two bins with one including such an event, the combined bin will also have its statistical power destroyed by that event, and therefore the combined bin will (most probably) be rejected with the same criterion that initially rejected the bin.

This assumption allows to express the constraint linearly as follows, either for the combination of two bins i and $i+1$ (with $1 \leq i$ and $i+1 < j$):

$$X_{(i,i+1),j}\left(c_i, c_{i+1}, c_j\right) = \left(\frac{\frac{N_i c_i + N_{i+1} c_{i+1}}{N_i + N_{i+1}} \mathcal{R}_{(i,i+1),j} - c_j}{\Delta\mathcal{R}_{(i,i+1),j}} \right)^2 \quad (4.52)$$

instead of

$$X_{i,j} + X_{i+1,j} = \left(\frac{c_i \mathcal{R}_{ij} - c_j}{\Delta\mathcal{R}_{ij}} \right)^2 + \left(\frac{c_{i+1} \mathcal{R}_{i+1,j} - c_j}{\Delta\mathcal{R}_{i+1,j}} \right)^2 \quad (4.53)$$

and similarly for the combination of two bins j and $j+1$ (with $i < j$ and $j+1 \leq N$):

$$X_{i,(j,j+1)}\left(c_i, c_j, c_{j+1}\right) = \left(\frac{c_i \mathcal{R}_{i,(j,j+1)} - \frac{N_j c_j + N_{j+1} c_{j+1}}{N_j + N_{j+1}}}{\Delta\mathcal{R}_{i,(j,j+1)}} \right)^2 \quad (4.54)$$

Then the analytic solution can be found, similarly to the reasoning done in Section 4.2.6.2, by:

1. Adding those terms $X_{(i,i+1),j}\left(c_i, c_{i+1}, c_j\right)$ and $X_{i,(j,j+1)}\left(c_i, c_j, c_{j+1}\right)$ in the function to be minimized S;
2. Computing the partial derivatives of $X_{(i,i+1),j}\left(c_i, c_{i+1}, c_j\right)$ and $X_{i,(j,j+1)}\left(c_i, c_j, c_{j+1}\right)$ with respect to c_i, c_{i+1}, c_j, c_{j+1};
3. Rearranging the terms to express the matrix U.

The full computation is shown in Appendix A.5.

One choice is left: which two bins should be combined? When one bin is lost, and when the 4 neighboring bins are also lost and also qualify as potentially giving fruitful combinations, then only one of the 4 possible combinations can be kept. Then, when one combination is made, it "blocks" other combinations for other bins; some combinations might be better than other ones, for instance in the gain of statistical uncertainty after the minimization. There is probably an optimal solution to this problem, but that is most certainly hard to find.

Robert Hankache made a choice based on the technical simplicity of implementation. Bins are looked at one by one in a certain order anyways. When two consecutive (and adjacent) bins are lost, it is checked whether their combination passes the selection criteria or not; if yes, then the combination is kept and enters the minimization. This way, not all possible bin combinations are tested, therefore the solution is clearly sub-optimal, but some constraints are recovered which was the initial goal. An example of two bins that are combined in one is shown in Figs. 4.9d to 4.9f.

Instead of simply combining two adjacent bins, one could imagine combining three or more bins in a row or column of the response matrix, or small blocks (for instance 2×2 blocks). Though possible in theory, this faces technical difficulties (the computations become a lot harder, and some choices should be made on which combinations to take), and it not necessarily desirable: indeed, the point of the η-intercalibration is to be granular enough to be sensitive to sharp structures in the calibration; if too many bins are combined, then the sharp structures might disappear. Furthermore, when combining bins that are too far apart, the linearity assumption might become harder to verify, as some bins might be affected by very different physics or detector effects, and therefore the asymmetry distribution in the combined bin might have multiple peaks.

4.2.6.4 Treatment of the central region

With the minimization of the Lagrange parameter $\frac{\partial S_\lambda}{\partial \lambda} = 0$ fulfilled, the average of all the intercalibration coefficients is equal to 1. However, the η-intercalibration is used to calibrate the forward regions with respect to the central region. This correction being relative (one region with respect to another one), the intercalibration coefficients are unique up to a multiplicative factor (in other words, the Lagrange parameter could have set the mean of the intercalibration to any real value, not just 1), thus a choice can be made of which region will receive a correction of 1; such a region would effectively only be corrected by the next steps of the *in situ* calibration chain. The central region $|\eta| < 0.8$ is chosen to receive an average calibration of 1, as the calibration is known to be rather flat in this region.

Therefore, the average of the intercalibration coefficients of the central region is set to 1 thanks to a multiplicative factor applied to all intercalibration coefficients.

4.2.6.5 Propagation of the statistical uncertainties

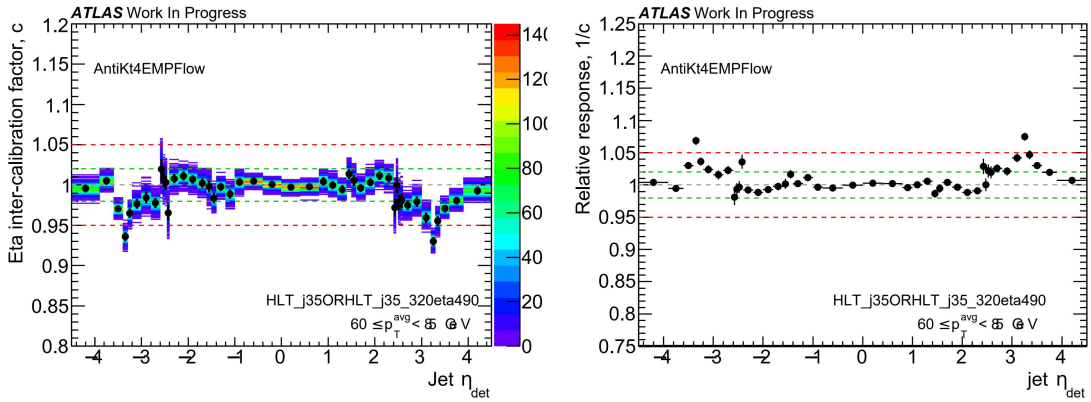
The minimization of S provides the central values of the intercalibration coefficients $\{c_{kl}\}$. The propagation of the statistical uncertainty on $\langle \mathcal{A} \rangle_{kij}$, $\Delta\langle \mathcal{A} \rangle_{kij}$, to $\mathcal{R}_{kij}$ is done linearly with Eq. (4.29) to obtain $\Delta\mathcal{R}_{kij}$ in each individual bin. However, the propagation of the statistical uncertainty from the relative responses $\{\mathcal{R}_{kij}\}$ to the intercalibration coefficients $\{c_{kl}\}$ is not done analytically as the computations are significantly harder than the computations above: it is done with "pseudo-experiments", also referred to as "replicas" or "toy experiments".[18]

For each p_T^{avg} bin k, a replica of the response matrix $\mathcal{R}_k$ is another response matrix $\mathcal{R}_k^r$ that is statistically equivalent to $\mathcal{R}_k$ and obtained by random Gaussian fluctuations around $\mathcal{R}_k$. $N_r = 1000$ such replicas are created. More precisely, this means that

$$\forall r \in [\![1, N_r]\!], \forall k \in [\![1, N]\!], \forall (i,j) \in [\![1, N_k]\!]^2, \begin{cases} \mathcal{R}_{kij}^r \hookrightarrow \mathcal{G}\left(\mathcal{R}_{kij}, \Delta\mathcal{R}_{kij}\right) \\ \Delta\mathcal{R}_{kij}^r = \Delta\mathcal{R}_{kij} \end{cases} \tag{4.55}$$

Then each of these replicas $\mathcal{R}_k^r$ is used with the "matrix method", *i.e.* with the minimization procedure on S, to obtain "replica" intercalibration coefficients $\{c_{kl}^r\}$. Finally, the statistical uncertainty on the intercalibration

[18] In ATLAS, this post-fit method to linearly propagate the statistical uncertainties is referred to as the "post-fit" method, sometimes also as the "simple method". The "bootstrap" method, described in Chapter 5, Section 5.2.3.2 is not used in the η-intercalibration.



(a) (b)

Figure 4.11: Example of the propagation of the statistical uncertainties with the linear post-fit method: each set of intercalibration coefficients c shown in **(a)** is obtained with the matrix method from a replica of the response matrix. The 1000 replicas are shown, and the density of replicas is indicated by the color scale on the Z axis. The Root Mean Square (RMS) of the replicas in each intercalibration bin gives the statistical uncertainty on the nominal value, shown in **(b)**.

coefficients is the root mean square of those replicas:

$$\forall k \in [\![1, N]\!], \forall l \in [\![1, N_k]\!], \Delta c_{kl} = \sqrt{\frac{1}{N_r} \sum_{r=1}^{N_r} {c_{kl}^r}^2} \tag{4.56}$$

The replicas are shown in Fig. 4.11a, and the resulting statistical uncertainty is shown in Fig. 4.11b.

4.2.7 From the intercalibration coefficients to the calibration

4.2.7.1 Ratio of intercalibration coefficients

The calibration is obtained in each (p_T^{avg}, η) bin (k, l) as the ratio of two intercalibration coefficients, as shown in Eq. (4.12):

$$\mathcal{C}_{kl} = \frac{c_{kl}^{\text{reco (data or MC)}}}{c_{kl}^{\text{MC (reco or truth)}}} \tag{4.57}$$

To simplify the explanations, the case of the relative implementation will be taken for the rest of the chapter, without loss of generality: with data at the numerator and MC reco at the denominator (unless otherwise noted).

As already stated, for historical reasons, the inverse intercalibration factor $R = \frac{1}{c}$ is used in computations, and what is shown in figures is $\mathcal{C}_{kl}^{\text{relative}} = \frac{R_{kl}^{\text{MC reco}}}{R_{kl}^{\text{data}}}$.[19] Examples are shown in Section 4.3, for instance in Fig. 4.18 where the upper part of the plots shows the inverse intercalibration coefficients and the lower part shows the calibration taken from the ratio of the inverse intercalibration coefficients.

The statistical uncertainty on the calibration in the (p_T^{avg}, η) bin (k, l) is:

$$\Delta\mathcal{C}_{kl} = \mathcal{C}_{kl} \sqrt{\left(\frac{\Delta R_{kl}^{\text{data}}}{R_{kl}^{\text{data}}}\right)^2 + \left(\frac{\Delta R_{kl}^{\text{MC reco}}}{R_{kl}^{\text{MC reco}}}\right)^2} \tag{4.58}$$

as both quantities are assumed to be statistically uncorrelated. This assumption is true when the numerator and the denominator are respectively data and MC (or an MC generator and another MC generator). However, it is not true in the case of the MC-specific calibration of the absolute implementation, that uses the same sample of events either at reco or at truth level. In this case, the two quantities cannot be assumed to be fully correlated either, as the selection criteria might not have kept the same sub-sample of events at reco and truth levels for

[19] For the central-reference method, it may happen that the determination of the intercalibration coefficient c was impossible in one bin, either for data or for MC (either the nominal MC generator or the alternative MC generator when studying the MC modeling uncertainty, see Section 4.2.8.7). In this case, the ratio cannot be taken, which means that the bin is lost, even if the determination of c worked in this bin for the other dataset(s). The calibration point is also lost when $c < 0$, or when the uncertainty on $\frac{1}{c}$ is bigger than 0.3.

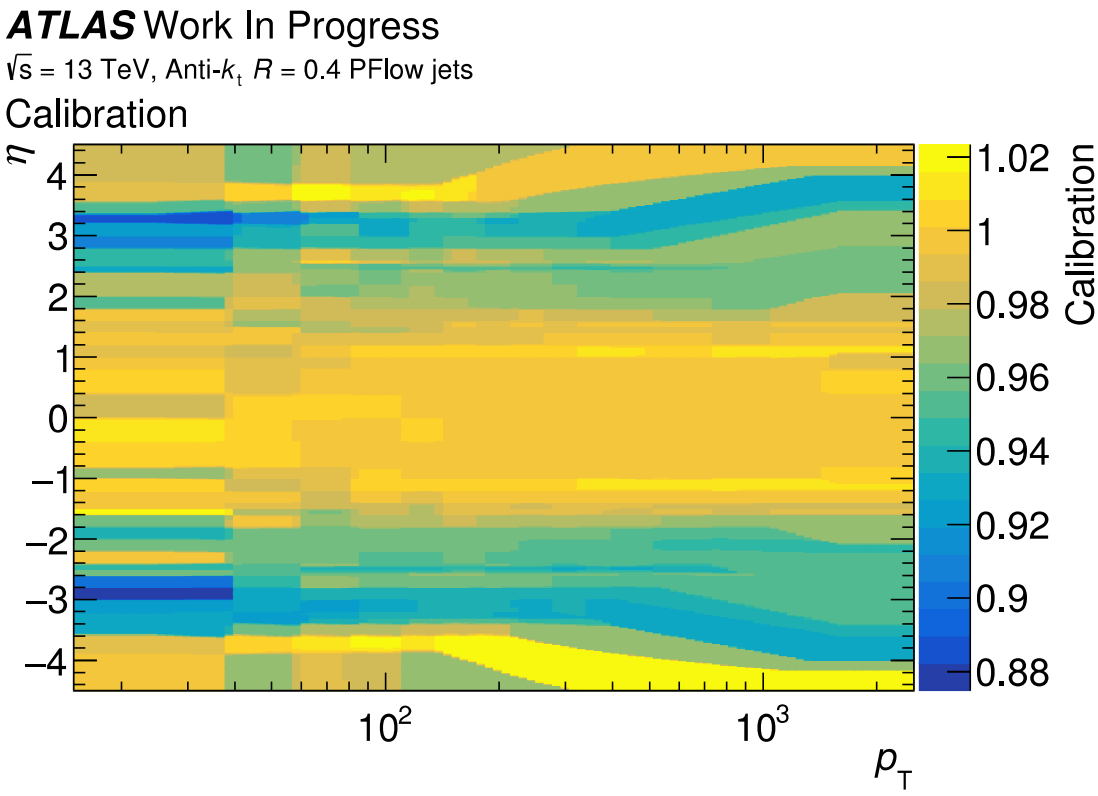


Figure 4.12: Full calibration (relative implementation), after smoothing and freezing. It is also displayed per rapidity or p_T slice in Section 4.3.

the determination of the mean asymmetry. Ideally, the bootstrap method should have been used to determine the statistical uncertainty on the MC-specific calibration for the absolute implementation, but I have not had enough time to make it work. Therefore, I used the formula above in this case: the statistical uncertainty is not optimal but is at least not underestimated.

4.2.7.2 From a p_T^{avg} binning to a p_T binning

At this stage, the calibration is obtained in (p_T^{avg}, η) bins. However, the calibration is not only applied to dijet events that have one jet in the central region, it should be applied onto any jet, regardless of the complete topology of the event. Therefore, the calibration should be obtained in (p_T, η) bins.

To know the correspondence between the two binnings, the following reasoning is used. If an uncalibrated jet with transverse momentum p_T was part of a dijet event with the associated jet in the central region of the detector, this associated reference jet would have the transverse momentum $p_T^{\text{ref}} = p_T \times c = \frac{p_T}{R}$. The p_T^{avg} of this dijet event would then be:

$$p_T^{\text{avg}} = \frac{p_T + p_T^{\text{ref}}}{2} = p_T \frac{1+R}{2R} = \frac{p_T}{sf} \tag{4.59}$$

with the scale factor $sf = \frac{2R}{1+R}$. This scale factor sf is used to move from p_T^{avg} to p_T or *vice versa*.

4.2.7.3 Smoothing

The calibration is binned, which means that there are sharp borders which are unphysical and would cause issues if used in a physics analysis. In order to have a smooth and continuous calibration, a smoothing is applied. This is done in two steps. The smoothed calibration is shown in Fig. 4.12.

Calibration points at the barycenters of the (p_T, η) bins.
The binned calibration is first expressed as a collection of N points $\left\{\left(\overline{p_T^{\text{probe}}}_i, \overline{\eta}_i, \mathcal{C}_i\right)\right\}_i$, with associated statistical uncertainties. Indeed, the calibration obtained in a given (p_T^{avg}, η) bin i is the most representative of the barycenter

of that bin, which is $\left(\overline{p_T^{\text{avg}}}_i, \overline{\eta}_i\right)$. The way the barycenters of the bins are computed is described in Appendix A.4. The scale factor sf_i is used to move from $\overline{p_T^{\text{avg}}}_i$ to $\overline{p_T^{\text{probe}}}$: $\overline{p_T^{\text{probe}}}_i = \overline{p_T^{\text{avg}}}_i \times sf_i$.

Note that the barycenter of the bin is computed both for data and for MC reco, but only the one of data is kept for the calibration point, as the calibration is to be applied onto data. In general, for the absolute calibration, the barycenter of the quantity at the numerator of Eq. (4.12) is kept, *i.e.* the barycenter of the bins of the reco quantity (either data or MC reco).

Gaussian smoothing.
As a second step, the calibration function $F(p_T, \eta)$ is constructed as the combination of all those N points, weighted by a two-dimensional Gaussian kernel:

$$F(p_T, \eta) = \frac{\sum_{i=1}^{N} \mathcal{C}_i w_i(p_T, \eta)}{\sum_{j=1}^{N} w_j(p_T, \eta)} \tag{4.60}$$

with the weights:

$$w_i(p_T, \eta) = \frac{1}{\Delta\mathcal{C}_i^2} \text{Gaus}\left(\frac{\log p_T - \log \overline{p_T^{\text{probe}}}_i}{\sigma_{\log p_T}} \oplus \frac{\eta - \overline{\eta}_i}{\frac{2\Delta\eta_i}{r}}\right) \tag{4.61}$$

with $\oplus$ denoting here the quadratic sum ($a \oplus b = \sqrt{a^2 + b^2}$), Gaus is the non-normalized Gaussian centered on 0 of variance 1: $\forall x \in \mathbb{R}, \text{Gaus}(x) = \exp\left(-\frac{1}{2}x^2\right)$, $\Delta\eta_i$ is the width in η of the bin i, and $\sigma_{\log p_T}$ and r are two free parameters.

This function performs smoothing, interpolation and extrapolation of the calibration at the same time. $\sigma_{\log p_T}$ and $\frac{2\Delta\eta_i}{r}$ control the width of the Gaussian for the smoothing. I tuned the two free parameters $\sigma_{\log p_T}$ and r empirically:

- In the case of the smoothing of the nominal calibration, the structure in η (and in p_T) should not be smoothed out, thus the smoothing is chosen to be "weak", otherwise the non-closure uncertainty would be significant. This is quantified by looking at the local χ^2, that evaluates the tension between the calibration point in a bin and the smoothed function in that bin. I used the values: $\sigma_{\log p_T} = 0.05$ and $r = 6$.

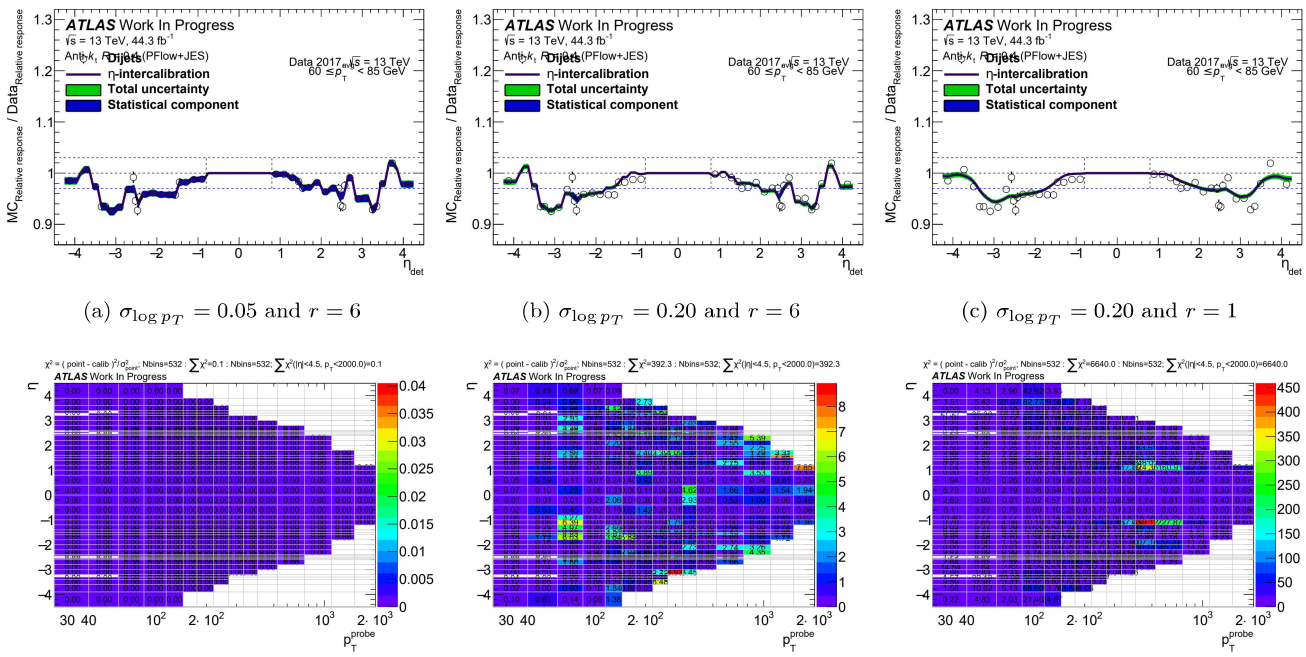
  Those new smoothing parameters are shown in Fig. 4.13 and compared to those of the previous calibration of [72]: the agreement, observed by eye with the calibration, and quantified with the χ^2, is much better. Using a lower value of $\sigma_{\log p_T}$ or a higher value of r does not improve the agreement when looking by eye for instance in the region $|\eta| \approx 2.5$, and with these values the χ^2 is at an acceptable level (it is negligible almost everywhere). Examples of the final smoothed calibration are shown in Section 4.3, for instance in the lower panels of Fig. 4.19: the red triangles are the calibration points, and the red curve is the smoothed calibration.
- In the case of the systematic uncertainties that are designed to cover for some physics effects, the fast fluctuations are expected to only be of statistical origin as the physics effects are smooth. Therefore, fluctuations are smoothed out by using a strong smoothing. I used the values: $\sigma_{\log p_T} = 0.25$ and $r = 1$. Examples of smoothed systematic uncertainties are shown in Section 4.3, for instance in Fig. 4.20.

The calibration points obtained in granular bins have a high statistical uncertainty, and therefore do not have a high weight in the combination due to the $\frac{1}{\Delta\mathcal{C}_i^2}$ factor. In order to preserve the sharp η structure of the calibration, the bin width $\Delta\eta_i$ is taken into account, such that the small bins have a locally stronger influence.

Extrapolation and freezing at high p_T.
The final smoothed calibration is produced with a granular binning in the range $-4.5 < \eta < 4.5$ and $20 < p_T < 2000$ GeV. Some of these bins are outside the range of the calibration points from which the calibration function F of Eq. (4.60) is derived. The Gaussian smoothing naturally allows to perform an extrapolation to these bins. However, one should care about using this extrapolation in regions where it is meaningful.

In the previous production of the η-intercalibration in [72], it had been decided to freeze the calibration in the bins above $p_T = 700$ GeV, because of a lack of statistics. This made the non-closure uncertainty high at high p_T. Since I did not have this constraint, I froze the calibration in the bins above the highest p_T calibration point, *i.e.* above $p_T^{\text{freeze}} = 1651$ GeV.

Note that the calibration is extrapolated down to 20 GeV. This is done for consistency with the other *in situ* calibrations, that go as low as 20 GeV, even though the η-intercalibration does not provide any calibration point that low (the lowest calibration point of the η-intercalibration is close to 25 GeV). Therefore, physics analyses that use the calibration below 25 GeV have to take extra care and to add an *ad hoc* uncertainty in this region.



(a) $\sigma_{\log p_T} = 0.05$ and $r = 6$

(b) $\sigma_{\log p_T} = 0.20$ and $r = 6$

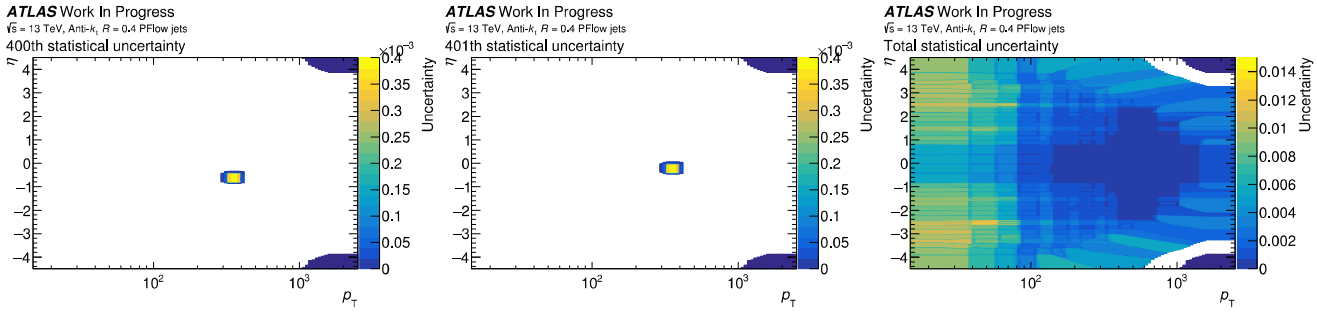
(c) $\sigma_{\log p_T} = 0.20$ and $r = 1$

(d) $\sigma_{\log p_T} = 0.05$ and $r = 6$

(e) $\sigma_{\log p_T} = 0.20$ and $r = 6$

(f) $\sigma_{\log p_T} = 0.20$ and $r = 1$

Figure 4.13: Examples of smoothing parameters on **(a-c)** the nominal calibration in the $25 \leq \mathrm{p_T} < 40$ GeV bin, and **(d-f)** the tension between the calibration point in a bin and the smoothed function in the same bin. **(a,d)** New smoothing parameters for the nominal calibration: $\sigma_{\log p_T} = 0.05$ and $r = 6$; the smoothed calibration is close to the calibration points, and the χ^2 is negligible almost everywhere and never exceeds 0.04 in any region. **(b,e)** Almost the same smoothing parameters as those of the previous calibration [72]: $\sigma_{\log p_T} = 0.20$ and $r = 6$ ($\sigma_{\log p_T}$ was 0.18); the smoothing is a bit too strong in p_T: the smoothed function is clearly far from the points in $1 < |\eta| < 1.6$ and in $2.4 < |\eta| < 2.8$, and the χ^2 shows a clear tension. **(c,f)** Almost the smoothing used for the systematic uncertainties: $\sigma_{\log p_T} = 0.20$ and $r = 1$ ($\sigma_{\log p_T}$ is 0.25 for the systematic uncertainties); this only serves as an example to give an idea of how much the systematic uncertainties are smoothed, where the goal is to get rid of detector-induced structures to only keep the (smooth) physics effects.



(a) (b) (c)

Figure 4.14: Statistical uncertainties. Each (p_T^{avg}, η) bin of the analysis comes with a statistical uncertainty, independent of that of the other bins and very localized. Due to the Gaussian smoothing, these uncertainties "spread" and overlap, as shown by the two independent statistical uncertainties in **(a,b)**. **(c)** Sum in quadrature of all the individual components.

4.2.8 Uncertainties

4.2.8.1 Statistical uncertainties on the calibration

The statistical uncertainties are propagated through the same Gaussian smoothing as the nominal one: a similar equation as Eq. (4.60) propagates the statistical uncertainty of the calibration point i:

$$F^{\text{stat i}}(p_T, \eta) = \frac{\Delta\mathcal{C}_i w_i(p_T, \eta)}{\sum_{j=1}^{N} w_j(p_T, \eta)} \tag{4.62}$$

with $w_i(p_T, \eta)$ the Gaussian weights of Eq. (4.61).

Since one calibration point contributes to multiple final bins due to the Gaussian smoothing, its statistical uncertainty also contributes to the statistical uncertainty of all those final bins, as shown in Figs. 4.14a and 4.14b. Therefore, there are as many statistical uncertainties as the number of calibration points: with the binning I used, this produces 530 statistical uncertainties for each data-taking period (2015+2016, 2017, 2018), *i.e.* 1590 statistical uncertainties in total for the full Run 2. All those individual statistical uncertainties can be summed in quadrature to make only one component per data-taking period, shown in Fig. 4.14c, by assuming no statistical correlations:

$$F^{\text{stat total}}(p_T, \eta) = \bigoplus_{i=1}^{N} F^{\text{stat i}}(p_T, \eta) = \frac{\sqrt{\sum_{i=1}^{N} \left(\Delta\mathcal{C}_i w_i(p_T, \eta)\right)^2}}{\sum_{i=1}^{N} w_i(p_T, \eta)} \tag{4.63}$$

This total statistical uncertainty loses the statistical correlations between the multiple components, which is why it is not used in precision measurements like the jet cross-section measurements (see Chapter 5).

4.2.8.2 Systematic uncertainties

The purpose of the η-intercalibration is to correct any residual systematic deviation of the detector response in the forward region with respect to the detector response in the central region, *i.e.* systematic deviations that would not already be corrected by the previous calibration steps. However, the detector effects that we want to correct are not the only reason for the calibration to deviate from unity. As said, physics effects can also cause an asymmetry in the dijet events, and therefore any mismodeling of the physics effects would also cause the calibration to systematically deviate from unity. Measures are taken to mitigate the physics effects:

- The effect of the additional radiations, and specifically of the third jet if any, are mitigated by the selection criteria on $\frac{p_T^{\text{jet 3}}}{p_T^{\text{avg}}}$ and $\Delta\phi^{1,2}$.
- The effect of pile-up is mitigated by a rejection of pile-up events thanks to the selection criterion on JVT.
- The effect of a mismodeling of physics is mitigated by choosing the MC generator that has the best agreement with data.

These measures are optimized while taking other criteria into account; for instance, the selection criteria remove some events, but the tightest values possible for those selection criteria might remove too many events and degrade the statistical uncertainty too much or even make the determination of the mean asymmetry impossible in some

bins. The criteria are therefore the result of a compromise, and thus do not remove the physics effects as much as possible.

Ultimately, the physics effects can never be completely removed, even if no such compromise was done. For instance, there will always be additional radiations: out-of-cone radiations will always be present by definition, and additional radiations (like a third jet) with too low energies might not be correctly measured, or not be measured at all: perfectly balanced dijet events are rare if not impossible, regardless of how tight the selection criteria are. More fundamentally, our knowledge of physics is incomplete, and our simulations are imperfect, so even the best MC generator does not perfectly describe the true physics.

Systematic uncertainties are estimated to account for the fact that the physics effects are not fully removed:

- These selection criteria on $\frac{p_T^{\text{jet 3}}}{p_T^{\text{avg}}}$, $\Delta\phi^{1,2}$ and JVT are varied (see Table 4.3) to obtain new calibrations, that are compared to the nominal calibration to obtain systematic uncertainties. The amount by which those selection criteria are varied has been studied in [103].
- The calibration is obtained with another MC generator, and compared to the nominal calibration to obtain the "MC modeling uncertainty".

In all cases, the nominal calibration $\mathcal{C}^{\text{nom}}$ is compared to the alternative calibration $\mathcal{C}^{\text{alt}}$ to obtain the relative systematic uncertainty:

$$\Delta\mathcal{C}^{\text{alt}} = \frac{\mathcal{C}^{\text{alt}}}{\mathcal{C}^{\text{nom}}} - 1 \tag{4.64}$$

In the context of the absolute implementation of the calibration, the JVT uncertainty has a slightly different expression. Indeed, there is no JVT variation at truth level, since there is no pile-up and thus no use of the JVT selection criterion at truth level. In this case, the nominal intercalibration coefficient for the MC at the denominator is used. The expression for the JVT uncertainty on the data-specific calibration is therefore:

$$\Delta\mathcal{C}^{JVT} = \frac{\mathcal{C}^{JVT}}{\mathcal{C}^{\text{nom}}} - 1 = \frac{\frac{\left(c_{\text{data}}^{\text{reco}}\right)^{JVT}}{\left(c_{\text{MC}}^{\text{reco}}\right)^{\text{nom}}}}{\frac{\left(c_{\text{data}}^{\text{reco}}\right)^{\text{nom}}}{\left(c_{\text{MC}}^{\text{reco}}\right)^{\text{nom}}}} - 1 \tag{4.65}$$

(which simplifies to $\frac{\left(c_{\text{data}}^{\text{reco}}\right)^{JVT}}{\left(c_{\text{data}}^{\text{reco}}\right)^{\text{nom}}} - 1$). A similar change is made for the JVT uncertainty on the MC-specific calibration. For clarity, the expression is also given in the case of the relative implementation:

$$\Delta\mathcal{C}^{JVT} = \frac{\mathcal{C}^{JVT}}{\mathcal{C}^{\text{nom}}} - 1 = \frac{\frac{\left(c_{\text{data}}^{\text{reco}}\right)^{JVT}}{\left(c_{\text{MC}}^{\text{reco}}\right)^{JVT}}}{\frac{\left(c_{\text{data}}^{\text{reco}}\right)^{\text{nom}}}{\left(c_{\text{MC}}^{\text{reco}}\right)^{\text{nom}}}} - 1 \tag{4.66}$$

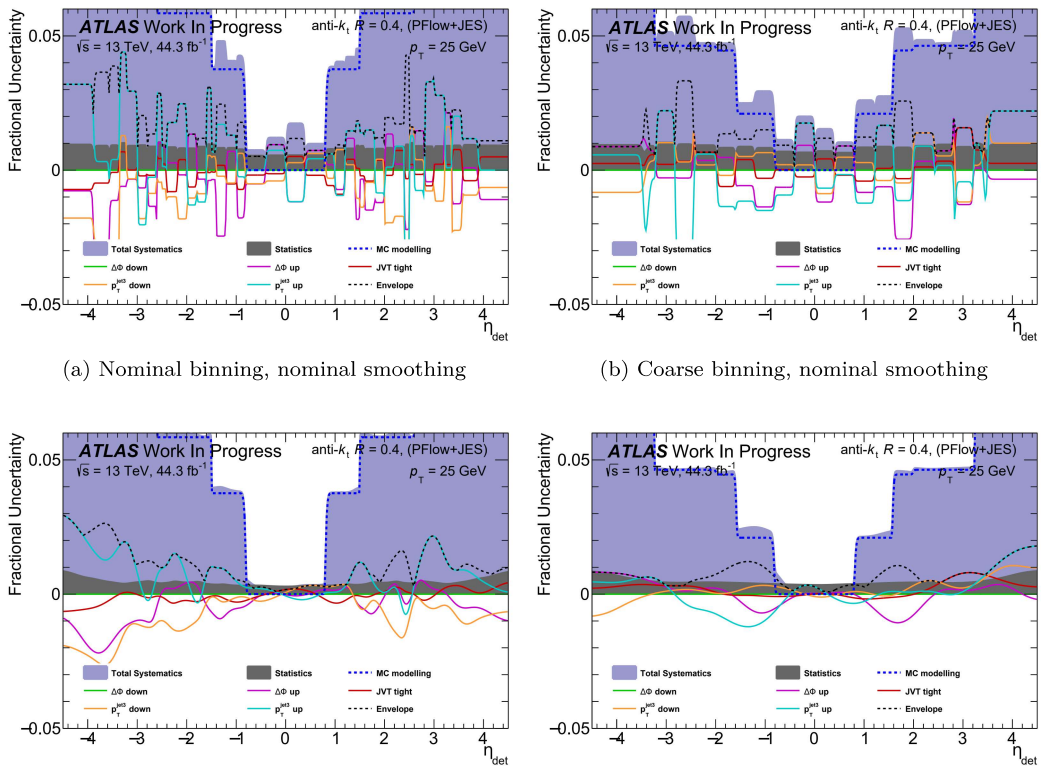
where one observes the presence of $\left(c_{\text{MC}}^{\text{reco}}\right)^{JVT}$.

4.2.8.3 Binning and smoothing

The systematic uncertainties are computed using a coarser binning than that of the nominal calibration (see Section 4.2.2.2), and using a stronger smoothing (see Section 4.2.7.3); a comparison of both those effects is shown in Fig. 4.15. There are two reasons for that.

First, the systematic uncertainties are designed to account for physics effects, which are assumed to be smooth. Therefore, no sharp structure is expected in the systematic uncertainties. However, if physics effects and detector effects do not completely factorize, as explained in Section 4.1.1, the structure of the detector can be observed in the systematic uncertainties. Such a structure is not desired in systematic uncertainties that do not account for detector effects; in fact, this could double-count detector effects, and therefore if the sharp structure of the calibration is observed in the systematic uncertainties, it should be smoothed away.

Second, the systematic uncertainties should be statistically significant, meaning that any non-zero point should indicate a systematic effect and not come from a statistical fluctuation. Large statistical fluctuations are expected (and observed) for the systematic uncertainties originating from the event selection. Indeed, a systematic uncertainty is computed using $\mathcal{C}^{\text{nom}}$ and $\mathcal{C}^{\text{alt}}$, which are obtained mainly using samples of events that have a large intersection. In fact, the majority of the events used to derive both calibrations are common. Both calibrations are therefore heavily statistically correlated. The difference between the two calibrations mostly comes from a few events: those



(a) Nominal binning, nominal smoothing

(b) Coarse binning, nominal smoothing

(c) Nominal binning, strong smoothing

(d) Coarse binning, strong smoothing

Figure 4.15: Comparison of the effects of using **(a,c)** the nominal binning or **(b,d)** the coarse binning, and **(a,b)** the nominal smoothing or **(c,d)** the strong smoothing.

that do not pass some selection criteria but pass the other ones. Based on Eq. (4.64), the statistical uncertainty on a systematic uncertainty $\Delta^{\text{syst}}\mathcal{C}^{\text{alt}}$ is expressed as follows:

$$\Delta^{\text{stat}}\Delta^{\text{syst}}\mathcal{C}^{\text{alt}} = \sqrt{\left(\frac{\Delta^{\text{stat}}\mathcal{C}^{\text{alt}}}{\mathcal{C}^{\text{alt}}}\right)^2 + \left(\frac{\Delta^{\text{stat}}\mathcal{C}^{\text{nom}}}{\mathcal{C}^{\text{nom}}}\right)^2 - 2\rho(\mathcal{C}^{\text{nom}}, \mathcal{C}^{\text{alt}})\,\frac{\Delta^{\text{stat}}\mathcal{C}^{\text{nom}}\Delta^{\text{stat}}\mathcal{C}^{\text{alt}}}{\mathcal{C}^{\text{nom}}\mathcal{C}^{\text{alt}}}} \tag{4.67}$$

where $\rho\left(\mathcal{C}^{\text{nom}}, \mathcal{C}^{\text{alt}}\right)$ is the statistical correlation between $\mathcal{C}^{\text{nom}}$ and $\mathcal{C}^{\text{alt}}$. This requires the evaluation of the statistical correlations, which I couldn't perform due to the lack of the bootstrap method, so I couldn't rely on automatic procedures to rebin or smooth the systematic uncertainties as is done for the cross-section measurements, see Section 5.3. Note that in the case of the MC modeling uncertainty, the two MC samples are uncorrelated.

The coarse η binning had been chosen by previous analyzers. I chose the smoothing parameters empirically to smooth out what appeared to be statistical fluctuations: I increased the strength of the smoothing until fluctuations did not make the uncertainty cross zero two frequently.

The η-intercalibration is set to 1 on average in the central region $|\eta| < 0.8$ (see 4.2.6.4), both for $\mathcal{C}^{\text{nom}}$ and $\mathcal{C}^{\text{alt}}$. This naturally makes the systematic uncertainties close to 0 in the central region. This is even enforced. The systematic uncertainties[20] are set to 0, with a smooth transition at the borders, just like the one used for the calibration: the calibration is multiplied by $\frac{1}{2}\left(\text{erf}\left(\frac{|\eta|-0.8}{0.05}\right)+1\right)$.

4.2.8.4 Symmetrization

Physics effects are symmetrical with respect to $\eta = 0$, therefore the systematic uncertainties should also be symmetrical. However, some systematic uncertainties for the relative calibration have been found not to be symmetrical. The decision has been taken to symmetrize them. Technically, multiple options are possible: $\frac{\Delta\mathcal{C}(p_T,\eta)+\Delta\mathcal{C}(p_T,-\eta)}{2}$, $\frac{|\Delta\mathcal{C}(p_T,\eta)|+|\Delta\mathcal{C}(p_T,-\eta)|}{2}$ and $\max\left(|\Delta\mathcal{C}(p_T,\eta)|, |\Delta\mathcal{C}(p_T,-\eta)|\right)$.

[20] Only the systematic uncertainties are set to 0, not the statistical uncertainties. Indeed, the η-intercalibration is applied in the central region: the four bins in $|\eta| < 0.8$ have an average set to 1, but are not set to 1. These bins are subject to statistical fluctuations, and therefore have statistical uncertainties.

This asymmetry could come from two sources: statistical fluctuations or detector effects. However, I do not know whether the uncertainties are statistically significant or not, the bootstrap method is required for that, which I have not had time to make work. It also seems plausible that the asymmetry comes from detector effects, because such an asymmetry is not observed in the truth-level intercalibration coefficients. This asymmetry is observed in the intercalibration coefficients primarily in the first p_T bin in data18 and in mc16d_reco.

These uncertainties are meant to cover physics effects; but, we know that using the MC at reco level convolves physics effects and detector effects. So without further investigation, I cannot assume that this asymmetry is only of statistical origin, and therefore the uncertainty should either not be symmetrized, or be symmetrized in a conservative manner.

Out of the three solutions tested, $\frac{\Delta\mathcal{C}(p_T,\eta)+\Delta\mathcal{C}(p_T,-\eta)}{2}$ is clearly the least conservative as it assumes that the asymmetry is solely of statistical origin, whereas $\max\left(|\Delta\mathcal{C}(p_T,\eta)|, |\Delta\mathcal{C}(p_T,-\eta)|\right)$ is the most conservative option. The latter option is chosen. Symmetrizing before smoothing (with the options that use absolute values) clearly overestimates the uncertainties, especially in the central region; symmetrizing after smoothing introduces cusps. I chose the second option.

4.2.8.5 Combination of the systematic uncertainties

The systematic uncertainties are combined to provide a single uncertainty to physics analyses instead of six uncertainties:

1. For the $\frac{p_T^{\text{jet 3}}}{p_T^{\text{avg}}}$ and $\Delta\phi^{1,2}$ uncertainties, the up and down variations are combined by taking the maximum in absolute value of either of them[21].
2. The four remaining uncertainties ($\frac{p_T^{\text{jet 3}}}{p_T^{\text{avg}}}$, $\Delta\phi^{1,2}$, JVT and MC modeling) are summed in quadrature.

4.2.8.6 Checking at truth level that systematic uncertainties originating from the event selection do account for the physics effects

Thanks to the intercalibration coefficients that I produced at truth level in the context of the absolute implementation of the η-intercalibration, I was able to study physics effects without any detector effect, and to compare them to the systematic uncertainties designed to cover them.

The physics effects induce a smooth and non trivial structure on the intercalibration coefficient, as can be seen in Figs. 4.16a to 4.16c. A similar structure with sharper variations due to convolution with detector effects is also present at the reconstruction level shown in the upper panels of Fig. 4.18, where red triangles show Powheg+Pythia8 at reco level. The systematic uncertainty is defined at truth level by $\Delta c = \left|\frac{c^{\text{alt}}}{c^{\text{nom}}} - 1\right|$, where c^{alt} is the intercalibration coefficient obtained with an alternative selection of the events (*i.e.* a different selection on $\frac{p_T^{\text{jet 3}}}{p_T^{\text{avg}}}$ or on $\Delta\phi^{1,2}$), and is shown in Figs. 4.16d to 4.16f.

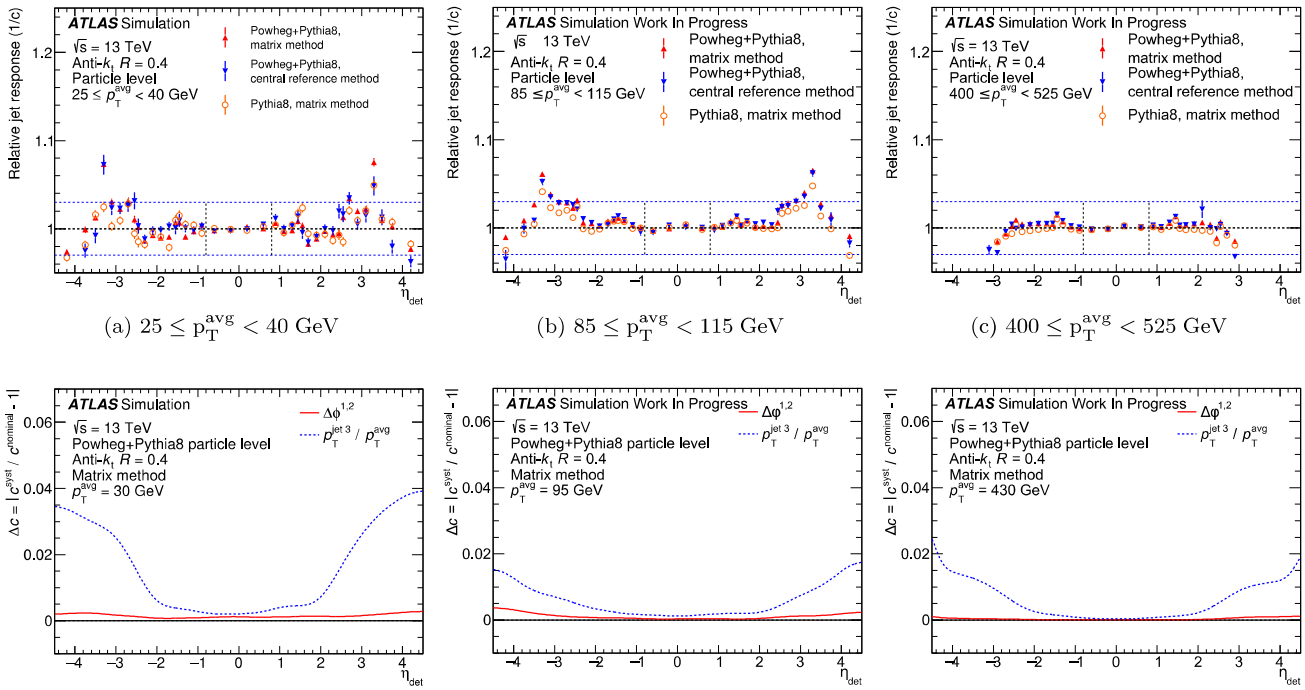
By comparing Figs. 4.16a to 4.16c and Figs. 4.16d to 4.16f, I observed that the magnitude of the physics effects, seen in the intercalibration coefficient at truth level, are of the same order of magnitude as the systematic uncertainties that are designed to cover them, in most phase space region: at low p_T and high p_T, both are of the order of 3 % in the $2.5 \leq |\eta| < 4.5$ region. Therefore, I was able to check that those uncertainties are not underestimated in most phase space regions. To my knowledge, this important check had never been done before.

However, at intermediate values of p_T ($p_T \approx 100$ GeV), the systematic uncertainties are of 0.8 % to 1.5 %, whereas the physics effects go as high as 3 % in the same η regions. The physics effects are 2 to 4 times larger than the associated uncertainties. Note that this discrepancy does not necessarily mean that the physics effects are underestimated by the systematic uncertainties. Indeed, the final calibration is a data over MC ratio, and therefore the physics effects cancel out. Hence, this study done at the level of MC truth only proves that the uncertainties are not underestimated in two phase space regions, low p_T and high p_T, but it does not conclude anything for the intermediate p_T region.

This provides a method to determine the values of the selection criteria for future analyses: by using selection criteria such that the magnitude of the physics effects observed at truth level is similar to the magnitude of the systematic uncertainties.

Another possible improvement for further investigations would be to use selection criteria that would have another dependence on p_T^{avg} than the current $\frac{p_T^3}{p_T^{\text{avg}}} < 0.25$. The value 0.25 and its variations ± 0.1 seem well-suited for low and high p_T, and other values could be used at intermediate p_T: a tighter nominal selection to better remove the physics effects (*i.e.* a degradation of the statistical uncertainty), or wider alternative criteria such that

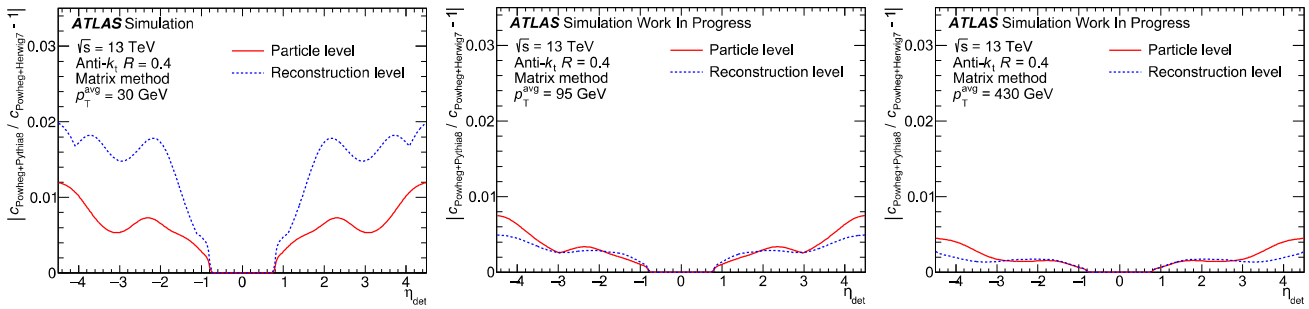
[21] Jargon of ATLAS: this is often referred to as "symmetrization" of the up and down variations.



(a) $25 \leq \mathrm{p_T^{avg}} < 40$ GeV (b) $85 \leq \mathrm{p_T^{avg}} < 115$ GeV (c) $400 \leq \mathrm{p_T^{avg}} < 525$ GeV

(d) $p_T^{\mathrm{avg}} = 30$ GeV (e) $p_T^{\mathrm{avg}} = 95$ GeV (f) $p_T^{\mathrm{avg}} = 430$ GeV

Figure 4.16: **(a,b,c)** Intercalibration coefficients obtained at truth level for Powheg+Pythia8 (for the MC16d campaign, corresponding to the 2017 data-taking period). **(d,e,f)** Systematic uncertainties obtained at truth level solely with the intercalibration coefficients: $\Delta c = \left|\frac{c^{\mathrm{alt}}}{c^{\mathrm{nom}}} - 1\right|$, using the same binning as the calibration, the same smoothing parameters, and symmetrized in the same way. The bins and values are chosen to be representative of the low-p_T, mid-p_T and high-p_T regions. A non-trivial structure is observed in the forward region. In the low-p_T and high-p_T regions, the physics effects observed in the truth level intercalibration coefficients are of the same order of magnitude as the systematic uncertainties of selection designed to cover those physics effects, thus proving that the systematic uncertainties are not underestimated in these regions.



(a) $p_T = 30$ GeV (b) $p_T = 95$ GeV (c) $p_T = 430$ GeV

Figure 4.17: MC modeling uncertainty obtained either at truth level (red curves) or at reco level (blue curves) for Powheg+Pythia8 (for the MC16d campaign, corresponding to the 2017 data-taking period). The truth-level uncertainty is a lot smaller in the $p_T < 40$ GeV and avoids any possible double-counting of the detector effects with the flavor uncertainties. The truth-level MC modeling uncertainty is therefore adopted.

the systematic uncertainties better encompass the physics effects (*i.e.* a degradation of the systematic uncertainties). It amounts to having a selection on $p_T^3 < f\left(p_T^{\text{avg}}\right)$, with f non linear in p_T^{avg}.

Non-monotonic MC modeling uncertainty

In previous analyses, the MC modeling uncertainty was forced to increase when $|\eta|$ increases, because physics effects were assumed to only increase when $|\eta|$ increases. Indeed, as a first and conservative hypothesis, the topology of events is increasingly more different than the one in the central region when we go more and more forward. Therefore, the MC modeling uncertainty was forced to increase (by using, in one bin, the maximum between the original value in that bin and the value in the adjacent bin with lower $|\eta|$).

However, what this study at truth level shows is that the physics effects do not only increase when going more forward: they increase and then decrease, as shown in Figs. 4.16a and 4.16b.

This constraint is therefore removed. This reduces the MC modeling uncertainty in the forward region for two reasons: first, as said, the uncertainty is allowed to decrease in the forward region, but also the uncertainty is less affected by statistical fluctuations (any upwards statistical fluctuation used to make the whole forward region have an artificially high MC modeling uncertainty).

4.2.8.7 Using the truth level for the Monte Carlo modeling uncertainty

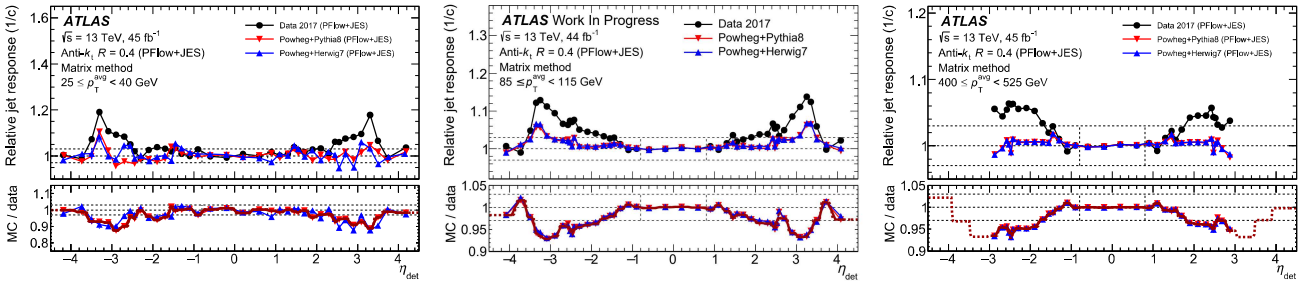
Variations in the parton showering and hadronization models can significantly affect the dijet balance, through:

- Physics effects with extra radiations like a third jet or out-of-cone radiations;
- The detector response, through the jet composition and substructure.

The impact on the detector response is already evaluated by the flavor uncertainties described in Section 3.3.1.4. Therefore, using the MC at reco level in the evaluation of the MC modeling uncertainty double-counts this impact on the detector response. I therefore chose to use the MC at truth level for the evaluation of the MC modeling uncertainty, since it is only impacted by physics effects.

This improvement is not only beneficial for the relative implementation, that uses the MC at reco level in the denominator of the calibration, it is also useful for the absolute implementation: the MC modeling uncertainty should only be evaluated on the data-specific correction, not on the MC-specific one.

This amounts to a reduction of the MC modeling uncertainty by a factor of 2 below 40 GeV, a region where this uncertainty used to be largely dominant, see Fig. 4.17.



(a) $25 \leq \mathrm{p_T} < 40$ GeV (b) $85 \leq \mathrm{p_T} < 115$ GeV (c) $400 \leq \mathrm{p_T} < 525$ GeV

Figure 4.18: The η-intercalibration *vs.* η for the data-taking period 2017, for the relative implementation. **(a-c)** The upper part of each plot shows the (inverse) intercalibration coefficients $1/c$; the black dots are data, the red triangles are Powheg+Pythia8, the blue triangles are Powheg+Herwig7. The lower part of each plot shows the calibration $\mathcal{C}$; the red points show the nominal calibration obtained with Powheg+Pythia8, the blue points show the alternative calibration obtained with Powheg+Herwig7 used to derive the MC modeling uncertainty; the red curve is the smoothed calibration.

4.3 Calibrations and their uncertainties: results

The calibrations are obtained for the relative implementation ($\mathcal{C}^{\mathrm{relative}}$) and for the absolute implementation, the latter including two calibrations: the data-specific one ($\mathcal{C}^{\mathrm{absolute}}_{\mathrm{data-specific}}$) and the MC-specific one ($\mathcal{C}^{\mathrm{absolute}}_{\mathrm{MC-specific}}$), as described in Section 4.1. The calibrations are computed independently for the three data-taking periods 2015+2016, 2017 and 2018. The calibrations use Powheg+Pythia8 as the nominal MC generator and Powheg+Herwig7 as the alternative MC generator to evaluate the MC modeling uncertainty. The matrix method is used to derive the intercalibration coefficients, along with the post-fit propagation method of the statistical uncertainties. The nominal calibration is obtained with the granular binning presented in Section 4.2.2.2, which goal is to capture the fast variations in the calibration coefficient, and with the weak smoothing presented in Section 4.2.7.3 not to smooth out those fast variations. The calibration is constrained to be close to unity in the central region $|\eta| < 0.8$.

The systematic uncertainties are obtained with the conditions described in Section 4.2.8.2: coarser binning and stronger smoothing for the systematic uncertainties to reduce the statistical fluctuations impacting them and because those uncertainties cover for physics effects and therefore are supposed to be smooth. The systematic uncertainties are symmetrized with respect to $\eta = 0$, and then set to 0 in the central region $|\eta| < 0.8$. The up and down systematic variations are "symmetrized" separately for the $\Delta\phi^{1,2}$ and for the $p_T^{\mathrm{jet}\ 3}/p_T^{\mathrm{avg}}$ selection criteria by taking the maximum in each bin. The four remaining systematic uncertainties are added in quadrature.

4.3.1 Precision calibration with the relative implementation

4.3.1.1 The calibration and its uncertainties

The final calibration with the relative implementation is obtained, and shown *vs.* η in Fig. 4.18 for the data-taking period 2017. The uncertainties *vs.* η are shown in Fig. 4.19. The calibration along with the uncertainties is shown *vs.* p_T in Fig. 4.20. The calibrations for the three data-taking periods display the same qualitative features; the other data-taking periods are shown in Appendix A.6.

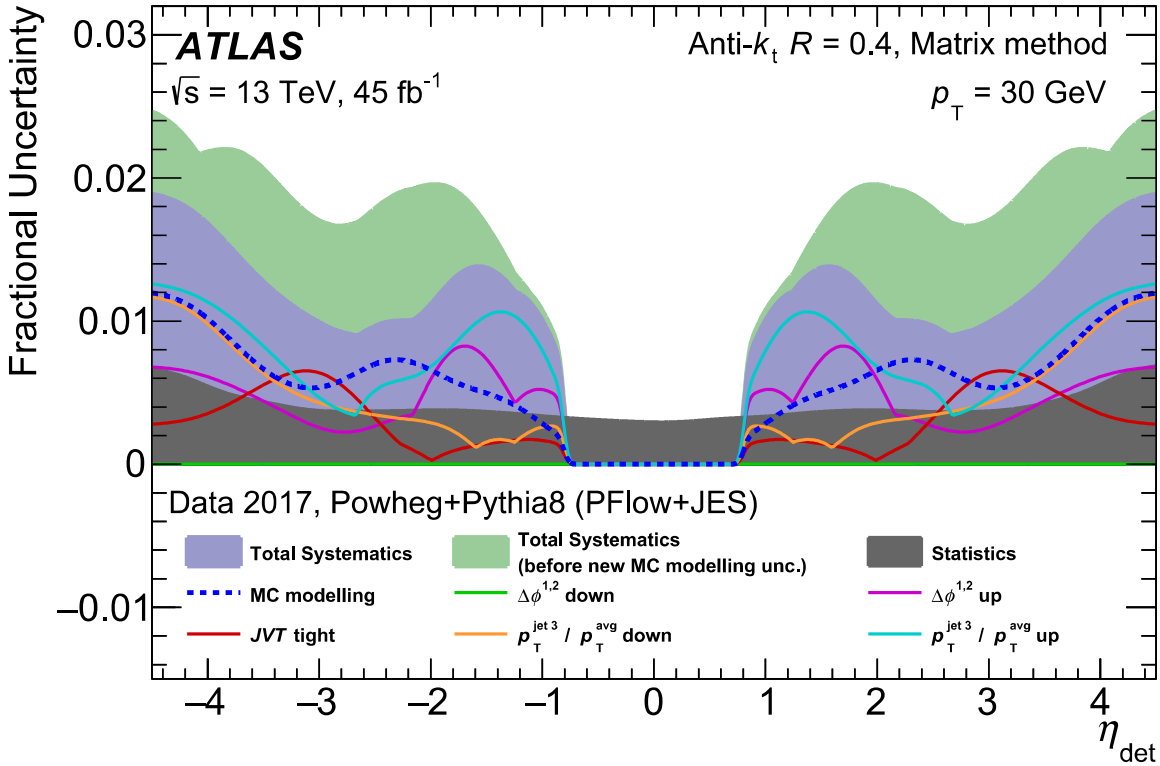
The calibration for large-R jets, to which I contributed by running part of the η-intercalibration framework (event selection, NTuple production, histogram production), is shown in Appendix A.7 for the data-taking period 2017.

The calibration

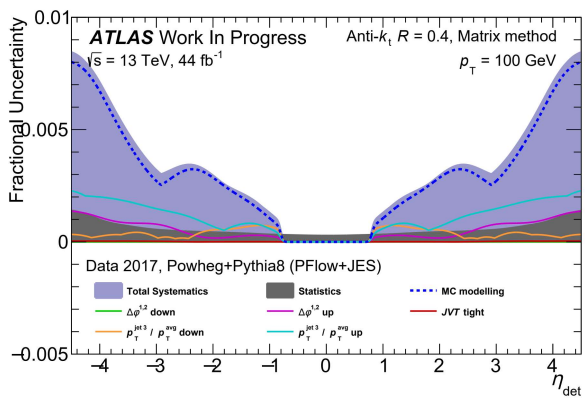
The calibration displays an expected smooth structure *vs.* p_T, and fast variations around $|\eta| \approx 1.4$, $|\eta| \approx 2.5$ and $|\eta| \approx 3.1$, which are also expected due to the detector structure (see Fig. 2.22). The sharpest variations are around the $|\eta| \approx 2.5$, which is the region that requires the smallest bins. These structures will be analyzed more in detail later in Section 4.3.2 in light of the deconvolution of the detector and physics effects brought by the absolute calibration.

The fast variations observed in the intercalibration coefficients and in the calibration points are not smoothed out thanks to the new values of the smoothing parameters.

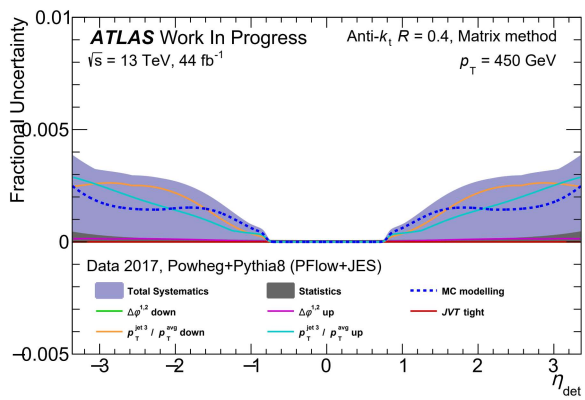
The forward region is always the region receiving the most important correction: up to 13 % in the lowest p_T



(a) $p_T = 30$ GeV

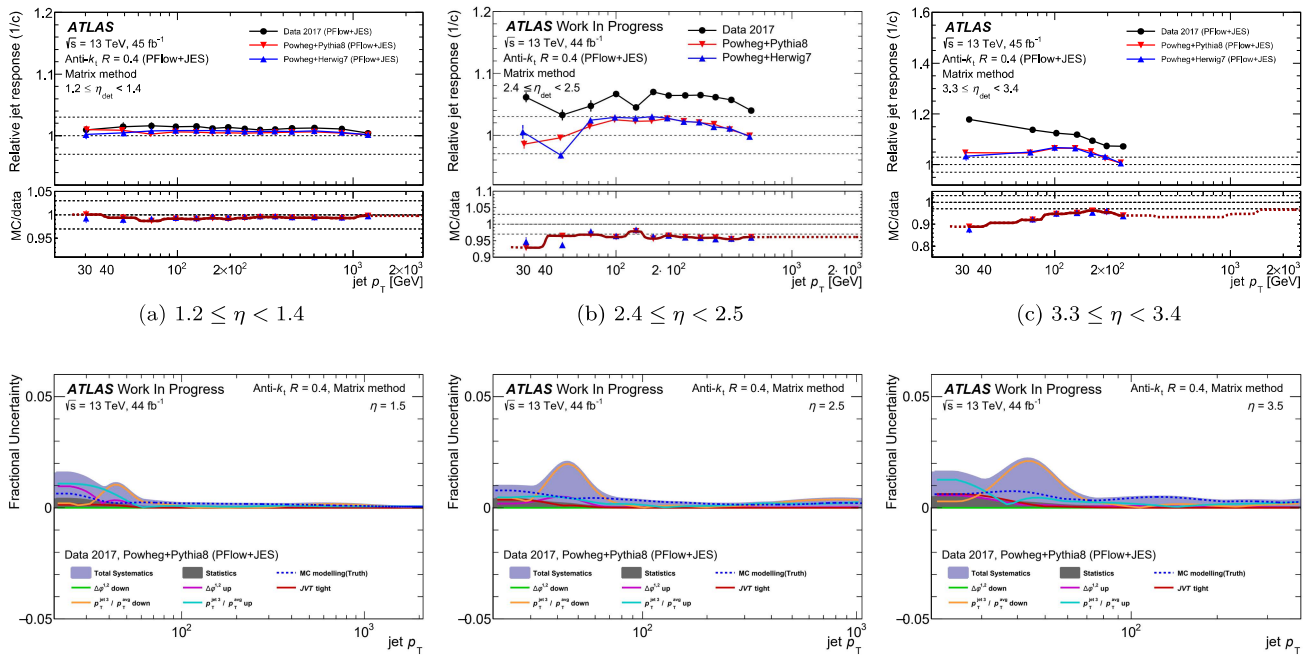


(b) $p_T = 100$ GeV



(c) $p_T = 450$ GeV

Figure 4.19: The uncertainties for the η-intercalibration *vs.* η for the data-taking period 2017, for the relative implementation. The MC modeling uncertainty (dashed blue curve) is obtained by comparing the calibration obtained with Powheg+Pythia8 and Powheg+Herwig7. The green band in **(a)** shows the total systematic uncertainties using the MC modeling at reco level, while the purple band uses the MC modeling at truth level: I chose the latter as it amounts to a reduction of 50 % of the MC modeling uncertainty, that used to be dominant below 40 GeV, while avoiding possible double-counting of detector effects with the flavor uncertainties (see discussion in Section 4.2.8.7).



(a) $1.2 \leq \eta < 1.4$ (b) $2.4 \leq \eta < 2.5$ (c) $3.3 \leq \eta < 3.4$

(d) $\eta = 1.5$ (e) $\eta = 2.5$ (f) $\eta = 3.5$

Figure 4.20: **(a-c)** The η-intercalibration *vs.* p_T for the data-taking period 2017 and **(d-f)** the associated uncertainties, for the relative implementation. **(a-c)** The upper part of each plot shows the (inverse) intercalibration coefficients $1/c$; the black dots are data, the red triangles are POWHEG+PYTHIA8, the blue triangles are POWHEG+HERWIG7. The lower part of each plot shows the calibration $\mathcal{C}$; the red points show the nominal calibration obtained with POWHEG+PYTHIA8, the blue points show the alternative calibration obtained with POWHEG+HERWIG7 used to derive the MC modeling uncertainty; the red curve is the smoothed calibration.

bin at $|\eta| \approx 3$, see figure Fig. 4.18a. Except for this region, the correction does not exceed 8 %. Near the central region ($0.8 < |\eta| < 1.5$), the correction is only of a few percents and never exceeds 3 %.

I have obtained the calibration for the 2018 data-taking period for the first time. Since this calibration displays features that are substantially similar to the ones of the 2017 calibration, it is not detailed here, though values of the calibration are larger by about one percent in the lowest p_T bin (which suffers from a large non-closure, as detailed later). This validates the approach of [72] to calibrate the 2018 data with the 2017 calibration by simply adding a "non-closure" uncertainty for 2018. The calibrations for the three data-taking periods can be seen for a few representative bins in Appendix A.6.

The statistical uncertainties

The statistical uncertainties mainly depend on p_T and slightly on η, and mostly reflect the statistics in data: the cross-sections fall when p_T and $|\eta|$ increase, and the prescales are very high for the low-p_T triggers.

In the central region, the statistical uncertainties start around 0.4 % at low p_T (around 30 GeV), go down to 0.02 % around 600 GeV, and then increase to 0.05 % at high p_T (around 1700 GeV). In the forward region, the statistical uncertainties start around 0.6 % at low p_T (around 30 GeV), go down to 0.03 % around 600 GeV, and then increase to 0.05 % at high p_T (around 1700 GeV). The statistical uncertainties are never dominant for the η-intercalibration.[22]

The systematic uncertainties

There are 6 systematic uncertainties described in Section 4.2.8.2: the MC modeling uncertainty that compares POWHEG+PYTHIA8 to POWHEG+HERWIG7, and 3 selection systematic uncertainties along with their up and down variations that aim at evaluating the physics effects that are mitigated thanks to the selection criteria on $\frac{p_T^{\text{jet } 3}}{p_T^{\text{avg}}}$, $\Delta\phi^{1,2}$ and JVT.

[22] Still, having more events selected by the low-p_T prescales would be useful for two reasons, as noted previously: it would avoid $\left(p_T^{\text{left}}, p_T^{\text{right}}\right)$ from being dropped due to the low effective statistics, especially in the forward-forward region; it also makes the asymmetry distributions more Gaussian, thus making the χ^2 of the Gaussian fit better and avoiding the use of the arithmetic mean.

I have dramatically reduced the MC modeling uncertainty, by showing that it is better to compute it at truth level instead of at reco level (as discussed in Section 4.2.8.7). This amounts to a reduction of a factor 2 in the $p_T < 40$ GeV, where it was largely dominant and is now of the same order of magnitude as other uncertainties. This improvement is shown by comparing the purple and the green bands in Fig. 4.19a, where the green band is computed with the MC modeling at reco level and the purple band is computed with the MC modeling at truth level.

The dominant uncertainty mostly depends on p_T:

- Below 40 GeV, all the systematic uncertainties are of a similar order of magnitude: approximately 1 %.
- From 40 GeV to 80 GeV, the $\frac{p_T^{\text{jet 3}}}{p_T^{\text{avg}}}$ uncertainties are dominant: approximately 1 % to 2 %.
- From 80 GeV to 200 GeV, the MC modeling uncertainty is dominant: from 0.1 % near the central region to 0.8 % at $|\eta| \approx 4.5$.
- Above 200 GeV, the MC modeling uncertainty and the $\frac{p_T^{\text{jet 3}}}{p_T^{\text{avg}}}$ uncertainties are of the same order of magnitude and dominant: from 0.1 % to 0.3 %.

Note that the down variation of the $\Delta\phi^{1,2}$ uncertainty is always 0. This comes from the fact that the quantities $\frac{p_T^{\text{jet 3}}}{p_T^{\text{avg}}}$ and $\Delta\phi^{1,2}$ are physically correlated: because of the "new" value of the nominal section on $\frac{p_T^{\text{jet 3}}}{p_T^{\text{avg}}}$ (chosen in [103]), all the events rejected by the nominal value of the $\Delta\phi^{1,2}$ selection are already rejected. Therefore, the down variation of the $\Delta\phi^{1,2}$ selection cannot add any new event, as it is a looser selection. This selection on $\Delta\phi^{1,2}$ is still kept for the up variation, that does remove some events.

Also note that the cusps in the systematic uncertainties come from the symmetrization around $\eta = 0$, as discussed in Section 4.2.8.4.

Qualitative comparison to the previous calibration

The new calibration that I obtained is compared to the previous calibration from [72] in Fig. 4.21, and the associated uncertainties are compared in Fig. 4.22. My new η-intercalibration is derived with jets calibrated with the new pile-up removal, the new MC JES and the new GNNC described in [89].

Qualitatively, the calibrations are of similar orders of magnitude in every phase space region, and similarly for the systematic uncertainties of selection (except the $\frac{p_T^{\text{jet 3}}}{p_T^{\text{avg}}}$ at 50 GeV, which could be a statistical fluctuation since this difference is not observed in the adjacent bins). This is expected, since the improved calibrations applied to jets prior to the η-intercalibration are applied both on data and MC, and the η-intercalibration looks at data-to-MC ratios. Still, not observing any degradation of the *in situ* calibrations is an important validation of the recent improvements in the previous steps of the calibration chain.

The MC modeling uncertainty is drastically reduced, for multiple reasons:

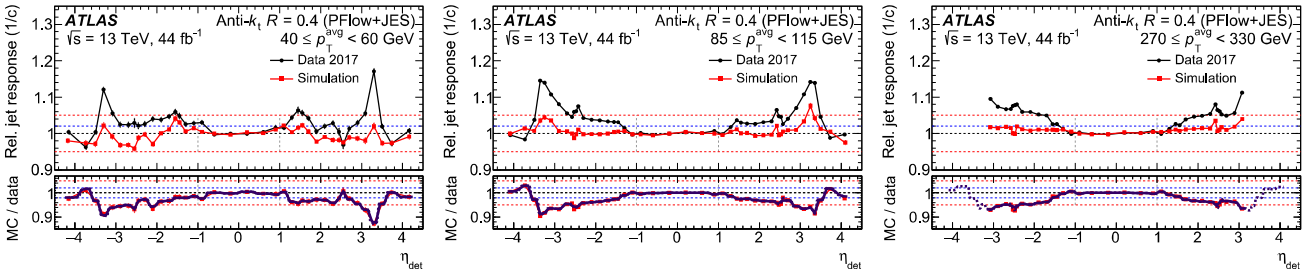
- The main reason is the change in alternative MC generator: it is now Powheg+Pythia8, a next-to-leading order generator, with a high number of events; whereas previously it was Sherpa, a leading order generator, with a low number of events;
- The use of the truth level instead of the reco level reduces it by a factor of 2 for $p_T < 40$ GeV (shown in Fig. 4.19a and discussed in Section 4.2.8.7);
- The removal of the constraint for the MC modeling uncertainty to necessarily increase when $|\eta|$ increases (discussed in Section 4.2.8.6) also reduces this uncertainty locally.

4.3.1.2 Closure test

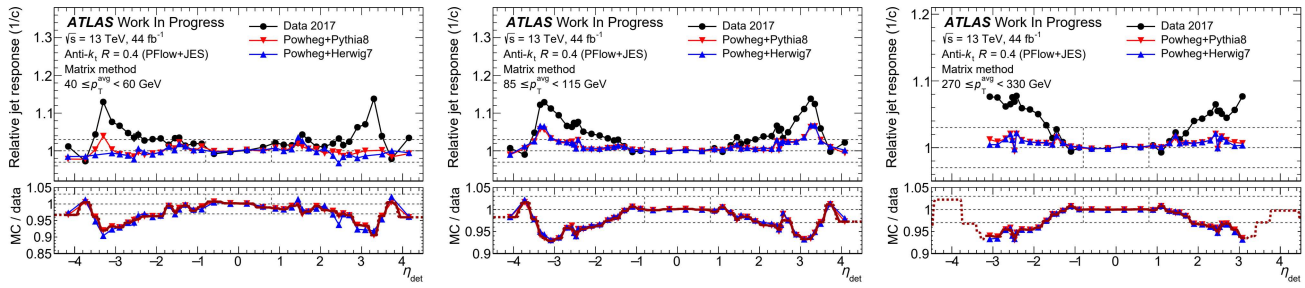
The "closure" is a methodological test: I apply the calibration that I derived to the events, and then re-derive the calibration that is then called the "non-closure". If the non-closure is compatible with 1 within statistical uncertainties, then the calibration is deemed to have a good closure. Otherwise, the non-closure is used as a methodological uncertainty in physics analyses, in phase-space regions where it is not compatible with unity.

In the case of the relative implementation, the η-intercalibration is only applied to data, whereas in the case of the absolute implementation, the data-specific calibration is applied to data and the MC-specific calibration is applied to MC. I did not look at the closure in the case of the absolute implementation, as the absolute implementation did not reach a state where it could be used in physics analyses.

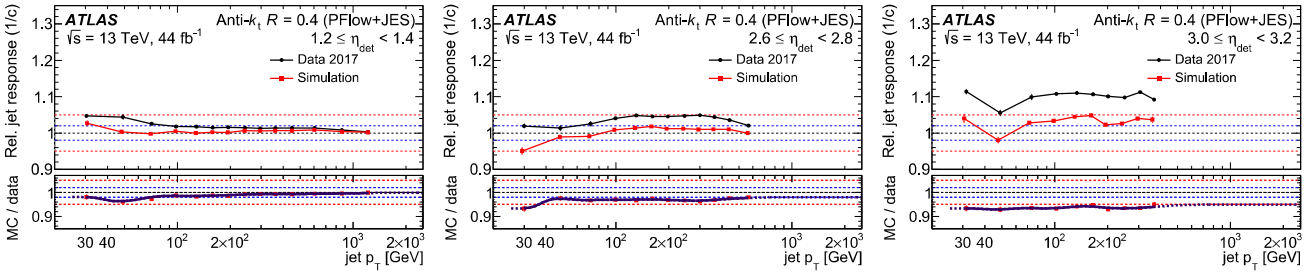
Note that the non-closure is derived in the fine binning and with no smoothing or very little smoothing, to avoid removing local methodological bias. The smoothing parameters used here are $\sigma_{\log p_T} = 0.01$ and $r = 20$. Ideally, the calibration should be applied on an event-by-event basis. For the sake of time, I applied the calibration to the p_T histograms of the two leading jets before computing the asymmetry (using the barycenter of each bin). This induces a small bias in the measurement of the non-closure, which is really negligible.



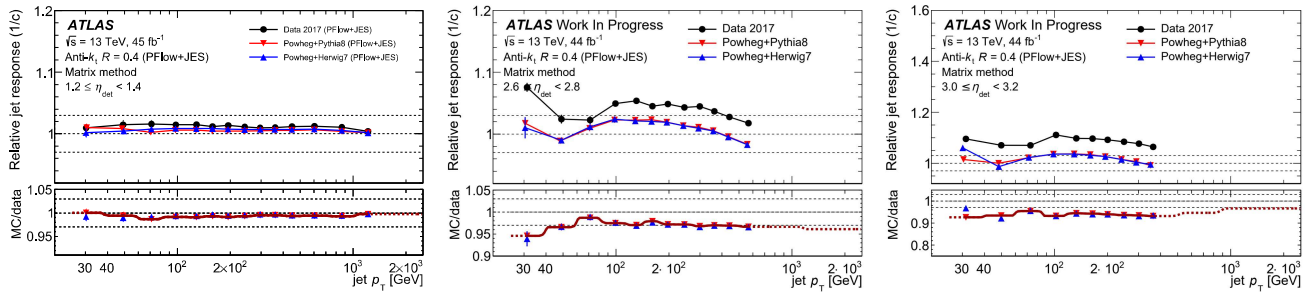
(a) $40 \leq p_T < 60$ GeV (b) $85 \leq p_T < 115$ GeV (c) $270 \leq p_T < 330$ GeV



(d) $40 \leq p_T < 60$ GeV (e) $85 \leq p_T < 115$ GeV (f) $270 \leq p_T < 330$ GeV

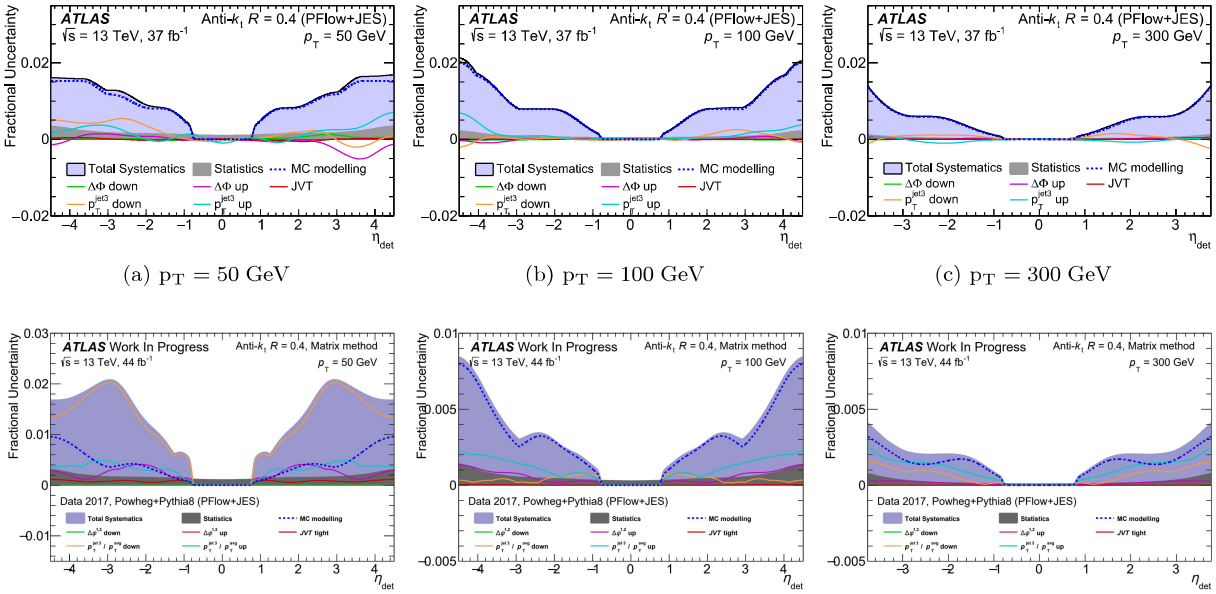


(g) $1.2 \leq \eta < 1.4$ (h) $2.6 \leq \eta < 2.8$ (i) $3.0 \leq \eta < 3.2$



(j) $1.2 \leq \eta < 1.4$ (k) $2.6 \leq \eta < 2.8$ (l) $3.0 \leq \eta < 3.2$

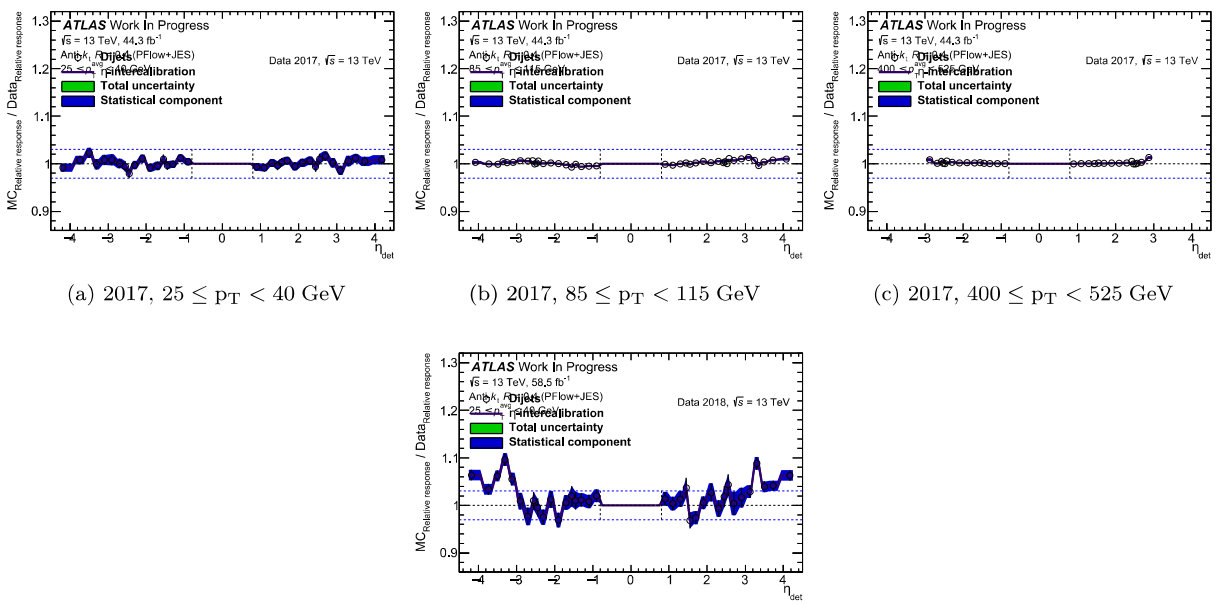
Figure 4.21: Comparison between the previous η-intercalibration (plots **(a-c,g-i)** reproduced from [72]) and my new η-intercalibration (plots **(d-f,j-l)**, derived with the new jet calibration applied, as described in [89]). Qualitatively, the corrections are of the same order of magnitude. Note that **(e,j)** have already been shown in Figs. 4.18 and 4.20, and are repeated here to ease the side-by-side comparison.



(a) $p_T = 50$ GeV (b) $p_T = 100$ GeV (c) $p_T = 300$ GeV

(d) $p_T = 50$ GeV (e) $p_T = 100$ GeV (f) $p_T = 300$ GeV

Figure 4.22: Comparison between **(a-c)** the uncertainties of the previous η-intercalibration (reproduced from [72]) and **(d-f)** my new η-intercalibration (derived with the new jet calibration applied, as described in [89]). Qualitatively, the systematic uncertainties of selection are of the same order of magnitude, except the $\frac{p_T^{\text{jet 3}}}{p_T^{\text{avg}}}$ that is bigger at 50 GeV. The MC modeling uncertainty is drastically reduced in every phase space region, mostly due to the change in alternative MC generator: it is now Powheg+Pythia8, a NLO generator, with a high number of events, whereas previously it was Sherpa, a LO generator, with a low number of events. Other improvements make the MC modeling uncertainty lower now: the use of the truth level reduces it by a factor of 2 for $p_T < 40$ GeV (shown in Fig. 4.19a and discussed in Section 4.2.8.7), and the removal of the constraint for the MC modeling uncertainty to necessarily increase when $|\eta|$ increases (discussed in Section 4.2.8.6). Note that **(e)** has already been shown in Fig. 4.19, and is repeated here to ease the side-by-side comparison.



(a) 2017, $25 \leq p_T < 40$ GeV (b) 2017, $85 \leq p_T < 115$ GeV (c) 2017, $400 \leq p_T < 525$ GeV

(d) 2018, $25 \leq p_T < 40$ GeV

Figure 4.23: Closure of the relative implementation of the η-intercalibration, **(a-c)** for 2017 data, where the non-closure is of 3 % maximum at low p_T, and compatible with unity within statistical uncertainties in most phase space regions; and **(d)** for the only bin in which the non-closure is above 3 %: the $25 \leq p_T < 40$ GeV bin of 2018.

The closure is shown for 2017 data for representative p_T bins in Fig. 4.23. The non-closure is below 3 % in all phase space regions and for all data-taking periods, and less than 1 % almost everywhere, except for one single bin: the $25 \leq \mathrm{p_T} < 40$ GeV bin of 2018. The non-closure is also always significantly smaller than the calibration, in every phase space region.

The non-closure is large for the $25 \leq \mathrm{p_T} < 40$ GeV bin of 2018 simply because of the lack of the central HLT_j15 trigger for 2018: only the existing forward trigger is used in this bin. It is known thanks to [103] that only combinations of central and forward triggers are fully efficient. Therefore, the lack of the central trigger induces a strong bias, measured by the non-closure. Ideally, the minimum-bias trigger should have been used in this bin for 2018: this would have resulted in a smaller non-closure, but a much bigger statistical uncertainty.

Note that I have not looked at the significance of the non-closure uncertainty.[23] The significance should be checked before turning the non-closure into a systematic uncertainty provided to physics analyses, to avoid double-counting the statistical uncertainties.

This closure is better than the one obtained for the previous calibration, presented in [103]. The non-closure had a significance greater than 2 in only three regions: $E > 4.5$ TeV (where it reached ± 3 %), and $2.4 < |\eta| < 2.6$ (where it reached ± 5 %). In all these regions, the new non-closure is better, especially in the high-energy region thanks to the new freezing point, as discussed in Section 4.2.7.3.

This amounts to an overall uncertainty reduction.

4.3.2 Comparison of the relative and absolute implementations

The absolute implementation of the η-intercalibration, described in 4.1, aims at a better deconvolution of the physics and detector effects, in order to understand them better individually. I have obtained the calibrations for the relative and absolute implementations of the η-intercalibration, which are compared in Fig. 4.24 for the 2017 data-taking period and for a few selected bins. A more thorough comparison for the three data-taking periods and for more bins is shown in Figs. A.2 to A.7.

The deconvolution of the effects is achieved.
The goal of deconvolving the physics effects and the detector effects is achieved: two separate calibrations are obtained for the absolute implementation instead of one for the relative implementation: $\mathcal{C}^{\mathrm{absolute}}_{\mathrm{data-specific}} = \frac{c_{\mathrm{data}}}{c_{\mathrm{MC\ truth}}}$ and $\mathcal{C}^{\mathrm{absolute}}_{\mathrm{MC-specific}} = \frac{c_{\mathrm{MC\ reco}}}{c_{\mathrm{MC\ truth}}}$, the former only containing the detector effects of data and the latter only containing the detector effects of MC, thanks to the use of the MC at truth level that allows to remove the physics effects from both quantities.

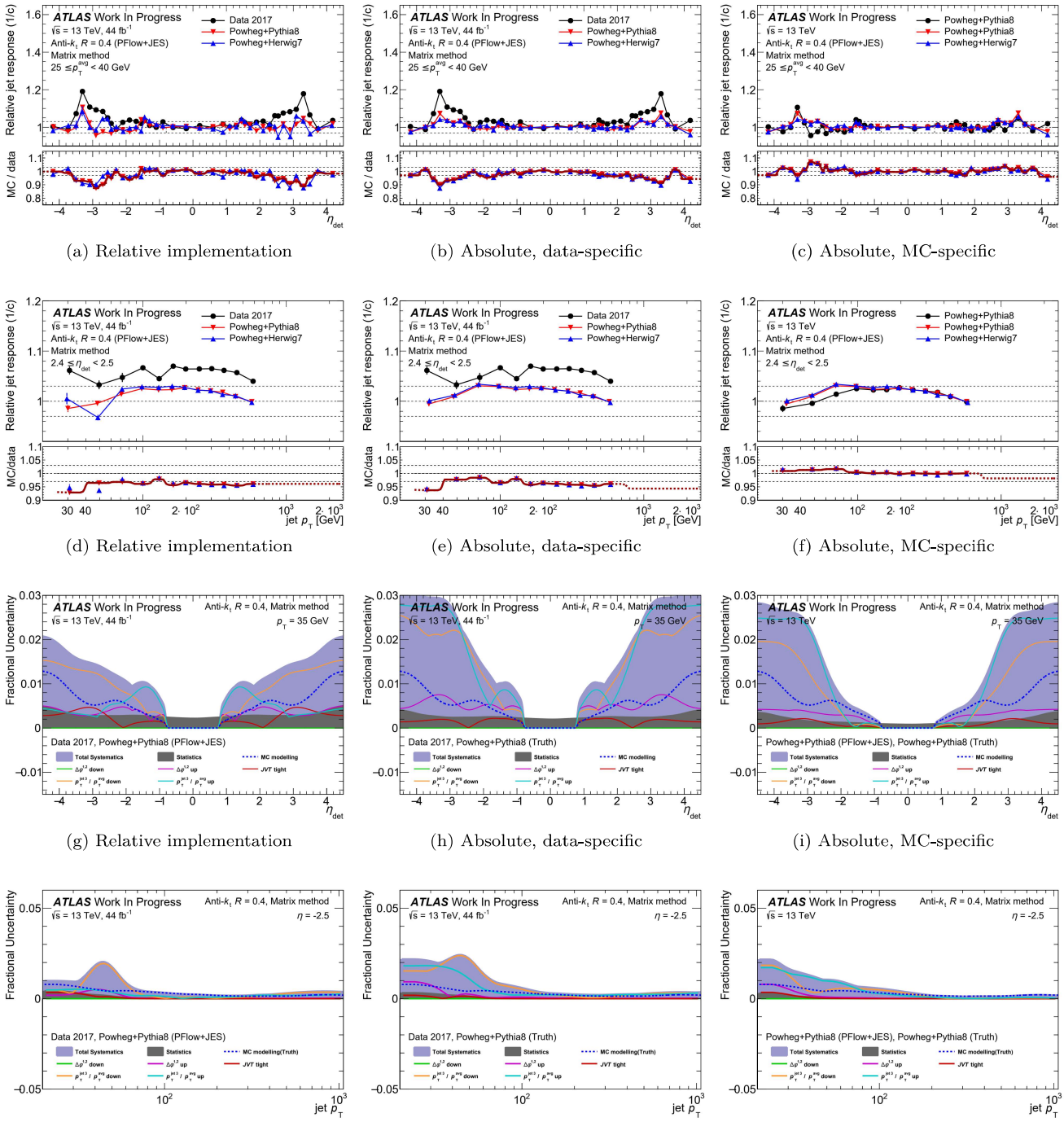
The calibrations for the relative and for the absolute implementations are different, especially in the lowest-p_T bin where the absolute MC-specific calibration is close to unity but not completely flat. The MC-specific calibration contains the residual simulated detector effects, that are corrected by the MC JES calibration step, so this result was not obvious, it was rather expected to be flat and compatible with unity within uncertainties.

The MC JES and the η-intercalibration are indeed a bit different:

- The selection of the events is not the same: the MC JES uses all the jets of all the events with a pretty loose selection, whereas the η-intercalibration only uses the two leading jets of dijet events, *i.e.* of events where the sub-sub-leading jet is negligible (if any).
- Since the topology of the events is not identical, the flavor composition of the jets used in both cases is different.
- The MC JES corrects the energy of the jet, whereas the η-intercalibration corrects its transverse momentum.
- The MC JES directly compares the truth and the reco levels of jets, whereas the η-intercalibration compares two different jets.
- The η-intercalibration uses data events, provided by triggers above their turn-on point. Observing this non-flatness of the detector-effects could point towards a bias in the selection of data, for instance the turn-on point of the triggers might be underestimated. However, this non-flatness is also observed in the MC-specific calibration and can only come from a miscalibration in the MC JES, so if such a bias in the turn-on point of the triggers exists, it does not explain everything.

The absolute implementation precisely aims at observing those kinds of effects, to isolate them and understand them better in future analyses. In this case, the non-flatness of the residual simulated detector effects at low p_T, which is not fully understood yet, could not have been observed with the relative implementation.

[23] Indeed, for the time being, my calibration has not been used yet due to bad timings, but discussions are ongoing about the optimization of the uncertainties of the cross-section measurements presented in Chapter 5 that would require my new calibration.



(a) Relative implementation (b) Absolute, data-specific (c) Absolute, MC-specific

(d) Relative implementation (e) Absolute, data-specific (f) Absolute, MC-specific

(g) Relative implementation (h) Absolute, data-specific (i) Absolute, MC-specific

(j) Relative implementation (k) Absolute, data-specific (l) Absolute, MC-specific

Figure 4.24: Comparison of the relative and absolute implementations of the η-intercalibration. **(a-f)** show the intercalibration coefficients in the upper pads and the calibrations in the lower pads, where the black dots are the reco quantities (data or Powheg+Pythia8), the red triangles are the nominal MC generator Powheg+Pythia8, the blue triangles are the alternative MC generator Powheg+Herwig7, the red curve is the smoothed calibration. **(g-l)** show the associated relative uncertainties, where the grey band is the statistical uncertainty, the purple band is the sum in quadrature of the symmetrized systematic uncertainties. Note that for the absolute calibration, the data-specific and MC-specific uncertainties are heavily physically correlated. Note that **(a,d,g,j)** have already been shown in Figs. 4.18 to 4.20, and are repeated here to ease the side-by-side comparison.

Conclusions on the physics effects thanks to the truth level.
Observations can also be made on the physics effects, by looking at the truth-level MC intercalibration coefficients, *i.e.* at the red and blue points in the upper pads of Figs. 4.24b, 4.24c, 4.24e and 4.24f.

I have drawn important conclusions by looking at the physics effects only, as already discussed in Section 4.2.8.6:

- The physics effects display a non-trivial structure: they increase when $|\eta|$ increases, and then decrease after $|\eta| \approx 3.3$. This allowed to remove a constraint previously imposed on the MC modeling uncertainty: it was assumed that physics effects only increase when $|\eta|$ increases, and therefore that the MC modeling uncertainty should also only increase.
- The physics effects are of the same order of magnitude as the systematic uncertainties that are designed to cover for them, both at low-p_T and at high-p_T, thus proving that the systematic uncertainties are not underestimated for sure in these regions.

These conclusions would not have been possible by solely looking at the reco-level intercalibration coefficients, since they convolve physics effects and detector effects.

Understanding the structures in the calibration.
The absolute implementation also allows to better understand the structures in the intercalibration coefficients.

For instance, a large peak is observed in the calibrations in the $2.7 < |\eta| < 3.5$ region. Such a structure is expected because of the detector structure around $|\eta| \approx 3.1$, that is the transition region in the detector between the endcap and forward calorimeters as shown in Fig. 2.22. A peak is also observed at truth level around $|\eta| \approx 3.3$. This peak, also observed at reco level, therefore comes from physics effects, and cancels out in the data-over-MC ratio, to only be left with detector effects.

The η regions where structure in the calibration is expected due to the detector structure are the following ones: $|\eta| \approx 1.4$, which is the transition between the barrel and endcap calorimeter, $|\eta| \approx 2.5$ which is the end of the tracker coverage and the transition between the precision region to the extended region in the electromagnetic endcap calorimeter, and $|\eta| \approx 3.1$ which is the transition between the endcap to forward calorimeter. No significant structure is observed in the detector effects around $|\eta| \approx 1.4$, meaning that the MC JES completely captured these effects (the MC JES in this region is large, with fast variations, as shown in Fig. 3.20). But fast variations are observed at $|\eta| \approx 2.5$ and $|\eta| \approx 3.1$ in the detector effects; the comparison to physics effects allows to make sure that the variations are indeed detector effects.

Another interesting example is the most forward region, where the intercalibration coefficients are close to unity. The reco-level intercalibration coefficients were already known to be close to unity in the forward region, but having this property could have come from a cancellation between detector and physics effects. Now it is known that the physics effects are close to unity in the most forward region, and that therefore the detector effects are also close to unity (*i.e.* well-calibrated by the MC JES) in this region.

A negative-η/positive-η asymmetry is observed in some low-p_T calibrations (for instance in Figs. 4.24a, 4.24c, A.4a and A.4b). This asymmetry is not observed in the truth-level intercalibration coefficients, and therefore does not come from physics effects (as expected). This asymmetry is sometimes observed in the MC reco intercalibration coefficient but not in the data one, as in Fig. 4.24a for the 2017 intercalibration coefficients. It is sometimes observed in the data intercalibration coefficient but not in the MC reco one, as in Fig. A.4a for the 2018 intercalibration coefficients. This points to the fact that the asymmetry does not come from a bias in the η-intercalibration method, but rather to a mismodeling of the detector effects.

Uncertainties At first glance, it looks like the uncertainties for the absolute calibration are larger, but this conclusion would be erroneous: there are 100 % physics correlations between the systematic uncertainties of the same type on the data-specific and MC-specific calibrations (*e.g.* the down variation of the selection on the third jet). In the relative implementation, the correlations are directly dealt with at the moment of the data over MC ratio of intercalibration coefficients. In a physics analysis, the correlations on the data-specific and on the MC-specific calibrations would therefore need to be taken into account, for the cancellations to occur.

It is expected that there should be no increase of the uncertainties, with a proper treatment of the physics correlations. Ultimately, the absolute implementation of the calibration should be tested in a physics analysis, because there the data-specific and MC-specific uncertainties are expected to cancel out due to the statistical correlations.

Conclusion

I have obtained the η-intercalibration for the three data-taking periods 2015+2016, 2017 and 2018, the latter being obtained for the first time. The MC modeling uncertainty has been obtained with the comparison of the next-to-leading order MC generators for the first time, leading to a large reduction of this uncertainty. A tweaking of the p_T freezing and smoothing parameters allowed me to obtain an improved closure with respect to the previous calibration.

I have implemented the absolute approach, explained in Section 4.1, in the η-intercalibration framework, with the intent of a better separation of the physics and detector effects, and I obtained the calibration in this context as well. This absolute implementation of the η-intercalibration was expected to have a lot of benefits, detailed in Section 4.1.2, some of which I have already been able to demonstrate.

The deconvolution of the physics and detector effects, which is the primary goal opening the way to all the other improvements, is achieved. In particular, this allowed me to study those effects independently, to draw important conclusions relevant both for the absolute and the relative implementations. The independent study of the physics and detector effects allowed to understand the source of the features observed in the calibration. The study of the physics effects alone thanks to the truth-level intercalibration coefficients allowed to observe a non-trivial η structure of the physics effects. This motivated the removal of a (conservative) constraint imposed on the MC modeling, allowing it to decrease when $|\eta|$ increases, thus reducing the MC modeling uncertainty.

The qualitative comparison in order of magnitude of the physics effects to the systematic uncertainties of selection allowed me to confirm that these uncertainties, designed to cover for these physics effects, are not underestimated at least at low and high p_T (which does not mean they are underestimated at intermediate p_T), as detailed in Section 4.2.8.6.

Finally, I have redefined the MC modeling uncertainty for the relative implementation by computing it at truth level instead of reco level, to avoid any possible double-counting of the detector effects with the flavor uncertainties. This also led to a reduction of the MC modeling uncertainty by a factor 2 in the $p_T < 40$ GeV region.

Some interesting studies are left for future analyzers. In particular, I have not had time to make the bootstrap method work for the η-intercalibration, which is important to make sure that the systematic uncertainties are statistically significant. Finally, now that the absolute approach of the η-intercalibration is implemented, the next step would be to validate in a physics analysis that it does not lead to an increase of the uncertainties, as expected if the physics correlations are taken into account.

Chapter 5

Jet Production Cross-section Measurements

> Quantum Field Theory [...] has led to the most fantastic agreement between theoretical predictions and experimental data in the history of science. [...] On the other hand, the subject is a mess.
>
> Matthew D. Schwartz, *Quantum Field Theory and the Standard Model*, 2014

Jets are collimated sprays of hadrons, stemming from the showering and hadronization of an outgoing parton (quark or gluon) after a high-energy particle collision, see Chapter 1, Section 1.3. They are one of the main observables at the LHC due to their production rate that is higher than any other particle, and allow for various physics studies, from low-energy non-perturbative QCD to Beyond the Standard Model searches, through high-energy QCD. The jet production cross-section measurements are one of the latter category. Roughly speaking, they are measurements of the probability of emission of a jet from a proton-proton collision at a certain energy in the center of mass. From the experimentalist's point of view, jet production cross-section measurements amount to a counting of jets, normalized by the luminosity.

These precision measurements have multiple uses:

- They can serve for precision evaluations of Standard Model parameters, like the determination of the Parton Density Functions (PDFs) or of the strong coupling constant α_S [128–131]. They are also used for the tuning of the free parameters of MC generators.
- They can be compared to theoretical predictions and constitute a test of the Standard Model and a probe for an indirect search of new physics Beyond the Standard Model: precision measurements are compared to precision predictions, and smooth deviations between the two could be explained by new physics that would manifest itself in the virtual loops. For instance, the exclusion in the low compositeness scale region of contact interactions between left quarks is obtained from the dijet mass spectrum in [132].

The jet production cross-section measurements can be done in multiple ways, with different observables. A cross-section can be total or differential: either the total number of jets is counted (above some p_T threshold and within an angular acceptance), or the number of jets is specified in bins of certain jet kinematic observables. In this chapter, events containing at least one jet are looked at, without caring about the additional possible other particles, but many other measurements have looked at the production of one jet or more along with specified other particles. This chapter focuses on one type of jets: small-R jets (with $R = 0.4$) reconstructed with the PFlow algorithm (see Chapter 3, Section 3.1), inclusively in jet flavor. Finally, the events containing jets can be counted in multiple ways, using the most appropriate observables to capture the kinematics of the events:

- "Inclusive jet production cross-section": every jet in the event is taken into account individually. It has been performed for instance as a function of p_T and $|y|$ in ATLAS [31, 133–136] and CMS [131, 137–141].
- "Dijet production cross-section": only the two leading-p_T jets are taken into account. It has been performed for instance as a function of many different kinematic variables in ATLAS [31, 132–134] and in CMS [139, 142–144].
- "Leading jet production cross-section": only the leading-p_T jet is taken into account. This is a new observable, introduced by Alexander Huss during the "ATLAS Standard Model" workshop in 2017 (Thesaloniki). It has been worked on by Robert Hankache during his Ph.D. in ATLAS but there is no public result yet [103].

- Other observables are possible, for instance the three-jet production cross-section in ATLAS [145] or in CMS [146, 147], or the four-jet production cross-section in ATLAS [148] or in CMS [147].

Strong tensions have been observed between experimental measurements and theory predictions for a long time. For instance, the inclusive jet production cross-section is found to have a good agreement for each individual $(p_T, |y|)$ bin, but when taking the full bin range and the correlations into account, the global $\frac{\chi^2}{N_{\text{dof}}}$ is very high [31, 136]. This points to the fact that the correlations play an important role, and multiple ideas have been expressed and tested to reduce this tension from the experimental side. For instance, the η-intercalibration analysis is performed in hundreds of (p_T, η) bins and having hundreds of local uncorrelated uncertainties allowed to reduce the χ^2 by more than 200 units compared to one single fully-correlated statistical uncertainty over the full (p_T, η) range. CMS also improved the agreement by modifying the correlation scheme [130].

I have had the opportunity to work on two cross-section analyses during this Ph.D.

I worked on the inclusive jet cross-section measurement jointly with Tancredi Carli: he made the dataset selection and the comparison to theoretical predictions; I did the unfolding, including the evaluation of the methodological bias, the statistical uncertainties and correlations and the propagation of the systematic uncertainties of the Jet Energy Scale (JES) and the Jet Energy Resolution (JER). I made the evaluation of the unfolding bias related to data/MC shape difference more quantitative, and I evaluated it in the forward region $3.0 < |y| < 4.5$ for the first time. I measured a new methodological uncertainty that evaluates the impact of the non-Gaussianity of the JER. I also worked on technical aspects to make this computationally-heavy analysis viable on the resources available: I made the Bootstrap library faster.

I worked on the leading jet cross-section measurement: I produced inputs for the nominal unfolding, and most notably I extended the analysis region to $3.0 < |y| < 4.5$ to take potential migrations to this region into account; but more time should be spent on adapting the unfolding code, to produce the inputs for the evaluation of the JES and JER systematic uncertainties, and to produce the theoretical predictions, in order to go to a publication.

This Chapter 5 is organized as follows, after a brief introduction to the inclusive and leading jet cross-section measurements:

- The dataset preparation is explained in Section 5.1, including the choice of data according to triggers, the choice of MC and the cleaning.
- The unfolding procedure is detailed in Section 5.2, with explanations on the Iterative Dynamically Stabilized (IDS) unfolding method used, how the transfer matrix is built, along with the binning choice and the purity and efficiency corrections, and how the number of iterations of the method is chosen, according to a trade-off between the statistical uncertainties and the bias related to data/MC shape differences.
- The systematic uncertainties are then presented in Section 5.3: the propagation of the JES/JER uncertainties, the new non-Gaussianity uncertainty, and the methodological uncertainties.
- The technical optimizations that make these analyses viable on the computational resources available are mentioned in Section 5.4.
- Finally, the results are shown in Section 5.5: the unfolded data and the associated uncertainties, along with a comparison with theoretical predictions.

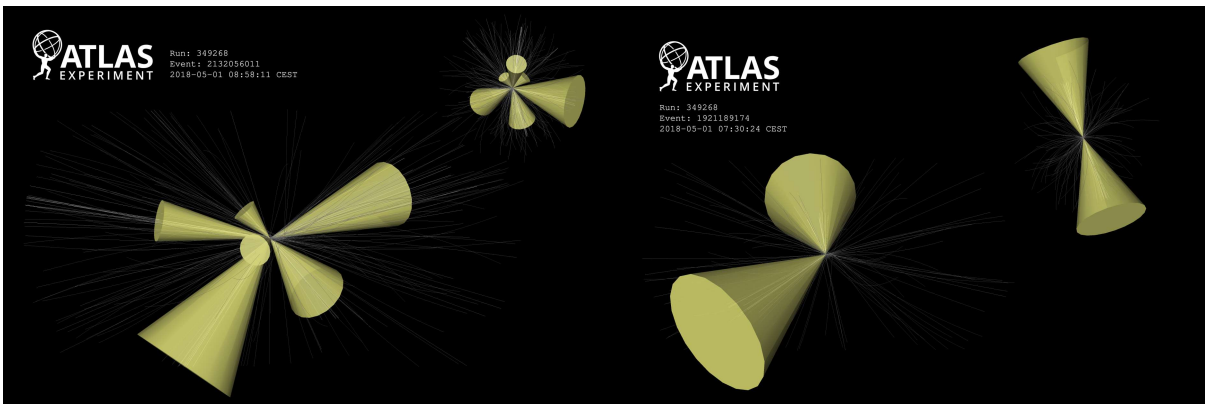
Introduction to the inclusive and leading jet production analyses

The two jet production cross-section measurements I have participated in are the inclusive and the leading jet ones. Both are measured double-differentially in bins of jet transverse momentum p_T and jet rapidity $|y|$, and count the number of jets that fall into those bins: the inclusive analysis counts every single jet of the event, while the leading jet analysis only counts one jet per event: the one with the highest p_T. The relationship between the number of jets N_{ij} in the $(p_T, |y|)$ bin indexed by (i, j) and the cross-section σ is given by:

$$N_{ij} = \mathcal{L} \int_{p_T^i}^{p_T^{i+1}} \int_{|y|^j}^{|y|^{j+1}} \frac{\mathrm{d}^2\sigma}{\mathrm{d}p_T\ \mathrm{d}|y|}\ \mathrm{d}p_T\ \mathrm{d}|y| \tag{5.1}$$

where $\mathcal{L}$ is the recorded luminosity. Dividing by $(\Delta p_T)_i$ and $(\Delta|y|)_j$ on both sides, which are respectively the sizes in p_T and $|y|$ of the bin, one gets the average of the differential cross-section in that bin:

$$\left\langle \frac{\mathrm{d}^2\sigma}{\mathrm{d}p_T\ \mathrm{d}|y|} \right\rangle_{ij} = \frac{N_{ij}}{\mathcal{L}\,(\Delta p_T)_i\,(\Delta|y|)_j} \tag{5.2}$$



(a) (b)

Figure 5.1: Displays of **(a)** a 6-jet event and **(b)** a dijet event, produced in pp collisions at $\sqrt{s} = 13$ TeV, data collected in 2018. Two views of the same event are provided: a side-view with the beam line running horizontally (lower image), and a view in the transverse plane with the beam line running directly into the image (upper image). Reproduced from the auxiliary figures of [149].

The difference between the inclusive and the leading jet cross-sections is that the inclusive one counts every single jet produced, thus having multiple entries from the same event. For instance, the event presented in Fig. 5.1a has 6 jets and therefore 6 entries in the cross-section measurement (in different kinematic regions). This implies that there are correlations between the bins, due to physics correlations between the multiple jets of the event, and this information is not preserved; for instance, the information on the number of jets per event is lost, the correlations between the kinematic variables of the different jets are lost. Focusing on one single bin, the jets in that bin can come from the leading-p_T jet, the sub-leading jet, *etc*; this makes the theoretical predictions harder.

On the contrary, the leading jet observable only counts each event once: in the bin where the leading-p_T jet falls (this is not quite true as explained in the next paragraph). This means that there is only one entry from each event, and so no physics correlation is missing. This also means that all jets in one bin are the highest-p_T jet of the event, thus making the choice of the renormalization and factorization scales less ambiguous: $\mu_R = \mu_F = p_T^{\text{jet}}$. This also has the advantage of focusing on the highest-p_T jet in the event, which is often the one containing the physics of interest.

However, the leading jet observable is not infrared-safe. Most events containing jets, especially at high energy, are dijet events like the one presented in Fig. 5.1b or in Fig. 4.1, with one jet having a (slightly) higher p_T than the other one and thus being the leading jet; however, any soft radiation emitted by the leading jet, that is not clustered in the jet, might make its p_T less than that of the sub-leading jet. This can also occur at the measurement, reconstruction or calibration stages: the jet p_T has a large resolution, and therefore at the reconstructed (reco) level, the p_T of the particle-level (truth-level) leading jet might be lower than that of the particle-level sub-leading jet, thus inverting the p_T hierarchy. Those migrations can also occur with the third highest-p_T jet (called the sub-sub-leading jet), with the fourth one, *etc*. These "order migrations" or "flips" are an additional challenge of the leading jet analysis: they are dealt with at the unfolding stage and by a regularization of the theory predictions, and require not only the leading jet to be accounted for but also the lower-order jets. Having migrations between the jet orders means that a jet in a certain $|y|$ region can become the leading-p_T jet at reco level, even though the truth-level leading-p_T jet is in another $|y|$ region: the migrations between orders also induce migrations between rapidity, that also have to be addressed at the unfolding stage.

The "unfolding" is the deconvolution of the remaining detector effects, statistically. The calibration already brings the jet-level variables as close as possible to their particle-level value, but residual detector effects remain, that can only be corrected for on average: apart from the imperfections in the calibration that also can be addressed at that stage, the unfolding allows to take the resolution migrations into account. Let's call r the reco distribution, t the truth distribution and R the resolution effects, they are related by $r = R \otimes t$ with $\otimes$ denoting the convolution.

For binned observables like the cross-section measurements, the unfolding methods use a resolution matrix R_{ij}, that gives the probability for a jet in a given truth bin to arrive in a given bin at reco level. In this case, the relationship between the two distributions is $r = R \times t$, or more explicitly $r_i = \sum_j R_{ij} t_j$. The resolution matrix, also called the "folding matrix" hereafter, is constructed using MC simulated events, accessible both at truth level and reco level. The unfolding step basically provides an estimate for the particle-level distribution associated to the reco data distribution d through $R^{-1} \times d$.

The analysis is performed as follows:

1. The events are selected to avoid any potential bias as much as possible while maintaining a high-enough selection efficiency, and are used to build both the histograms of the cross-section spectra and the transfer matrix.
2. Since the transfer matrix is built from jets that are geometrically matched between the truth and the reco levels, some jets are not matched, and therefore do not enter in the construction of the transfer matrix, so an impurity correction is applied to the data spectrum, such that the number of jets used to build the transfer matrix is consistent with the number of events in the corrected data spectrum.
3. The unfolding method is applied to obtain the unfolded data spectrum, for different values of the free parameters of the unfolding method; in the case of the IDS unfolding method used here, detailed in Section 5.2.1, the free parameter is a number of iterations. The nominal unfolded result is corrected for the matching efficiency losses, to get back a number of jets and not just a number of matched jets.
4. For each value of the iteration parameter, the statistical uncertainties and the unfolding bias related to the data/MC shape differences are evaluated and compared to settle on a value for the number of iterations. This provides the nominal result.
5. The statistical correlations between the bins are evaluated.
6. The uncertainties from the JES and the JER are propagated by varying respectively the calibration or the Gaussian smearing, then by unfolding using the same number of iterations, and by comparing with the nominal unfolded spectrum.
7. The methodological uncertainty related to the non-Gaussianity of the JER is evaluated.
8. Finally, unfolded data are compared to the theoretical predictions.

The inclusive jet cross-section measurement in CMS

In the CMS collaboration, the other experiment at LHC studying similar physics as ATLAS (see Chapter 2), the inclusive jet cross-section measurement has also been performed, with a similar goal, as already mentioned [131, 137–141].

There are a few notable differences that are worth mentioning:

- The main difference is the unfolding method: in CMS, either a simple matrix inversion (maximum likelihood) is used [131], without any regularization; or the D'Agostini method is used, which has constraints on the smoothness of the distribution. In ATLAS, the IDS method is used, as described in Section 5.2.1; this method does not impose a constraint on the smoothness of the distribution, but requires a hyper-parameter (the number of iterations).
- For its 13 TeV analysis, CMS has created its folding matrix the same way as ATLAS does [131], but in the previous analyses the folding matrix was created by a convolution of the MC truth distribution with a Gaussian (or Gaussian-like function), to account for the migrations caused by the JER; inefficiencies were taken into account by adding an uncertainty.
- CMS does not estimate an uncertainty for the bias related to the sensitivity of the unfolded method to the data/MC shape differences. In ATLAS, this uncertainty is shown not to be negligible, see Section 5.2.3.3.
- CMS takes into account the migrations between rapidity bins. This is not done in ATLAS for the inclusive jet measurement, as these migrations have a negligible impact in the central region $|y| < 3.0$ (but this is done for the leading jet analysis because in this case the migrations are non negligible, especially in the forward region $3.0 < |y| < 4.5$).
- The non-Gaussian tails of the resolution are taken into account when smearing the MC events to match the data distribution. In ATLAS, the effect of the non-Gaussian tails in MC is estimated with a new uncertainty, derived in this work, see Section 5.3.2.
- The geometrical matching is performed within $\Delta R = 0.2$, instead of $\Delta R = 0.3$ in ATLAS; this makes almost no difference, apart from having more migrations taken into account by the inefficiencies instead of the unfolding.
- The statistical fluctuations are smoothed out with Chebyshev polynomials. In ATLAS, this is not done, as this would amount to a modification of data without proper physical justification. The statistical fluctuations do play a crucial role in ATLAS, which is why there are so many calibration uncertainties that are propagated: one for every single calibration bin. The systematic uncertainties are rebinned and smoothed according to their statistical significance, obtained with the bootstrap method, to avoid adding uncertainties that are not statistically significant, and to avoid double-counting potential statistical fluctuations in the χ^2 computation.
- In CMS, the correlation scheme between the uncertainties seems *ad hoc* (it is not justified in detail in the papers).

5.1 Dataset preparation

Dataset description
Experimental data (referred to as "data") used in these analyses have been recorded during the full Run 2 of the LHC by ATLAS from proton-proton collisions at 13 TeV energy in the center of mass, high pile-up conditions ($\langle\mu\rangle = 33.7$) and corresponding to the per-year and total integrated luminosities given in Table 5.1.

2015	2016	2017	2018	Total
3.2 fb^{-1}	33.4 fb^{-1}	44.6 fb^{-1}	58.8 fb^{-1}	140.1 fb^{-1}

Table 5.1: Per-year and total integrated luminosities of the Run 2 of LHC, recorded by ATLAS, obtained through `https://atlas-lumicalc.cern.ch/`. Uncertainty on the total integrated luminosity is 1.2 fb^{-1}, *i.e.* 0.83 % [62].

Simulated data (referred to as "MC") are produced using the Monte Carlo generators presented in Chapter 1, Section 1.3.1.6, and more specifically Pythia8 (LO generator) and Powheg+Pythia8 (NLO generator). The theoretical predictions at fixed order are obtained through NNLOJET++, adding non-perturbative and electroweak corrections, as described in Section 1.3.1.7.

Particle-like inputs are reconstructed from detector-level information using the PFlow algorithm, see Chapter 3, Section 3.1.1.2. Jets are reconstructed with the anti-k_t algorithm using a radius parameter of $R = 0.4$, either at truth level (for MC) using the particle-level information or at reco level (for data and MC), see Section 3.1.2.

At reco level, jets are calibrated with the recommendations of [72] presented in Section 3.3, but not using the improvements of [89] (because the calibration using the new techniques has not been fully computed by the JetETmiss group due to time constraints):

1. Area-based pile-up subtraction followed by the residual pile-up subtraction.
2. MC JES and Global Sequential Calibration (GSC).
3. The full *in situ* calibration chain (η-intercalibration, γ+jet, Z+jet, multijet).
4. The improved high-p_T calibration thanks to the E/p calibration.
5. The MC-to-MC calibration (for the calibration of MC generators other than Pythia8), to have access to the improved flavor uncertainties.
6. A Gaussian smearing is applied to the MC jets in case the resolution in MC is better than the resolution in data; in the opposite case, since we do not want to degrade data, no smearing is applied and a systematic uncertainty is measured. This induces subtleties for the propagation of the JER systematic uncertainties, discussed in Section 5.3.1.

Data	MC reco	MC truth	Selection criteria
✓			Good-Runs List (GRL) v89 (2015-2016), v99 (2017), v102 (2018)
✓			Errors in LAr/Tile/core (corrupted events)
✓	✓		Primary vertex with at least two tracks, and at least two track particles
✓			Triggers
✓	✓	✓	Timing selection: $\lvert t_{\text{jet}}\rvert < 10$ ns
✓	✓		Jet-level cleaning
	✓		MC cleaning: require at least one truth jet and $p_T^{\text{avg,reco}}/p_T^{\text{jet 1,truth}} \leq 1.4$
✓	✓	✓	55 GeV $< p_T$ (inclusive jet) or 70 GeV $< p_T$ (leading jet)

Table 5.2: Selection. All the selection criteria are shown for each type of dataset (data, MC reco and MC truth). When a selection is not required for one type of dataset, there is no check-mark (✓) in the associated cell.

Events are selected using multiple selection criteria, ensuring the good quality of the sample, see Table 5.2. Some of the selection criteria have already been described in the chapter about the η-intercalibration, like the Good Runs Lists (GRLs), the errors in the LAr/Tile/core calorimeters, the triggers, the MC cleaning. The weighting of the data events is simpler than for the η-intercalibration though, as only the leading jet is required to pass a given single-jet trigger, and if it does, all the jets in that event receive the weight of the trigger. The timing selection and the jet-level cleaning are described respectively in Chapter 3, Section 3.2.3 and in Section 3.2.4. Note that no JVT selection is applied, since the high pile-up activity is mostly below 60 GeV, which is below the lowest-p_T bin of the analyses (the JVT likelihoods are not computed above 60 GeV).

These selection criteria (mostly the triggers and the jet-level cleaning) introduce inefficiencies of the order of the percent. Considering the precision reached by the cross-section analyses, they have to be corrected for.

5.2 Unfolding

The goal of "unfolding" is to provide an estimator of a distribution by removing detector effects, in the present case the double-differential jet cross-section spectrum *vs.* p_T and $|y|$. This is essential to compare the measurements with those of other experiments or with theoretical predictions. Here, the detector effects are mostly bin migrations due to the large p_T resolution of jets, but in general it can take into account miscalibrations, the finite acceptance of the detector, or a reconstruction efficiency.

The following notations are used: the data distribution will be written d, with d_i the content if the i-th bin, the MC reco distribution is r, the MC truth distribution is t, the unfolded data distribution is u, the transfer matrix T with T_{ij} the content of the bin labeled (i, j), the folding matrix F and the unfolding matrix $\tilde{F}$. The transfer matrix is filled by numbers of jets: if a jet falls in the $(p_T, |y|)$ bin labeled i at reco level and in the $(p_T, |y|)$ bin labeled j at truth level, its weight is filled in the bin (i, j) of the transfer matrix (added to T_{ij}). The folding matrix (also called the "resolution matrix") and the unfolding matrix are defined from the transfer matrix as:

$$F_{ij} = \frac{T_{ij}}{\sum_k T_{kj}} \text{ and } \tilde{F}_{ij} = \frac{T_{ij}}{\sum_k T_{ik}} \tag{5.3}$$

where F_{ij} gives the probability for a jet in the truth bin j to arrive in the reco bin i, and $\tilde{F}_{ij}$ gives the probability for a jet in the reconstructed bin i to come from the truth bin j.

Since the MC truth and reco distributions are related through $r = F \times t$, the naive unfolding method is to invert F and to apply it to data to obtain the unfolded spectrum $u = F^{-1} \times d$. This method is not ideal, because the statistical uncertainties from the low-entry bins are propagated to the high-entry ones, inducing large fluctuations in those bins and large statistical uncertainties. Multiple methods exist to "regularize" the unfolding and improve the unfolded spectrum, mostly two categories with multiple variants that differ with respect to the regularization:

- The methods that use a matrix inversion, like the Singular Value Decomposition (SVD) method that suppresses the effects of small eigenvalues (that are the most affected by noise) and adds a constraint on the smoothness of the unfolded distribution [150].
- The iterative, Bayes-inspired methods, that do multiple unfoldings using the relation $u \approx \tilde{F} \times d$, improving the truth distribution (the prior) and thus the unfolding matrix $\tilde{F}$ at each iteration [151]. In the unfolding methods, the number of iterations serves as a regularization parameter.

The choice of the binning is also a strong regularization parameter: having a very fine binning induces more migration, hence a more under-determined problem (there is a higher degeneracy of the solutions), and *vice versa* with a coarser binning. The unfolding methods are also tasked to reduce the dependency on simulations as much as possible for the unfolded spectrum to be data-driven.

5.2.1 The Iterative Dynamically Stabilized unfolding method

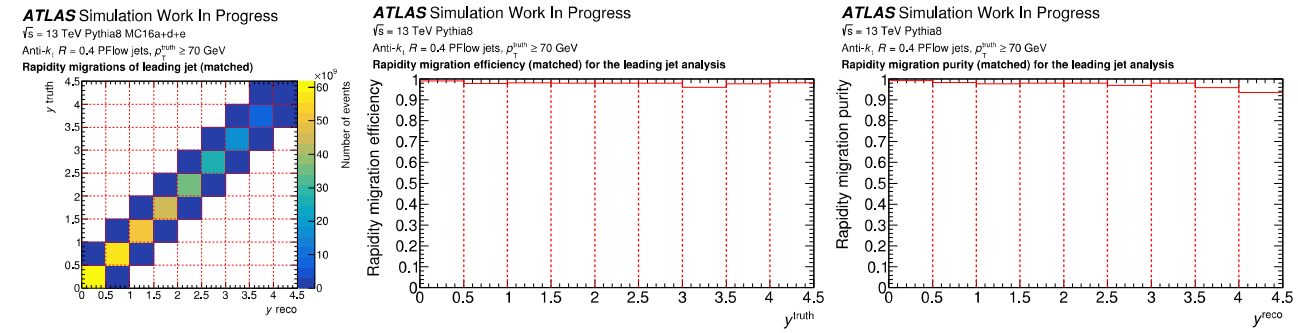
The method used in the cross-section analyses is a powerful iterative method called the Iterative Dynamically Stabilized (IDS) unfolding [152, 153]. This toolbox relies on the folding matrix F correctly describing the migrations due to the resolution, and on the truth-level MC spectrum correctly describing the unfolded data. The latter requirement can be imperfect as the truth-level spectrum is iteratively improved during the unfolding procedure. The IDS method is able to deal with background events or with local non-modeled physics effects (like a mass peak of an unknown particle), though neither of these features is used in the present case (thus the formulas hereafter simplified accordingly). The regularization provided by the IDS method has proven to be efficient for the jet cross-section analyses by requiring less iterations than other methods. The main regularization parameters are the choice of the number of iterations and the binning. The IDS unfolding also provides a local regularization to prevent migrations of events from bins with a high uncertainty to those with a low uncertainty during the unfolding; no global requirement on the shape of the spectrum is imposed (contrarily to D'Agostini's regularization scheme [154, 155] for instance).

At each iteration, the unfolded spectrum is given by:

$$u_j = t_j \frac{N_{\text{data}}}{N_{\text{MC}}} + \sum_k f_k \Delta d_k \tilde{F}_{kj} + (1 - f_j)\Delta d_j \tag{5.4}$$

where $\frac{N_{\text{data}}}{N_{\text{MC}}}$ is a normalization coefficient (the number of entries in the data spectrum over the number of entries in the MC spectra), $\Delta d_k = d_k - \frac{N_{\text{data}}}{N_{\text{MC}}} r_k$, and f_k a regularization term that dictates the fraction of events that migrate from the bins k, based on the uncertainty in that bin:

$$f_k = 1 - e^{-\left(\frac{\Delta d_k}{\sigma_k}\right)^2} \tag{5.5}$$



(a) (b) (c)

Figure 5.2: **(a)** Rapidity migrations, **(b)** rapidity migration efficiency (*i.e.* fraction of truth jets staying in the same reco bin) and **(c)** rapidity migration purity (*i.e.* fraction of reco jets coming from the same truth bin), for the leading jets that are geometrically matched between the truth and reco levels, such that $p_T^{\text{truth}} \geq 70$ GeV. The migrations are therefore due to the rapidity resolution only.

with σ_k the statistical uncertainty on Δd_k. This unfolding equation propagates a fraction of the entries in the bin k into the bin j, based on the uncertainty; the rest of the events stay in the bin j as seen in the last term.

Between each iteration, the truth spectrum is improved by a reweighting according to the data/MC differences, which can equivalently be performed as a modification of the transfer matrix:

$$T'_{ij} = T_{ij} + \Delta u_j F_{ij} \frac{N_{\text{MC}}}{N_{\text{data}}} \tag{5.6}$$

with $\Delta u_j = u_j - \frac{N_{\text{data}}}{N_{\text{MC}}} t_j$ the difference between the unfolded spectrum at the current iteration step and the truth spectrum.

Each iteration of the IDS unfolding reduces the dependence on the truth spectrum and the sensitivity to the data/MC shape differences, but degrades the statistical uncertainties. The choice of the number of iterations is therefore a trade-off, discussed hereafter.

5.2.2 Preparing histograms for the unfolding

5.2.2.1 Choice of the binning

Rapidity binning

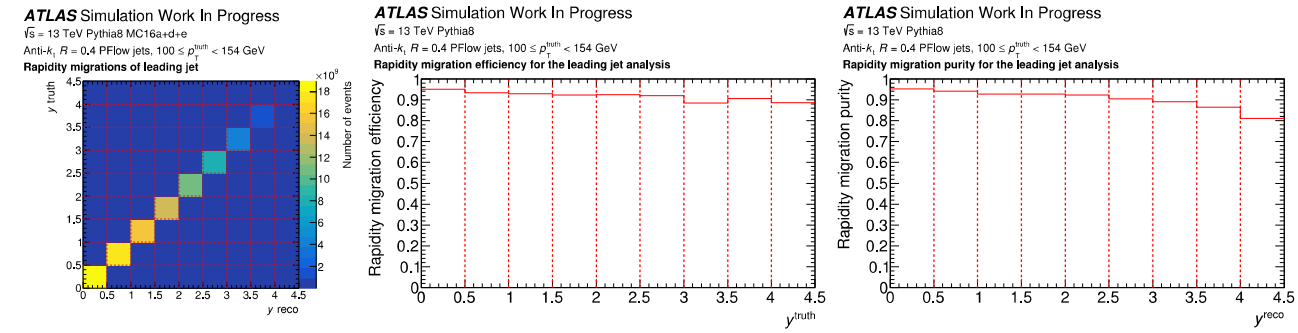
The rapidity binning is the same for the inclusive and leading jet analyses. After the η-intercalibration, no detector structure should affect the spectrum in rapidity, and the rapidity resolution is very good. Since small-R jets have a radius parameter of $R = 0.4$, a rapidity binning that is larger than 0.4 reduces the migrations even more. A regular spacing of 0.5 is observed to provide a good granularity while keeping the statistics high enough and making the migrations between rapidity bins negligible. The rapidity migrations of jets due to the rapidity resolution are shown in Fig. 5.2, for jets that are geometrically matched between truth and reco level: they are of the order of 1 % to 4 %, reaching 7 % in the most forward rapidity bin.

Therefore, the unfolding for the inclusive jet cross-section measurement is only one-dimensional: only the p_T migrations need to be corrected for, while the small rapidity migrations are accounted for in the efficiency and purity corrections. However, for the leading jet analysis, as previously mentioned, the jet order flip forces to take the rapidity migrations into account, thus the unfolding for this analysis has to be three-dimensional: migrations in p_T but also migration in jet order and rapidity. The rapidity migrations due to jet order flips are of the order of 5 % in the most central bin, and 10 % in general, as observed in Fig. 5.3.

Considering the strong data/theory disagreement observed both in the inclusive and leading jet cross-section measurements, the rapidity range considered has been extended compared to the previous works to the forward region $3.0 < |y| < 4.5$, to take potential migrations between the central region and the forward region into account.

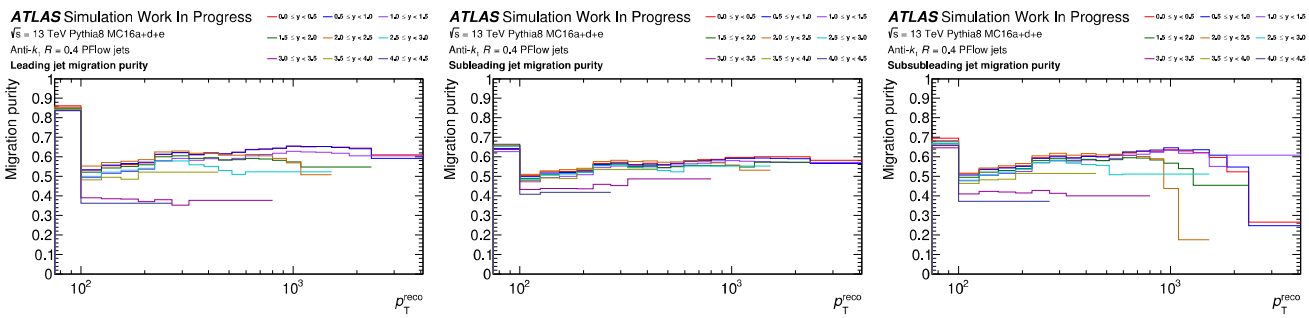
p_T binning

The p_T binning is chosen to be granular enough for interesting data/theory comparisons, while still being coarse enough to avoid having too many migrations: the fewer the migrations, the easier the unfolding. To start with, the binning is therefore chosen to be rather granular. Then bins are merged until the resulting bins are almost equally



(a) (b) (c)

Figure 5.3: **(a)** Rapidity migrations, **(b)** rapidity migration efficiency (*i.e.* fraction of truth jets staying in the same reco bin) and **(c)** rapidity migration purity (*i.e.* fraction of reco jets coming from the same truth bin), for the leading jet of the event such that $100 \geq \mathrm{p_T^{truth}} < 154$ GeV. No geometrical matching condition is imposed, so the rapidity migrations are mostly due to the jet order flips and a little bit due to the rapidity resolution.



(a) Leading jet (b) Sub-leading jet (c) Sub-sub-leading jet

Figure 5.4: Migration purity for the leading, sub-leading and sub-sub-leading jets. MC16a+d+e. The migration purity is generally between 50 % and 70 %, and never goes below 36 %, a value that is reached in the forward region. Note that the migration purity for the inclusive jet analysis is lower, of the order of 40 %, because the bins are finer.

spaced in $\log(p_T)$ and such that the migration purity is high enough in all the bins. The migration purity is defined as:

$$\varepsilon_i^{\text{migration purity}} = \frac{T_{ii}}{\sum_j T_{ij}} \tag{5.7}$$

The migration purity is the fraction of reco events that come from the same truth bin. A high enough migration purity helps having the eigenvalues of the folding matrix not too small (as the matrix is more diagonal in the first place): the migration purity is required to be larger than 0.4, the bins are merged until this condition is met. If the migration purity is larger than that, the bins can be made smaller; in the case of the inclusive jet cross-section analysis, the binning is therefore chosen such that the migration purity is constant around 0.4, while for the leading jet analysis the purity is chosen higher, around 0.6 in the central rapidity bins, to reduce migrations.

The constant purity binning for the inclusive jet analysis has recently been obtained by Ota Zaplatilek. For the leading jet analysis, I used the binning already obtained by Robert Hankache in [103], and I extended it to the forward bins; the resulting binning and migration purities are shown in Fig. 5.4 for the different jet orders and the different rapidity bins. Note that the migration purity for the sub-sub-leading jet does not matter too much, because only the leading jet will be kept at the end, the other two orders are only used for the unfolding; having a bad migration purity for the sub-sub-leading jet affects the bias on the sub-sub-leading jet (maybe a bias of 10 %), and the migrations between leading and sub-sub-leading jets are of 0.1 % maximum, hence making the effect of this bias onto the leading jet negligible.

Both the inclusive and the leading jet analyses extend down to 100 GeV, but not further down due to the complications that the high pile-up activity introduce. However, to take into account the possible migrations of jets from or to the region below 100 GeV, at least one additional bin has to be introduced; it has been shown that extending to 70 GeV is enough. Since the region with a really high pile-up activity is below 60 GeV, the extra bins are not affected by pile-up. The extra bin added below 100 GeV are therefore only used during unfolding, but are

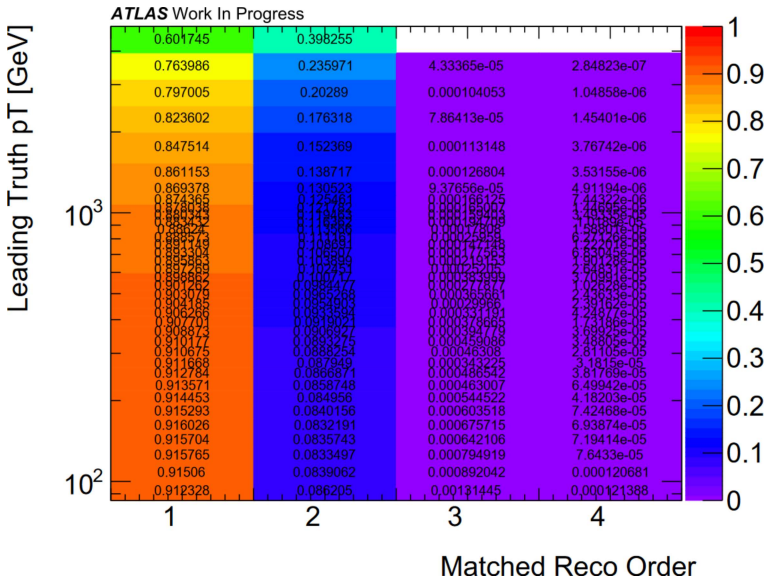


Figure 5.5: Probability of a jet order flip. The y axis shows the p_T of the leading truth jet while the x axis shows the order of the reco jet geometrically matched to it. Reproduced from [103].

removed from the final results.

Choice of the number of jet orders for the leading jet analysis

The jet order flip is to be taken into account for the leading jet analysis, and the number of jet orders to consider has been studied in [103]. In an event, the jets are sorted according to their p_T, which defines their "order": the highest-p_T jet is called the "leading jet", or the "first order jet"; the second highest-p_T jet is the "sub-leading jet" or the "second order jet", *etc.* The jet order flip has been shown in to mostly take place between the first two leading jets, where at very high p_T the probability of a flip is observed to be as high as 40 %, as shown in Fig. 5.5. The flip probability is lower when p_T decreases, but remains high: it is of 9 % in the lowest-p_T region considered in this analysis.

The probability for a leading jet at truth level to become a sub-sub-leading jet is very low though, below 0.1 % in the lowest-p_T region and lower in the higher p_T regions. The probability of a flip with a fourth order jet is at least one order of magnitude lower: completely negligible. The unfolding using only the leading jet has been compared to the unfolding using the three highest-order jets, and a bias of the order of 5 % has been observed [103]. When comparing the unfolding with two and three jet orders, only a bias of the order of the permille level is observed, confirming that the potential bias of using only three jets compared to four jets is negligible compared to the uncertainties of the analysis (which are at the percent level).

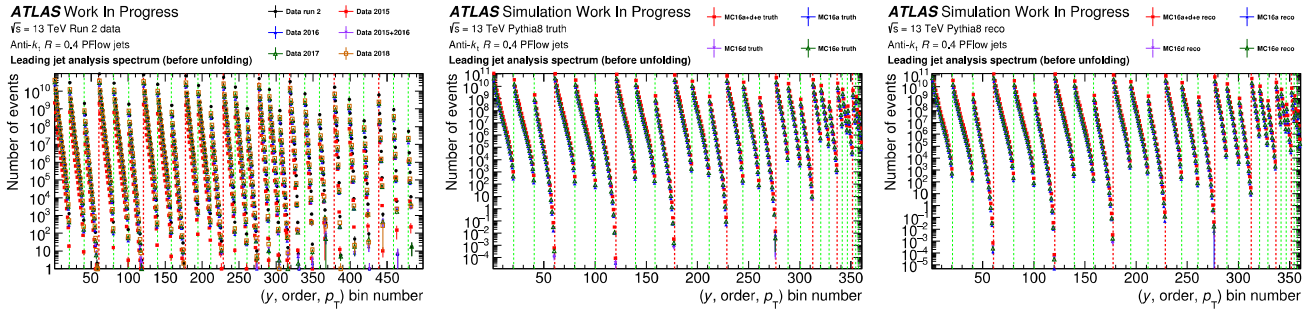
Therefore, the number of jet orders to consider has been chosen to be 3.

5.2.2.2 The transfer matrix

For the inclusive jet analysis, the unfolding is one-dimensional, so the binning is straightforward: there is one transfer matrix per rapidity bin, and the axes of the matrix are simply "p_T of the truth jet" and "p_T of the reco jet". The unfolding for the inclusive jet analysis is performed individually for each of the 9 rapidity bins. In the case of the leading jet analysis, the unfolding is three-dimensional, so a mapping between the $(|y|, o, p_T)$ bins and $\mathbb{N}$ is required to use a one-dimensional unfolding technique. The spectra before unfolding of data, MC truth and MC reco using this bin correspondence are shown in Fig. 5.6.

The transfer matrix is built using the jets that are both present at truth and reco levels. They are required to be geometrically matched: the angular distance between the two has to be lower than $\Delta R = 0.3$. The geometrical matching should be symmetrical: if the closest truth jet to the reco jet labeled i is the one labeled j, then the closest reco jet to the truth jet j must be i. A jet that satisfies this geometrical matching criterion enters the transfer matrix in the appropriate bin. Note that two definitions of the angular distance have been tested, either using the rapidity y or the pseudo-rapidity η, and the difference between the two approaches has been found to be minor, as detailed in the next section.

An example of a transfer matrix for the inclusive jet analysis and the full transfer matrix for the leading jet analysis are shown in Fig. 5.7, along with their associated folding matrices.



(a) Run 2 data (b) MC truth (c) MC reco

Figure 5.6: Spectra of data, MC truth and MC reco with numbers of events. All $(|y|, o, p_T)$ bins. The vertical red lines separate the $|y|$ bins, the vertical green lines separate the jet orders, the spectra between two green vertical lines are *vs.* p_T.

5.2.2.3 Choice between rapidity and pseudo-rapidity

For the geometrical matching used in the building of the transfer matrix, the angular distance can be defined either using the rapidity y or the pseudo-rapidity η:

$$\Delta R_y = \sqrt{\Delta\phi^2 + \Delta y^2} \tag{5.8}$$

$$\Delta R_\eta = \sqrt{\Delta\phi^2 + \Delta\eta^2} \tag{5.9}$$

There are arguments in favor of both. The physical quantity is the rapidity, and the rapidity is shown to be better suited for the definition of the global jet quantities from the jet's constituents than the pseudo-rapidity [156]. The bins of the previously published cross-section analyses are rapidity bins, thus the bins of the analyses presented in this thesis are rapidity bins, to make comparisons possible. However, the computation of the rapidity requires the mass of the jet while the pseudo-rapidity doesn't, and no mass calibration is available for small-R jets (and therefore no associated uncertainty).

In reality, this choice is not expected to play an important role, because:

- The difference between using y or η for the jets changes the barycenter of the jet (it shifts it towards higher $|\eta|$ if we were using η) only if the mass of the jet is big enough, which is not the case of small-R jets in general (it would be different for large-R jets).
- The mass of the small-R jets being low in general, the potential additional mass uncertainties would probably be very low.
- If this choice changes anything for the geometrical matching, this would essentially be taken into account by the efficiencies.

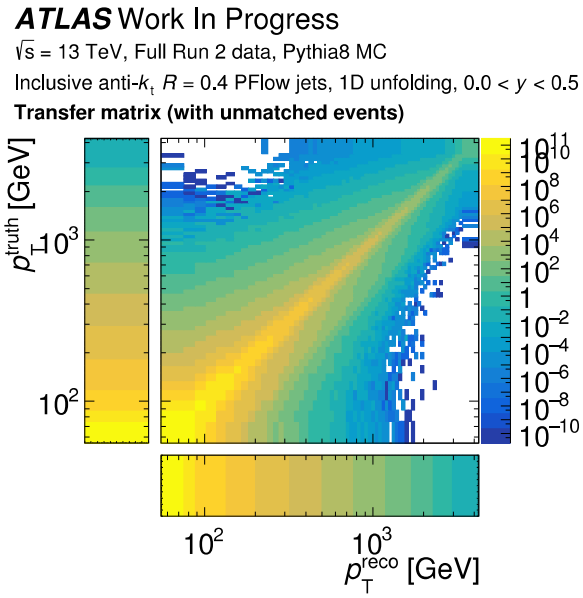
To be sure that this effect is negligible two approaches are possible:

- Compare the ΔR_y and ΔR_η distributions, choose a variation of the selection criterion that encapsulates the 1 σ variations (say ± 0.1), choose between y and η, and obtain a systematic with those variations;
- Unfold once using y and once using η, and compare the unfolded results, yielding a two-point systematic.

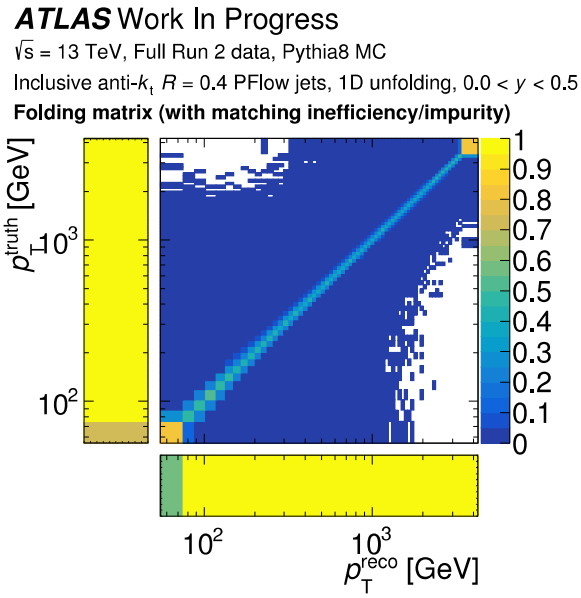
The second approach is chosen.

Figure 5.8a has been made with all the jets of all the PYTHIA8 events (even the lowest-p_T ones) and shows that ± 0.1 variations of the $R < 0.3$ geometrical matching criterion largely encompass the "ΔR_y to ΔR_η migrations", regardless of the definition of ΔR. So the simple ± 0.1 variation is retained.

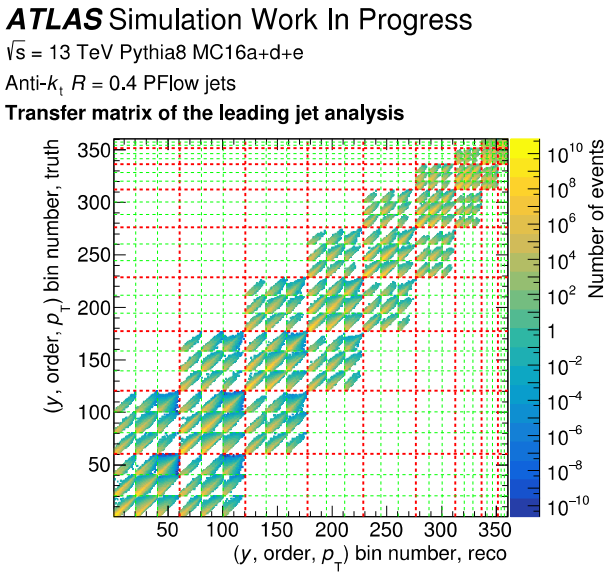
I made this comparison for multiple p_T^{truth} slices: the effect of the definition of ΔR is minor for jets of 100 GeV or more, and clearly taken into account by the ± 0.1 variation, variations of ± 0.05 would still be enough by far. This was expected, as the calculation of the rapidity takes the mass into account, and the higher the p_T the lower the role of the mass, so the lower the differences between ΔR_y and ΔR_η. Even for jets at a p_T lower than the threshold of the analyses, the ± 0.1 variations cover all possible differences between the definitions.



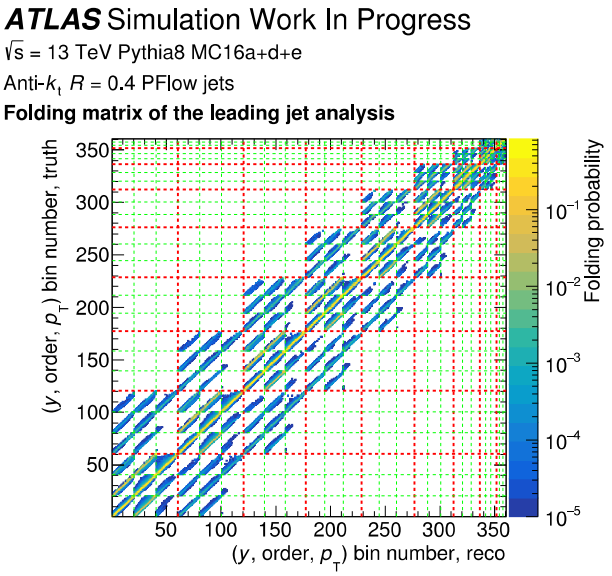
(a) Inclusive jet, TM



(b) Inclusive jet, FM

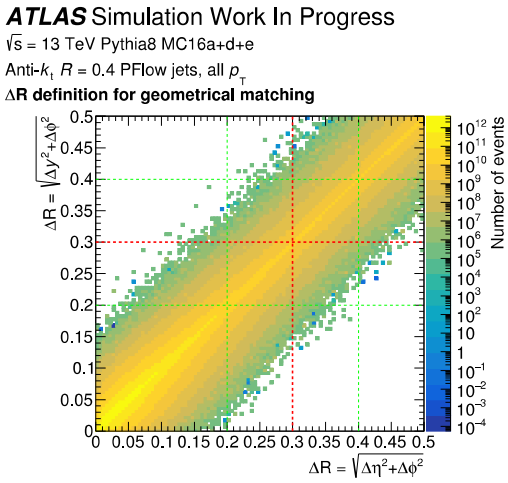


(c) Leading jet, TM

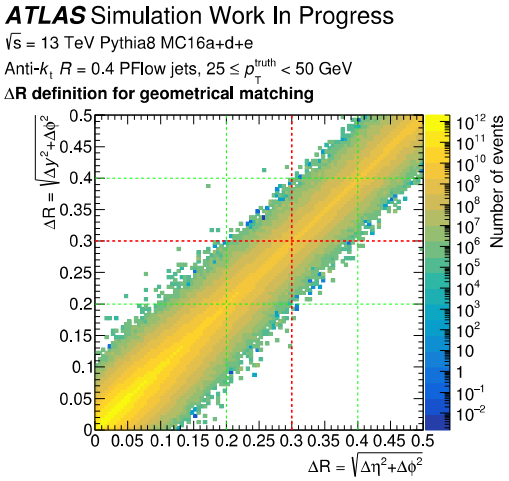


(d) Leading jet, FM

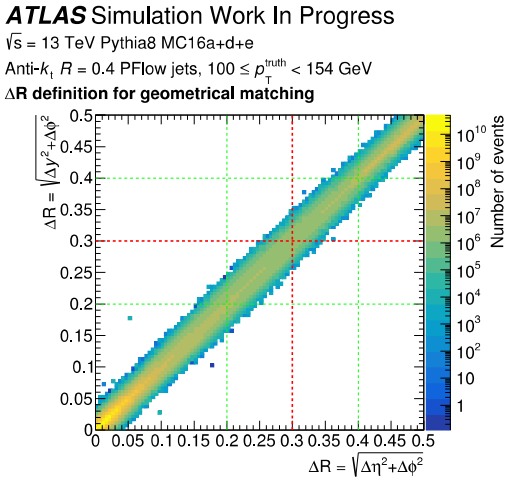
Figure 5.7: **(a)** Example of a transfer matrix for the inclusive jet analysis in the rapidity bin $|y| < 0.5$ and **(b)** the associated folding matrix. **(c)** Transfer matrix for the leading jet analysis and **(d)** the associated folding matrix. The vertical red lines separate the $|y|$ bins, the vertical green lines separate the jet orders.



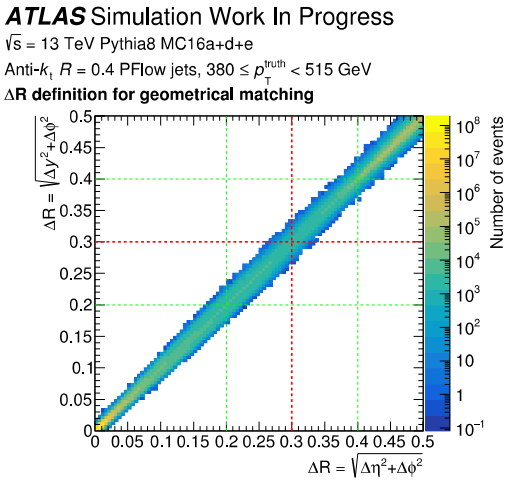
(a) Inclusive in p_T^{truth}



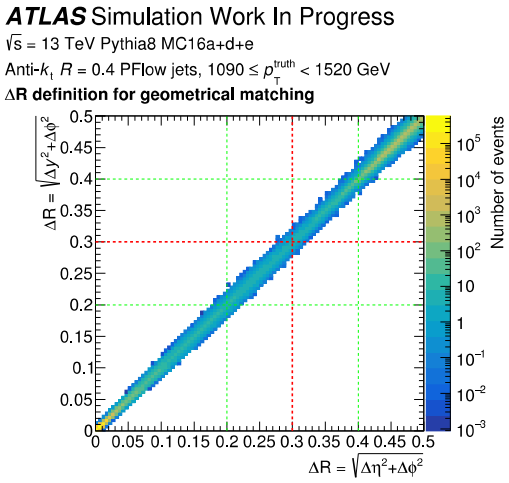
(b) $25 \le \mathrm{p_T^{truth}} < 50$ GeV



(c) $100 \le \mathrm{p_T^{truth}} < 154$ GeV

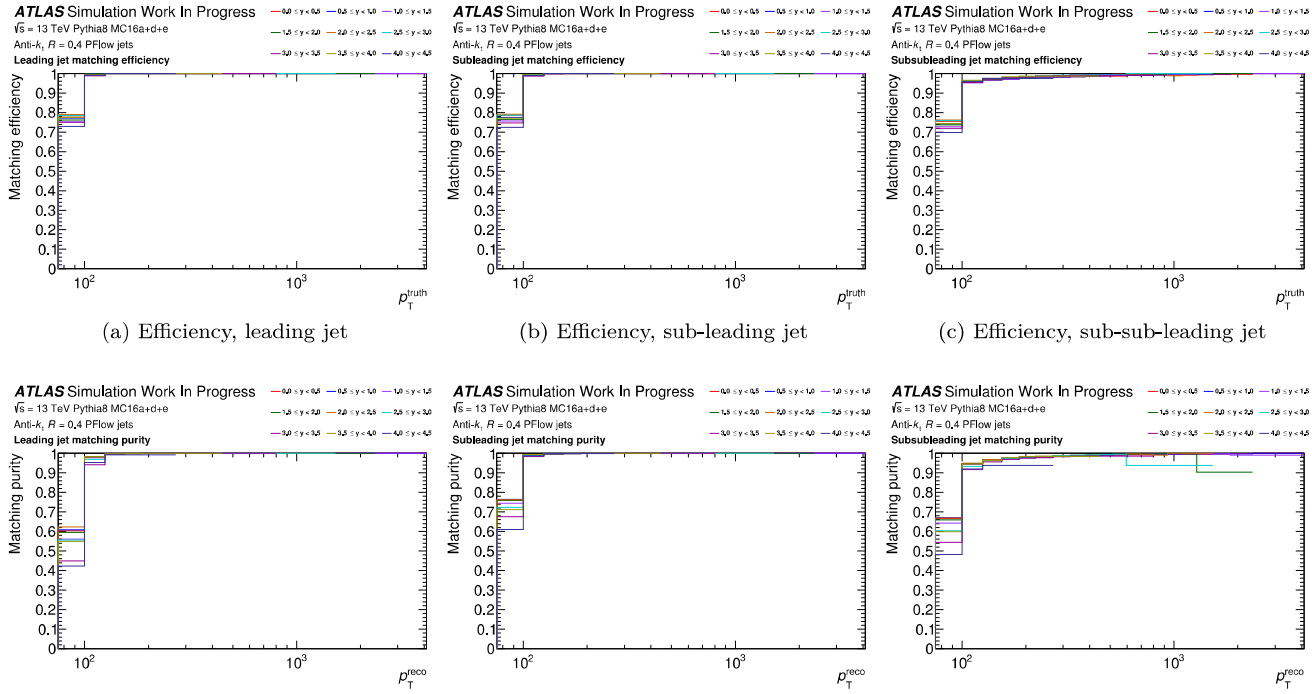


(d) $380 \le \mathrm{p_T^{truth}} < 515$ GeV



(e) $1090 \le \mathrm{p_T^{truth}} < 1520$ GeV

Figure 5.8: ΔR_y *vs.* ΔR_η for the geometrical matching. Multiple p_T^{truth} slices. MC16a+d+e.



(a) Efficiency, leading jet (b) Efficiency, sub-leading jet (c) Efficiency, sub-sub-leading jet

(d) Purity, leading jet (e) Purity, sub-leading jet (f) Purity, sub-sub-leading jet

Figure 5.9: **(a,b,c)** Matching efficiencies and **(d,e,f)** matching purities for the leading, sub-leading and sub-sub-leading jets.

5.2.2.4 Matching efficiency and purity

Some reco or truth jets are not geometrically matched to a truth or reco counterpart, or are matched to a jet that is below the low-p_T threshold, and therefore do not enter the transfer matrix. These jets are accounted for in the matching efficiency and matching purity, defined as follows:

$$\begin{aligned} \varepsilon_i^{\text{matching purity}} &= \frac{N(\text{matched reco jets})_i}{N(\text{reco jets})_i} &= \frac{\sum_j T_{ij}}{r_i} \\ \varepsilon_i^{\text{matching efficiency}} &= \frac{N(\text{matched truth jets})_j}{N(\text{truth jets})_j} &= \frac{\sum_i T_{ij}}{t_j} \end{aligned} \tag{5.10}$$

They are shown for the leading jet analysis in Fig. 5.9. The largest effect by far is the low-p_T jets that are matched to a lower-p_T jet and that do not enter the transfer matrix; for the sub-sub-leading jet, the absence of a geometrically matched jet starts to be significant.

The matching purity and efficiency are taken into account in the following way:

1. Apply a purity correction to data, to have a "matched data" distribution: $\forall i, md_i = d_i \cdot \varepsilon_i^{\text{matching purity}}$;
2. Do the unfolding on the matched data distribution md to obtain a "matched unfolded data" distribution mu;
3. Apply an efficiency correction to the matched unfolded data distribution to obtain the unfolded data distribution: $\forall i, u_i = \frac{mu_i}{\varepsilon_i^{\text{matching efficiency}}}$

5.2.3 Application of the IDS unfolding method to the inclusive jet cross-section analysis

Until now, I have only had time to perform the unfolding for the inclusive jet analysis. The method described here also applies for the leading jet analysis, though the number of iterations found to be necessary might be different.

5.2.3.1 First spectra

An example of unfolded spectrum for the inclusive jet production cross-section analysis is shown in Fig. 5.10, for the rapidity bin $|y| < 0.5$. This plot clearly shows the effect of the p_T migrations due to the p_T resolution. The corrections are stronger in the low-p_T region due to the larger resolution, and in the high-p_T region due to the steeper

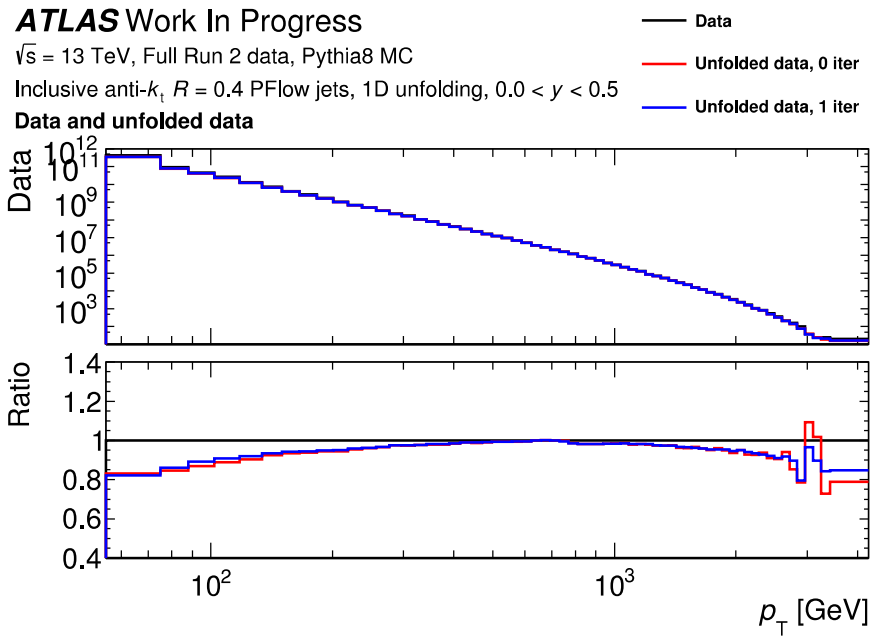


Figure 5.10: Inclusive jet cross-section spectrum in the $|y| < 0.5$ rapidity bin, either not unfolded (black) or unfolded with the IDS method without iteration (red) or with one truth-distribution improvement, *i.e.* with one iteration (blue). The lower panel shows the ratio of the unfolded spectra to the non-unfolded one.

fall of the cross-section. This plot also shows that the unfolded spectrum changes with the number of iterations; as expected, the changes are minimal when the number of iterations increases: the IDS method converges very fast, which is the reason why it has been chosen in the first place.

The number of iterations has to be chosen. This number of iterations will be used for the final nominal result, but also for the propagation of the systematic uncertainties, that have to use the exact same unfolding procedure. As already mentioned, when the number of iterations increases, the dependency on the MC decreases but the statistical uncertainties increase. The number of iterations is therefore chosen as a trade-off.

5.2.3.2 Statistical uncertainties and correlations with the bootstrap method

The bootstrap method

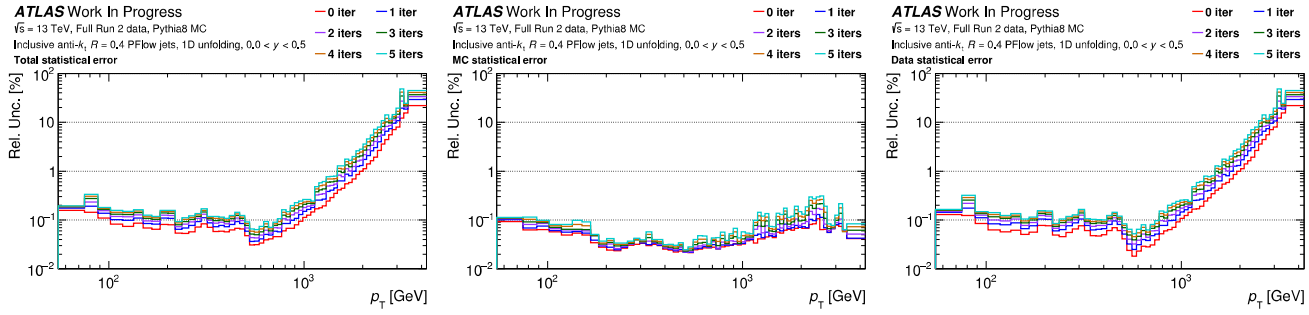
The determination of the statistical uncertainties is not straightforward:

- An analytic linear propagation of the statistical uncertainties on the inputs to the outputs is either difficult or impossible, due to the mathematical complexity of the unfolding method, and due to the complex correlations.
- The correlations between the bins of a spectrum, or between multiple measured quantities in general, arise from multiple reasons. An event is never used only once, for instance one MC event is used both at truth and reco levels, and is used to fill multiple histograms: the truth spectrum, the reco spectrum, the transfer matrix (and others), making those histograms not statistically independent. A data event is also used in multiple places, for instance to obtain the nominal unfolded spectrum and the unfolded spectrum resulting from a systematic variation (of the JES for instance), making both spectra correlated; hence the evaluation of the statistical uncertainty on a systematic uncertainty is difficult. Finally, a given bin, or measurement in general, is used multiple times: a bin in the unfolded spectrum is a (weighted) combination of all the input bins; so the unfolding introduces correlations between the bins. A smoothing would also introduce such correlations.

The solution to this problem is the "bootstrap" method of statistical uncertainty estimation [157]. This powerful technique uses a set of replicas[1] of the initial distributions, that are statistically equivalent to it. Each replica is analyzed in the same way as the nominal distribution, and provides a set of replica measurements, from which one can extract the statistical uncertainties on the nominal measurement as well as correlations between the bins. The method is based on the simple question: what if a given event had not been observed, how would it affect the distributions? What if it had been recorded twice, or three times? One replica is created by filling each event a random number of times, according to a Poisson distribution of mean 1.

More in details, the bootstrap method works as follows:

[1] Other words for "replica" are "pseudo-experiment" or "toy".



(a) Total stat uncertainty (b) Stat uncertainty from MC (c) Stat uncertainty from data

Figure 5.11: Statistical uncertainties obtained with the bootstrap method for the inclusive jet cross-section spectrum in the $|y| < 0.5$ rapidity bin, for different numbers of iterations of the IDS unfolding. **(a)** Total statistical uncertainties, using replicas of both data and MC. **(b)** Statistical uncertainties obtained with bootstrap replicas of the MC distributions only. **(c)** Statistical uncertainties obtained with bootstrap replicas of the data distributions only.

1. Compute the nominal value of interest x (say the unfolded data spectrum in the i-th bin u_i, or the systematic uncertainty in the i-th bin Δu_i) from the input histograms filled normally with the events.
2. Generate one replica: for each event, instead of simply filling it in the histograms, generate a random POISSON-distributed number $n \hookrightarrow \text{Pois}(1)$ and fill the event n times.
3. Run the unfolding on the pseudo-experiment, compute the quantity of interest
4. Repeat N times to get N replicas: for $r \in [\![1, N]\!]$, each replica r provides a different value x^r of the quantity of interest. There is a distribution of values x^r around the nominal value x.
5. Determine the half-width Δx of this distribution to obtain the statistical uncertainty on the quantity of interest: $x \pm \Delta x$. It is computed from the RMS of the replicas (or possibly more refined options like the half-width of a Gaussian fit).

The uncertainty $\Delta\Delta x$ on this estimation of the stat uncertainty Δx is proportional to $\frac{1}{\sqrt{N}}$. For the nominal result, the number of replicas is chosen to be 1000, to have an uncertainty on the stat uncertainty of 3.2 %. For the systematic uncertainties detailed hereafter, 100 replicas are enough: they provide an uncertainty on the statistical uncertainty on the systematic uncertainty of 10 %.

The bootstrap method has two more advantages: it provides statistical correlations between the bins, and it can provide correlations between the measurements of two different analyses. The correlations between two observables a and b are obtained as

$$\rho(a, b) = \frac{1}{\Delta a \Delta b} \frac{1}{N} \sum_{i=1}^{N} (a_i - \bar{a}) \left(b_i - \bar{b}\right) \tag{5.11}$$

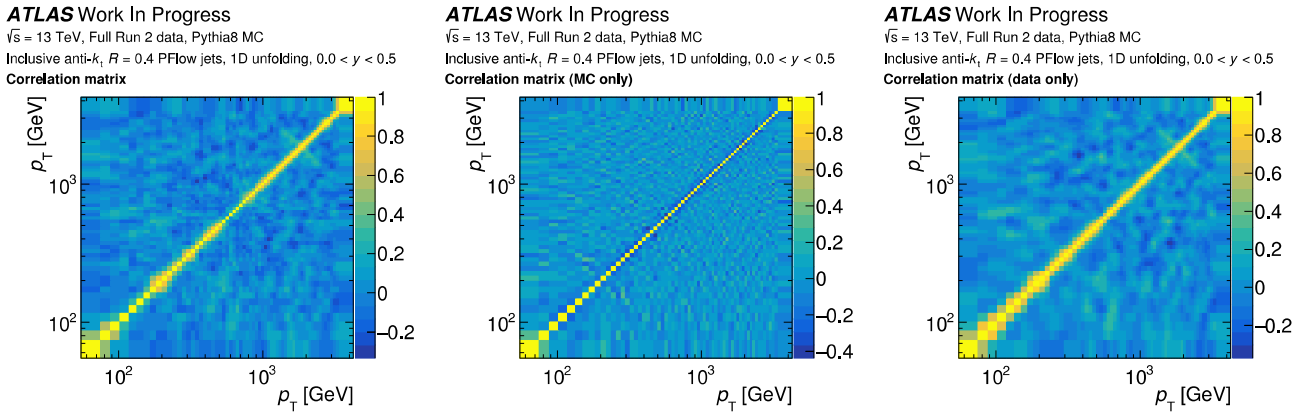
Two separate analyses might use input datasets that partially overlap, and therefore the observables a and b can come from different analyses, potentially years apart. The above correlation formula can be used as long as the replicas have been created in the same conditions: the two analyses have to use the same pseudo-random numbers to create the replicas, which is done by seeding the random number generator with the unique event identifier and with the identifier r of the replica.

Statistical uncertainties and correlations for the inclusive jet analysis

The statistical uncertainty on the unfolded spectrum is shown in Fig. 5.11a, for one rapidity bin ($|y| < 0.5$).

It is possible to vary the weights of the events from data only or from MC only, to identify where the statistical uncertainties come from, as shown in Figs. 5.11b and 5.11c. In the present case, the uncertainty at high p_T is dominated by the lack of events in data, due to the falling jet cross-section production spectrum, as expected. In the low-p_T region, the statistical uncertainty from data is close to 0.1 % due to the use of triggers and variable binning; similarly for the statistical uncertainty from MC on the whole p_T range due to the use of slices, as described in Chapter 4, Section 4.2.1.3.

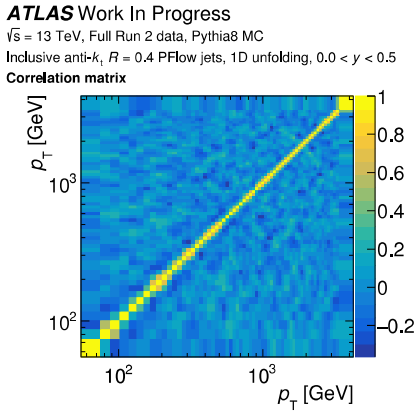
The statistical correlations are obtained and shown in Fig. 5.12 for the rapidity bin $|y| < 0.5$ and for 0 and 1 iteration of IDS unfolding, either using replicas for both data and MC, or only for MC or only for data, indicating the origin of the statistical correlations. Correlations and anti-correlations are observed near the diagonal, as expected, as the unfolding makes events move from a bin to the adjacent bins. Outside the diagonal, a rather



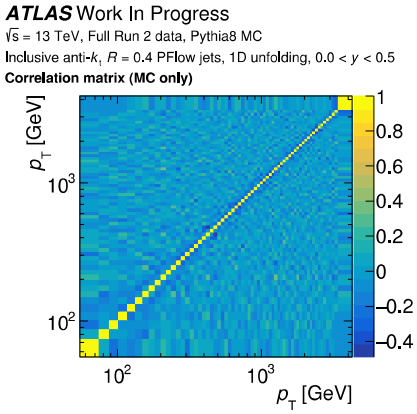
(a) 0 iteration, data and MC

(b) 0 iteration, MC only

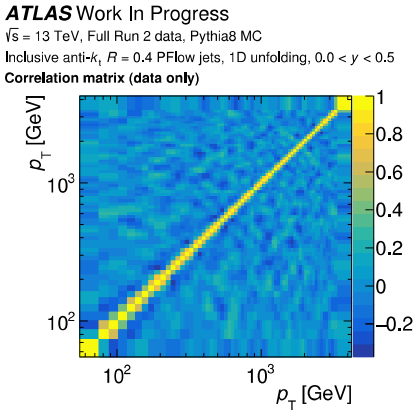
(c) 0 iteration, data only



(d) 1 iteration, data and MC



(e) 1 iteration, MC only



(f) 1 iteration, data only

Figure 5.12: Statistical correlations obtained with the bootstrap method for the inclusive jet cross-section analysis, in the $|y| < 0.5$ rapidity bin, for **(a,b,c)** 0 iteration of the IDS unfolding (*i.e.* only unfolding, no improvement of the truth spectrum) and for **(d,e,f)** 1 iteration of IDS unfolding. **(a,d)** Using replicas of both data and MC. **(b,e)** Using replicas of the MC distributions only. **(c,f)** Using replicas of the data distributions only. (Note that plot **(d)** is identical to Fig. 5.24a, reproduced here to ease the side-by-side comparison with the other plots, discussed in the text.)

uniform background of ±20 % is the statistical noise in the determination of the correlations. The range of the (anti-)correlations extends with the number of iterations, which is expected since the reweighting is based on the intermediate unfolded spectrum:

- For 0 iteration (unfolding only), the MC-only anti-correlations of 10 % to 40 % are located solely in the bins directly adjacent to the diagonal. The data-only correlations extend to two bins around the diagonal: around 50 % to 80 % in the bins adjacent to the diagonal, 20 % to 50 % in the bins adjacent to them. Due to the constant purity binning, the correlations always spread on a constant range in terms of number of bins (in this case, two bins). Due to the logarithmically growing bin sizes, the correlations span a range of approximately 10 GeV at low p_T around 100 GeV, while at high p_T around 2 TeV they span a range of approximately 100 GeV.
- For 1 iteration (unfolding, reweighting of the truth level and unfolding again), the MC-only anti-correlations have not changed much, while the figure is quite different for the data-only (anti-)correlations: the p_T range of the (anti-)correlations has extended, the correlations in the bins near the diagonal are lower, and the anti-correlations have appeared in the 1 or 2 bins after the bins adjacent to the diagonal. Those anti-correlations are of the order of 20 % to 35 %, and the p_T range is of 30 GeV around 100 GeV and a few hundreds of GeV around 2 TeV.

For the MC, a fluctuation of an entry in the transfer matrix amounts to a fluctuation of migrations between the bins: an upwards fluctuation of the diagonal (dominant) term in the transfer matrix amounts to less migrations to the adjacent bins, therefore to anti-correlations near the diagonal. In data, when a bin fluctuates up, it means that it contains more events, that will migrate to the adjacent bins with unfolding, hence the (positive) correlations in the neighboring bins. The distribution of data also impacts the transfer matrix through the reweighting, hence the appearance of anti-correlations after one iteration.

5.2.3.3 Bias related to the sensitivity of the unfolded method to the data/MC shape differences

MC simulations do not perfectly render the shape of the data distribution. Therefore the unfolding procedure might over-correct or under-correct some regions of data, resulting in a bias of the estimated unfolded data spectrum. This bias is iteratively reduced by the improvement of the MC truth spectrum, and therefore the IDS method is less and less sensitive to the data/MC shape differences when the number of iterations increases. A quantitative and data-driven way to estimate the bias related to the sensitivity of the unfolding method to the data/MC shape differences is presented here [152].[2]

The idea is to apply the unfolding procedure to a distribution r' that displays the same global features as the data distribution and for which a corresponding truth distribution t' is known, and to compare the unfolded data-like reco distribution u' to the associated truth distribution t'.

More in detail, the following procedure is applied:

1. Create the modified distributions r' and t' by a reweighting of the truth distribution t by a smooth function f, $t'_i = t_i \cdot f(p_T^i)$, and by applying the folding to obtain $r' = F \times t'$.
2. r' and t' are normalized to have the same number of events as d, and the statistical uncertainties of r', $\Delta r'_i$, are set to be those of data Δd_i.
3. Do the unfolding on the modified reco distribution r' (acting as pseudo-data) using the unchanged transfer matrix T, to obtain the unfolded modified reco distribution u'.
4. Compare to the modified truth distribution t': the bias is $b = \frac{u'-t'}{t'}$.

The reweighting function f

Obtaining the function f is tricky though. f is taken to be a polynomial, which degree N depends on the number of bins of the distribution:

$$f(p_T) = \sum_{i=0}^{N} c_i \left(\log p_T - \log p_T^0\right)^i \tag{5.12}$$

with p_T^0 the lower bound of the lowest-p_T bin of the analysis: $p_T^0 = 55$ GeV in case of the inclusive jet analysis, $p_T^0 = 70$ GeV for the leading jet analysis.

Fitting the $\frac{d}{r}$ distribution is not enough[3] to obtain f: f should be the optimal reweighting function of t such that r' is close to d, with r' obtained through folding. The reweighting function obtained with those two different

[2] It is often referred to as "basic unfolding uncertainty", "data-driven (non-)closure test" or "unfolding bias", while I will rather call it "shape bias" for clarity.

[3] This is sufficient in cases where the level of migrations is lower.

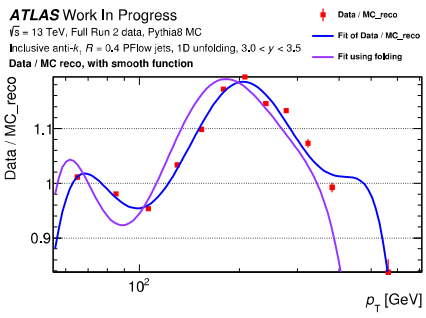


Figure 5.13: Reweighting function to obtain a modified reco distribution r' that minimizes the data/MC reco shape differences, in the $|y| < 0.5$ rapidity bin. $\frac{d_i}{r_i}$ distribution (before reweighting) directly fitted by the polynomial (blue) and compared to the reweighting function f obtained through folding with the fine bins transfer matrix (purple). The differences between the two illustrate why f cannot be fitted on $\frac{d}{r}$ and why folding is necessary.

methods, shown in Fig. 5.13, are quite different. What is minimized is therefore

$$\tilde{\chi}^2 = \sum_i \frac{(d_i - r'_i)^2}{\Delta d_i{}^2 + \Delta r'_i{}^2} \tag{5.13}$$

where i is the index of the p_T bin, and the index of the rapidity bin is dropped for clarity, since the unfolding is one-dimensional in p_T. Note that the $\tilde{\chi}^2$ is used to provide a reweighting function f, but that different functions than the $\tilde{\chi}^2$ could be used since different criteria are used to evaluate whether the minimum is satisfactory or not, as discussed later. The data/MC reco shape differences before and after reweighting of the truth distribution by f are shown in Fig. 5.14a and Fig. 5.14d respectively, for the rapidity bin $|y| < 0.5$; the local $\tilde{\chi}^2$ before and after minimization is shown in Fig. 5.14b and Fig. 5.14e respectively.

For simplicity, since a transfer matrix gives access both to the reco and truth distributions by projection along its axes, the reweighting function f can be applied directly to the transfer matrix, yielding the modified transfer matrix T', and the $\tilde{\chi}^2$ is computed as follows:

$$r_i = \sum_k F_{ik} \cdot t_k = \sum_k T_{ik} \tag{5.14}$$

$$\implies \Delta r_i = \sqrt{\sum_k \Delta T_{ik}{}^2} \tag{5.15}$$

$$r'_i = \sum_k F_{ik} \cdot t'_k = \sum_k F_{ik} \cdot t_k \cdot f\left(p_T^k\right) = \sum_k T_{ik} \cdot f\left(p_T^k\right) = \sum_k T'_{ik} \tag{5.16}$$

$$\implies \Delta r'_i = \sqrt{\sum_k \Delta T'_{ik}{}^2} = \sqrt{\sum_k \Delta T_{ik}{}^2 \cdot f\left(p_T^k\right)^2} \tag{5.17}$$

$$\implies \tilde{\chi}^2 = \sum_i \frac{(d_i - r'_i)^2}{\Delta d_i{}^2 + \Delta r'_i{}^2} = \sum_i \frac{\left(d_i - \sum_k T_{ik} \cdot f\left(p_T^k\right)\right)^2}{\Delta d_i{}^2 + \sum_k \Delta T_{ik}{}^2 \cdot f\left(p_T^k\right)^2} \tag{5.18}$$

One value is lacking to use the previous formula: the value of p_T^k at which to evaluate f. Ideally, the reweighting would be applied directly on a jet-by-jet basis. This is technically complicated. But the function cannot be directly computed and applied on the large bins of the analysis either! Indeed, both the cross-section and the reweighting function vary rapidly, especially in the high-rapidity bins $3.0 < |y| < 4.5$. So the mean in one p_T bin $\langle t \cdot f \rangle$, which is what we would have if the function was applied on a jet-by-jet basis, is different from $\langle t \rangle \cdot \langle f \rangle$ which is (approximately) what we have (more precisely $\langle t \rangle \cdot f(p_T^k)$, with p_T^k still unspecified yet, that could be the bin center or its barycenter for instance). This results in a bias.

The compromise is to apply the reweighting function on a transfer matrix obtained with fine bins, that are then merged into the normal larger binning. The bins of the "fine bins transfer matrix" should be finer than those of the "normal bins transfer matrix" by at least an order of magnitude, so equally-spaced 1 GeV bins are satisfactory. Such a transfer matrix with fine bins is shown in Fig. 5.15 for one rapidity bin. Then, f is simply evaluated at the center of the fine bin.

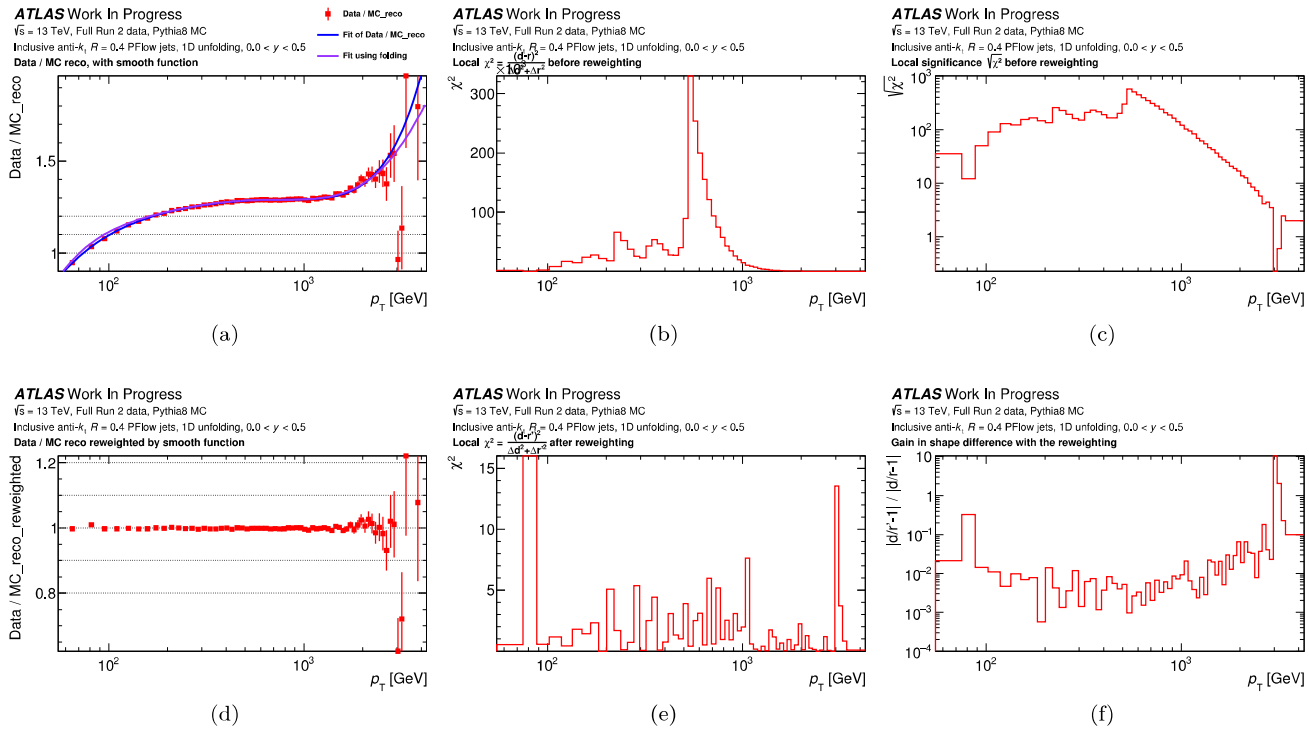
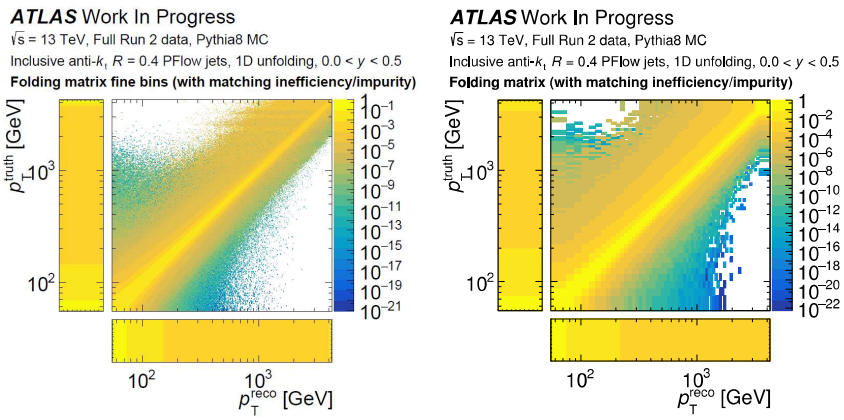


Figure 5.14: Reweighting to obtain a modified reco distribution r' that minimizes the data/MC reco shape differences, and satisfaction criterion for the bias estimation for the inclusive jet cross-section analysis, in the $|y| < 0.5$ rapidity bin. **(a)** $\frac{d_i}{r_i}$ distribution (before reweighting) directly fitted by the polynomial (blue) and compared to the reweighting function f obtained through folding with the fine bins transfer matrix (purple). The differences between the two illustrate why folding is necessary and are better seen on other rapidity bins. **(b)** Local $\tilde{\chi}_i{}^2 = \frac{(d_i - r_i)^2}{\Delta d_i{}^2 + \Delta r_i{}^2}$ before reweighting. The spikes reflect the available statistics in the JZ slices (*i.e.* the number of events in the p_T slices of MC events, explained in Section 4.2.1.3), and the smoothly falling right part of the distribution reflects the falling cross-section. **(c)** Local significance $\sqrt{\tilde{\chi}_i{}^2}$ before reweighting, where the regions above 2 are those in which the gain in shape difference must be better than a factor of 10. **(d)** $\frac{d_i}{r'_i}$ distribution after reweighting of the truth distribution by f and folding through the transfer matrix to obtain r'. **(e)** Local $\tilde{\chi}_i{}^2 = \frac{(d_i - r'_i)^2}{\Delta d_i{}^2 + \Delta r'_i{}^2}$ after reweighting, indicating that there is no region that seems incorrectly fitted, the remaining spikes look like statistical fluctuations. **(f)** Relative residual data/MC reco difference $\left|\frac{d_i}{r'_i} - 1\right| \Big/ \left|\frac{d_i}{r_i} - 1\right|$; the inverse of this is the gain in shape difference with the reweighting: a residual difference of 0.1 is equivalent to a gain of a factor 10. For the function found by the minimization on $\tilde{\chi}^2$ to be satisfactory, any bin i in **(c)** for which $\sqrt{\tilde{\chi}_i{}^2} > 2$, the gain in shape difference in **(f)** has to be larger than 10.



(a) (b)

Figure 5.15: Folding matrices for the inclusive jet cross-section measurement in the $|y| < 0.5$ bin, with **(a)** fine bins or **(b)** normal "large" bins.

I have optimized the choice of the number of parameters of f with an old version of the histograms, using EMTopo jets. However, the number of parameters is not expected to change with the new version of the histograms using PFlow jets; furthermore, this choice is somewhat arbitrary.

The number of parameters has been increased until either the $\frac{\tilde{\chi}^2}{N_{\text{dof}}}$ is not improved any more or a minimum of $N_{\text{dof}} = 1$ is reached, with $N_{\text{dof}} = N_{\text{bins}} - N_{\text{parameters of } f}$. Eight parameters have been chosen for $0.0 < |y| < 4.0$ because the $\frac{\tilde{\chi}^2}{N_{\text{dof}}}$ does not improve for more parameters, and four parameters have been chosen for $4.0 < |y| < 4.5$ because there are only 5 bins in this rapidity region.

Is the function f found by the minimization on $\tilde{\chi}^2$ satisfactory?

The satisfaction criterion is the following one: in the p_T bins in which the data and MC reco points were significantly different, the difference must be reduced by about a factor 10. This means that the uncertainty on the shape bias is of the order of 10 %, which is acceptable for a bias estimation that is rather qualitative, and not dominant anyways. The bins that are far from 1 but not significantly might get an overestimated bias, but it is fine because their statistical uncertainty is bad already (it is a data-driven test), because the non-closure shape bias is small anyways, and because it is conservative.

Quantitatively, this means that in the p_T bins where the $\frac{d}{r}$ points were away from 1 with a significance higher than 1.5 to 2 σ, the $\frac{d}{r'}$ ratio should be at least 10 times closer from 1 than $\frac{d}{r}$ was. In this context, the significance is defined as

$$\sqrt{\frac{\left(\left(\frac{d}{r'}\right)_i - 1\right)^2}{\Delta\left(\frac{d}{r'}\right)_i^2}} = \sqrt{\frac{\left(\frac{d_i - r'_i}{r'_i}\right)^2}{\frac{1}{r'^2_i}\left(\Delta d_i^{\,2} + \frac{d_i^2}{r'^2_i}\Delta r'^2_i\right)}} = \sqrt{\frac{(d_i - r'_i)^2}{\Delta d_i^{\,2} + \frac{d_i^2}{r'^2_i}\Delta r'^2_i}} \approx \sqrt{\chi_i^{\,2}} \text{ if } d_i \approx r'_i \tag{5.19}$$

Therefore, the criterion is:

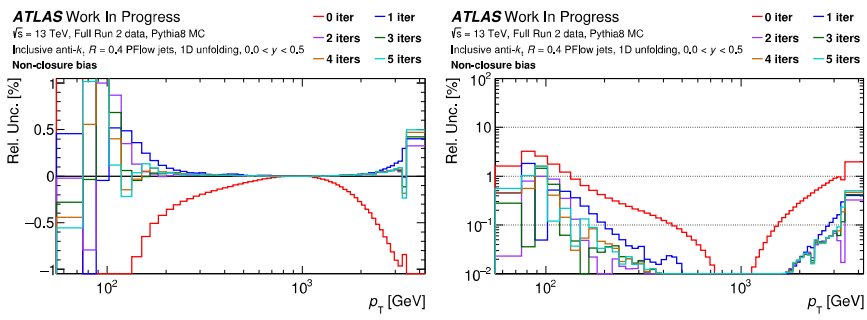
$$\forall i, \sqrt{\tilde{\chi}_i^{\,2}} > 2 \Rightarrow \frac{\left|\frac{d_i}{r'_i} - 1\right|}{\left|\frac{d_i}{r_i} - 1\right|} < \frac{1}{10} \tag{5.20}$$

The local significance is shown in Fig. 5.14c, and the gain in shape difference is shown in Fig. 5.14f.

Another important criterion is that the truth distribution should also remain smooth; this criterion has not been made quantitative, and is thus simply evaluated by looking at the t' distributions.

Since the satisfaction criterion is not having a $\frac{\tilde{\chi}^2}{N_{\text{dof}}} \approx 1$, the function that is minimized could be improved to be the actual satisfaction criterion. This is however not necessary: the approach of minimizing on the $\tilde{\chi}^2$ to find a function f and then to check that the function is satisfactory using another criterion is fine, and is shown to provide good enough reweighting functions.

As a side note, the $\tilde{\chi}^2$ computed here is not meaningful to provide information on the tensions between data and MC, because the bin-to-bin correlations are not taken into account when computing $\tilde{\chi}^2$, and because only the statistical uncertainty is considered and not all the systematic uncertainties. This $\tilde{\chi}^2$ is really only used as an intermediate variable to find a reweighting function, without any deep meaning.



(a) (b)

Figure 5.16: Estimation of the bias b related to the sensitivity of the unfolded method to the data/MC shape differences for the inclusive jet cross-section analysis, in the $|y| < 0.5$ rapidity bin, for different values of the number of iterations of the IDS unfolding. **(a)** Linear scale and **(b)** $|b|$ in logarithmic scale.

There are a few bins for which the gain is less than a factor of 10 in regions where the deviation was significantly far from 1. They all have been checked not to be problematic. Indeed:

- If the gain is of a factor of 5, it is still acceptable if those regions are not too extended.
- For the high p_T bins, they all have an important statistical error already, so having an incorrectly evaluated bias is not problematic, since the bias uncertainty will not end up being dominant.
- For the mid-range p_T bins, the gain is not of a factor 10 but they were already close to 1 and the data/MC shapes were already similar.
- For the lowest-p_T bins concerned by large data/MC differences, they are below 100 GeV, so their bias will not impact the region of interest for physics studies (*i.e.* the one included in the publication).
- The goal is to capture the shape of the distribution, so the absolute value is not so important. This last criterion could be made more quantitative: the derivative of the $\frac{d}{r'}$ distribution could be looked at locally, and globally the curvature could be computed. One could also look at the minimum and maximum differences between $\frac{d}{r}$ and 1 and check that all the differences after reweighting are of the order of magnitude of the minimum from before reweighting.

All these reasons show that the bias estimation is, in the end, rather qualitative.

To conclude, the reweighting of the truth distribution in fine bins by a 7-degree polynomial (or 3-degree in the last rapidity bin) correctly captures the shape differences between the data and reco distributions. This makes me confident that the shape bias is correctly evaluated, and not underestimated.

The bias related to the sensitivity of the unfolded method to the data/MC shape differences is shown in Fig. 5.16 for the rapidity bin $|y| < 0.5$. As expected, the bias is reduced when the number of iterations increases.

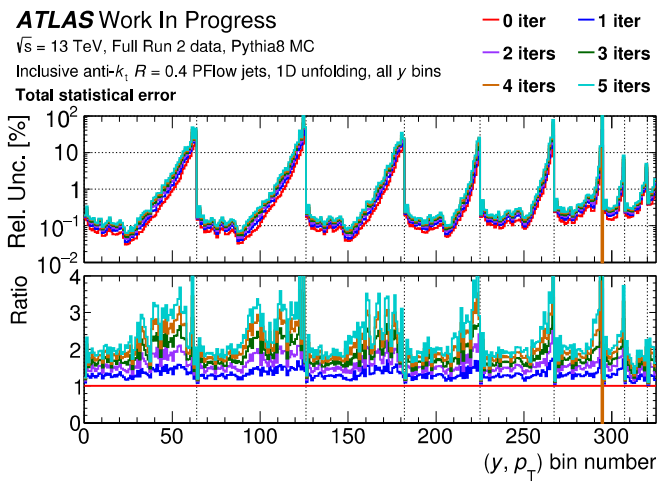
5.2.3.4 Choice of the number of iterations

As discussed earlier, increasing the number of iterations reduces the bias related to the sensitivity of the unfolded method to the data/MC shape differences, as shown in Fig. 5.16, and increases the statistical uncertainty, as shown in Fig. 5.11a. The sum in quadrature of the two, shown in Figs. 5.17 and 5.23, reveals that the improvement in shape bias is dominant at low p_T while the increase in statistical uncertainty is dominant at high p_T.

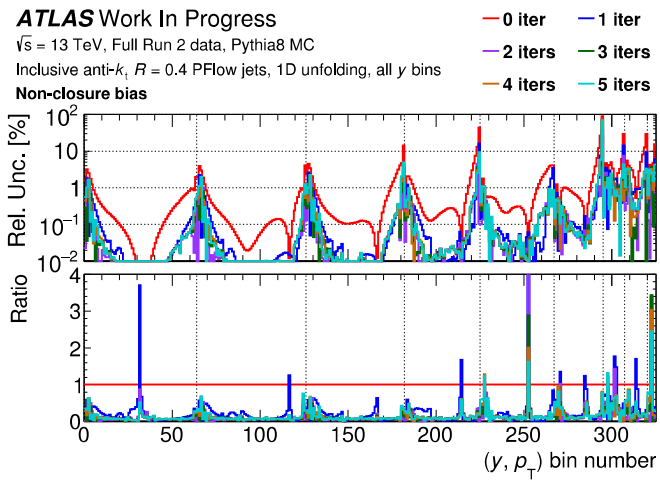
The balance between the two is as follows.

1. The bias is negligible in most phase space regions compared to the overall total uncertainties, including the other systematic uncertainties discussed hereafter, that are of the order of 1 % minimum. The bias is a lot smaller after one iteration: one order of magnitude on a large portion of the p_T range in each rapidity bin. The gain in bias is negligible after one iteration, 2 iterations or more are unnecessary. So the choice is between no iteration and one iteration.
2. It is preferable to have the analysis dominated by statistical uncertainties, that are determined in a rather reliable way, with a clear statistical meaning compared to the bias which estimation is more qualitative (although here it is data-driven, which makes it quite reliable). So decreasing the shape bias is preferable, as long as the increase in statistical uncertainty is acceptable, which it is because the increase in statistical uncertainty with 1 iteration is not too big (a factor 2 maximum, approximately 1.3 on average) and because the analysis will be dominated by the JES and JER uncertainties anyways.

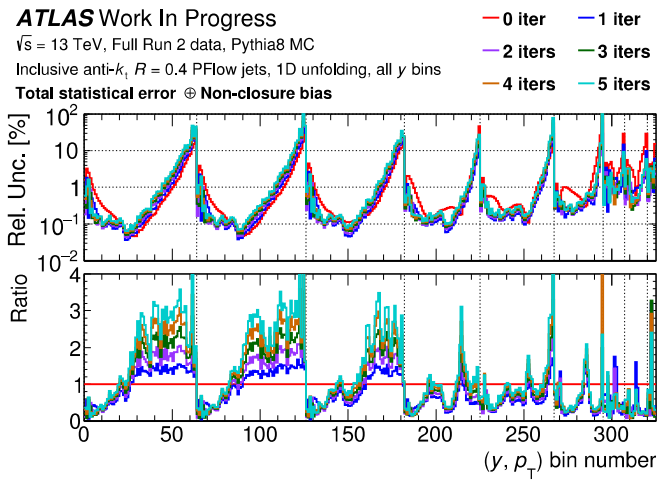
Therefore, 1 iteration of Iterative Dynamically Stabilized unfolding is chosen for the inclusive jet analysis.



(a)



(b)



(c)

Figure 5.17: Impact of the number of iterations of IDS unfolding on the statistical uncertainties estimated with the bootstrap method and on the bias related to the sensitivity of the unfolded method to the data/MC shape differences, for the inclusive jet cross-section analysis, for all the rapidity bins separated by red vertical dashed lines. **(a)** Statistical uncertainties. **(b)** Shape bias. **(c)** Sum in quadrature of the statistical uncertainties and of the shape bias; these sums in quadrature are presented bin-by-bin, in a more readable way, in Fig. 5.23.

5.3 Experimental uncertainties

Multiple sources of uncertainties affect the cross-section measurements, of different types and origins:

- Statistical uncertainties: the number of events in the data and MC samples is not infinite, limiting the precision. They are evaluated with the bootstrap method, and already discussed in Section 5.2.3.2.
- The bias related to the sensitivity of the IDS unfolding method to the data/MC shape differences is evaluated with a data-driven test already discussed in Section 5.2.3.3.
- The jet observables are calibrated with the Jet Energy Scale (JES), and the data/MC Jet Energy Resolution (JER) agreement is improved with the smearing of MC. Both the JES and the JER come with uncertainties, that have to be propagated through the unfolding.
- The JER is evaluated and applied as the first Gaussian of the resolution distributions, but the tails are non-Gaussian, which has an impact for which there is no associated uncertainty from the *in situ* measurements. This impact has been measured at the moment of unfolding for the first time in CMS [143, 147]. I have evaluated this uncertainty for the first time in ATLAS cross-section measurements.
- Event selection uncertainties: the selection criteria can be varied, and an uncertainty on the efficiency corrections (for the trigger inefficiency and for the jet cleaning inefficiency) can be determined.
- The luminosity intervenes in the measurement in Eq. (5.2), and therefore the associated uncertainty plays a role. An uncertainty on the knowledge of the beam energy is also evaluated.

Note that this is work in progress: my update of the leading jet cross-section measurement, done by Robert Hankache for the first time in his Ph.D. thesis in 2019 [103], has not reached the stage of the unfolding, and the inclusive jet cross-section measurement has not evaluated all the uncertainties yet.

Having already talked about the statistical uncertainties and the bias due to the sensitivity to shape differences, this section will detail the other sources of uncertainties, with a focus on the new JER non-Gaussianity uncertainty.

5.3.1 Propagation of the uncertainties on the JES and the JER

Category	Statistical uncertainty		Systematic uncertainty	
	2015+2016	2017 + 2018	all	2018
η-intercalibration	532×2	532×2	4×2	$+2$
"*In situ* propagation"	2			
Flavor			8×2	
γ+jet	16×2		8×2	
Multijet balance	28×2		6×2	
pile-up			4×2	
Punch through			2	
Single particle at high p_T			2	
Z+jet	28×2		12×2	
JER	12×2		22×2	
Total: $1216 \times 2 = 2432$	1149×2		67×2	

Table 5.3: Origin of the systematic uncertainties. The uncertainties of the JER are isolated from the uncertainties from the JES. The η-intercalibration has two sets of uncertainties for the different data-taking periods and a non-closure uncertainty to apply the 2017 calibration on 2018 data. The "*in situ* propagation" refers to the propagation of the total statistical uncertainty of the η-intercalibration through the other *in situ* analyses. The $\times 2$ everywhere accounts for the up and down variations.

All components of the JES and JER uncertainties, detailed in Chapter 3, Section 3.3 and mentioned in Table 5.3, are individually propagated through the unfolding with the procedures described hereafter. A detailed list of all the 2432 propagated components is provided in Appendix B.1. The resulting uncertainties grouped by category (summed in quadrature within each category) are shown in Fig. 5.26.

Jet Energy Scale

The JES is simply varied by $+1\ \sigma$ or $-1\ \sigma$, the histograms are filled again, the unfolding is done again, then the uncertainty is simply the $\Delta u = \frac{u^{\text{varied}}}{u^{\text{nominal}}} - 1$. Although the JES uncertainties are symmetrical, both the up and down variations are propagated individually because the unfolding makes them asymmetrical, mostly due to the non-linearly falling shape of the cross-section. Each variation is made coherently on the whole $(p_T, |y|)$ range, and is

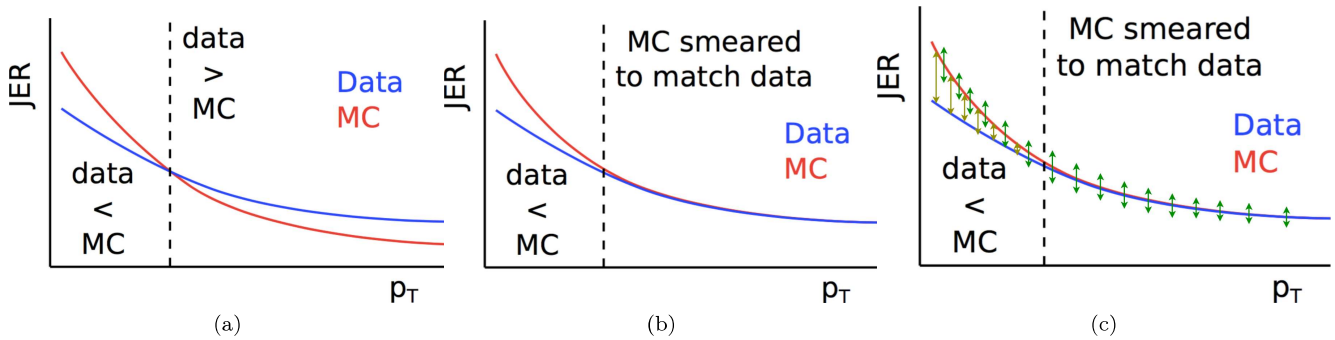


Figure 5.18: **(a)** Illustration of the case where the JER of MC and data cross each other: in some p_T region the resolution of data is better than that of MC, and *vice versa*. **(b)** In case the JER is better in MC than in data, the resolution of MC is degraded by a Gaussian smearing to match the resolution of data, for a better agreement between the two, but nothing is done when the JER is better in data than in MC. **(c)** In the latter case, the difference between the JER of data and MC is taken as a JER uncertainty, illustrated by the brown arrows, that needs to be propagated through the unfolding. In both cases, the uncertainties on the JER illustrated by the green arrows also need to be propagated through the unfolding, either by a degradation of the resolution in MC or in pseudo-data. Reproduced from [158].

treated as fully correlated across the phase space since the variation is caused by a unique source (be it of physical origin, statistical origin, *etc*), but independent of the other systematic uncertainties.[4]

These uncertainties evaluate the relative effect of a change in the calibration, and therefore do not necessarily need to be evaluated on data but can be evaluated with the MC sample (acting as "pseudo-data"). This allows to have a better statistics, especially at high p_T. In this case, MC reco is unfolded as if it were data, then MC reco is calibrated with a varied calibration and unfolded again as if it were data, and finally both unfolded spectra are compared to obtain the systematic uncertainty.

The complete set of uncertainties of the JES is used, instead of the "reduced schemes" proposed by the JetETmiss group. Indeed, this precision analysis has been shown to have a strong dependency on individual components, and this is especially true for the thousands of statistical uncertainties from the η-intercalibration. Each (p_T^{avg}, η) bin of the η-intercalibration is affected by a statistical uncertainty, then the smoothing (described in Section 4.2.7.3) makes those local uncertainties "leak" onto the adjacent bins. Only using the sum in quadrature of the statistical uncertainties of the η-intercalibration has been shown to increase the data/theory χ^2 disagreement by 200 units, because it would then be only one single fully-correlated uncertainty. All the individual components allow to take the local fluctuations into account and thus reduce the χ^2. As will be discussed later in Section 5.5.2, using non-smoothed η-intercalibration statistical uncertainties helps to reduce the χ^2 even further.

Jet Energy Resolution

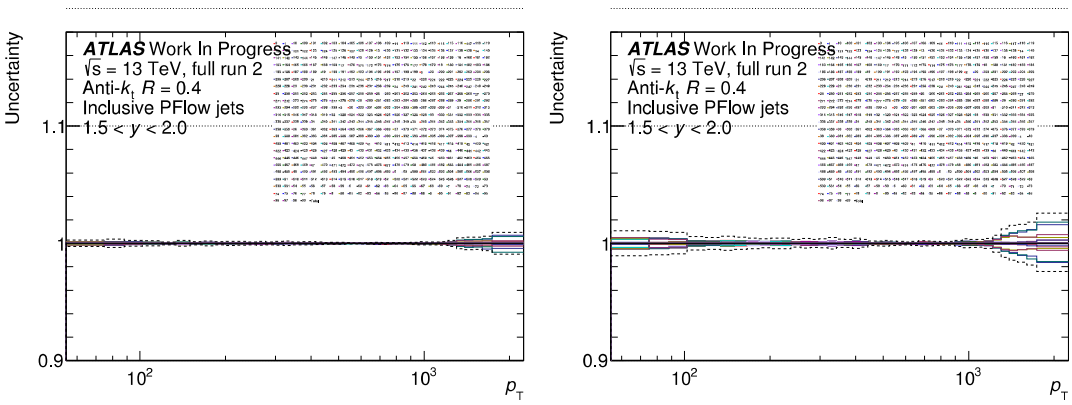
The JER provides uncertainties from two different sources.

First, a "JER uncertainty" comes from the difference between the nominal JER measured in data and in MC. Indeed, during the calibration phase described in Section 5.1, the resolution of MC reco has been smeared in case it was better in MC than in data, but not smeared in case it is better in data than in MC, as illustrated in Fig. 5.18. In the latter case, the difference $\sigma_{\text{nominal,data}} - \sigma_{\text{nominal,MC}}$ illustrated by the brown arrows in Fig. 5.18c is simply taken as an additional JER uncertainty, and is propagated using the procedure described hereafter.

Second, the systematic uncertainties on the resolution are propagated. They are illustrated by the green arrows in Fig. 5.18c. This is done in a different way from the JES uncertainties, as the resolution is accounted for in the folding matrix. The jet p_T is "smeared" by a Gaussian function of variance $\sigma^2_{\text{smear}} = (\sigma_{\text{nominal}} - \Delta\sigma_{\text{uncertainty}})^2 - \sigma^2_{\text{nominal}}$. More precisely, a random number is drawn according to this Gaussian distribution and multiplied with the p_T of the jet:

- In case the resolution needs to be degraded compared to the nominal JER, *i.e.* when $\Delta\sigma_{\text{uncertainty}} > 0$, the p_T of the events entering the transfer matrix is smeared.
- In case the resolution needs to be improved compared to the nominal JER, *i.e.* when $\Delta\sigma_{\text{uncertainty}} > 0$, the p_T of the events in the transfer matrix cannot be "anti-smeared". An equivalent effect is reached by smearing

[4]Note that a "statistical uncertainty" on the JES is now called a "systematic uncertainty" for the cross-section analysis; this is a standard convention. Hopefully it will not get confusing when we will start to evaluate statistical uncertainties on these systematic uncertainties propagated through the unfolding... Let alone when we will talk about the uncertainty on the statistical uncertainty on these systematic uncertainties... I tried to keep it clear.



(a) (b)

Figure 5.19: Systematic uncertainties coming from the statistical uncertainties of the η-intercalibration, propagated through the unfolding for the inclusive jet cross-section measurement, in the $1.5 < |y| < 2.0$ rapidity bin. Per-year uncertainties are given the appropriate weight which is the luminosity recorded during the data-taking period, which is why the uncertainties for 2017 (applied onto the 2017 and 2018 datasets) are larger than those of 2015+2016.

the p_T of the jets in the pseudo-data distribution.

Therefore, each JER systematic uncertainty requires not only a modified pseudo-data spectrum, but also a modified transfer matrix, that need to be used jointly to evaluate the systematic uncertainty.

Per-year uncertainties

The Run 2 period is covered by two sets of η-intercalibrations, one for the 2015+2016 period and one for the 2017+2018 period, each having their sets of statistical uncertainties, plus one extra non-closure uncertainty for 2018.[5]

The recorded luminosity is different for the different data-taking periods and is given in Table 5.1, so the uncertainties for the 2015+2016 period should not be given the same weight as those for the 2017+2018 period; more precisely, the uncertainty on one year must have an impact that is proportional to the luminosity recorded in that year. The goal being to evaluate the impact of the per-year uncertainty not on the per-year dataset but on the total dataset. The technical solution I retained is simply to multiply the MC weights by the luminosity (which is possible because MC events are weighted to cross-section units), such that the weights of MC events is expressed in numbers of events, which is the natural unit of data. Some uncertainties only exist for one period for the η-intercalibration; in this case, the histograms produced for one year with a variation of the calibration are summed with the nominal histograms of the other periods. The impact of the per-year weight on the statistical uncertainties of the η-intercalibration is exemplified by the comparison between the 2016+2016 and the 2017 uncertainties on the inclusive jet cross-section measurement in Fig. 5.19.

Statistical uncertainties on the systematic uncertainties

The systematic uncertainties are rebinned until they become statistically significant in each bin, or discarded if they never become significant by rebinning. Finally, the systematic uncertainties are smoothed, to avoid local statistical fluctuations, that have an impact on the computation of the data/theory χ^2.

The statistical uncertainty on a systematic uncertainty propagated through the unfolding is evaluated thanks to the bootstrap method (described in Section 5.2.3.2) that allows to take into account the correlations between the results with nominal and varied calibrations. 100 bootstrap replicas are used for the systematic uncertainties, both for the JES and the JER (including replicas of the modified transfer matrices in the latter case), such that the uncertainty on the statistical uncertainty on the systematic uncertainty is of 10 %, which is good enough for the rebinning and the smoothing. In case of systematic uncertainties that come from the η-intercalibration statistical uncertainties, a problem arises: 100 replicas for those 2124 uncertainties is not computationally feasible. This is actually not a problem: the η-intercalibration statistical uncertainties are very localized in $(p_T, |y|)$, as shown in Fig. 4.14, and therefore do not need rebinning or smoothing, especially in the case of the non-smoothed η-intercalibration statistical uncertainties. Therefore, the bootstrap method is not used for the η-intercalibration uncertainties of statistical origin.

[5] I have derived the η-intercalibration independently for 2015+2016, 2017 and 2018. Discussions are ongoing on how to use my new η-intercalibration for this analysis, which is not obvious since the rest of the *in situ* calibration chain after the η-intercalibration has not been fully run due to time constraints.

5.3.2 Non-Gaussianity of the JER

The JER is evaluated *in situ*, as the width of the first Gaussian of distributions, as detailed in Chapter 3, Section 3.3.3, and is used for the smearing of MC quantities to make them match those of data quantities, again using the assumption that the distributions are Gaussian. But we know that the tails of the distributions are non-Gaussian, and there is no *in situ* evaluation of their effect. Is the impact negligible or not? Is a systematic uncertainty enough to cover for this effect? How to estimate that systematic uncertainty?

I have evaluated a systematic uncertainty for this effects for the first time for ATLAS jet cross-sections.

The transfer matrix is constructed with jets that are geometrically matched between truth and reco levels within $\Delta R < 0.3$. It gives access to non-Gaussian effects in the off-diagonal terms. The idea is then to compare the result of the unfolding using a "Gaussian transfer matrix" to the nominal unfolding using the normal transfer matrix that contains the non-Gaussian effects, which will provide an estimate of this effect, used as an uncertainty.

A few technical details matter:

- The Gaussian transfer matrix is built from the convolution of the JER to the MC truth distribution. More precisely, to get the normalization right, the same truth distribution has to be used as the one in the normal transfer matrix, which is the "matched truth" distribution, *i.e.* the distribution of truth events that have been geometrically matched to reco events, or in other words the projection of the normal transfer matrix onto the truth axis.
- Ideally, each event would be convolved with the JER to fill a transfer matrix, but this is technically difficult to do. The same problem as for the estimation of the shape sensitivity bias arises, and the solution is to use the "matched truth" distribution in fine 1 GeV bins, which is provided by the fine bins transfer matrix.
- It is not necessary to draw events from the truth distribution that would be smeared by the Gaussian JER, where the number of generated events would limit the statistical precision of the non-Gaussianity uncertainty. It is actually possible to not introduce extra statistical uncertainties. The solution is simply to convolve each truth bin by a Gaussian distribution. It just amounts to having a Gaussian distribution centered on the center of the bin, of width given by the JER in that bin, and normalized by the number of entries in that bin.

Technically, the building of the "Gaussian transfer matrix" is done as follows:

1. Take the matched truth distribution in fine bins (1 GeV);
2. Smear each truth bin $\left[p_T^{\text{truth},1}, p_T^{\text{truth},2}\right]$ by a Gaussian distribution $\mathcal{N}(\mu, \sigma)$ where $\mu = \frac{1}{2}\left(p_T^{\text{truth},1} + p_T^{\text{truth},2}\right)$ and $\sigma = R \cdot \mu$ with R the JER at the center of this fine $(p_T, |y|)$ bin, with $R = \max\left(R_{\text{data}}, R_{\text{MC}}\right)$;
3. Integrate the Gaussian on the reco bin $\left[p_T^{\text{reco},1}, p_T^{\text{reco},2}\right]$ (expression is analytic, given hereafter) to obtain the contribution to the bin of the Gaussian transfer matrix $\left[p_T^{\text{truth},1}, p_T^{\text{truth},2}\right] \times \left[p_T^{\text{reco},1}, p_T^{\text{reco},2}\right]$ (note that in principle the reco bins could already be large, but I used 1 GeV reco bins for technical simplicity);
4. Merge the small bins to obtain the Gaussian transfer matrix in normal bins.

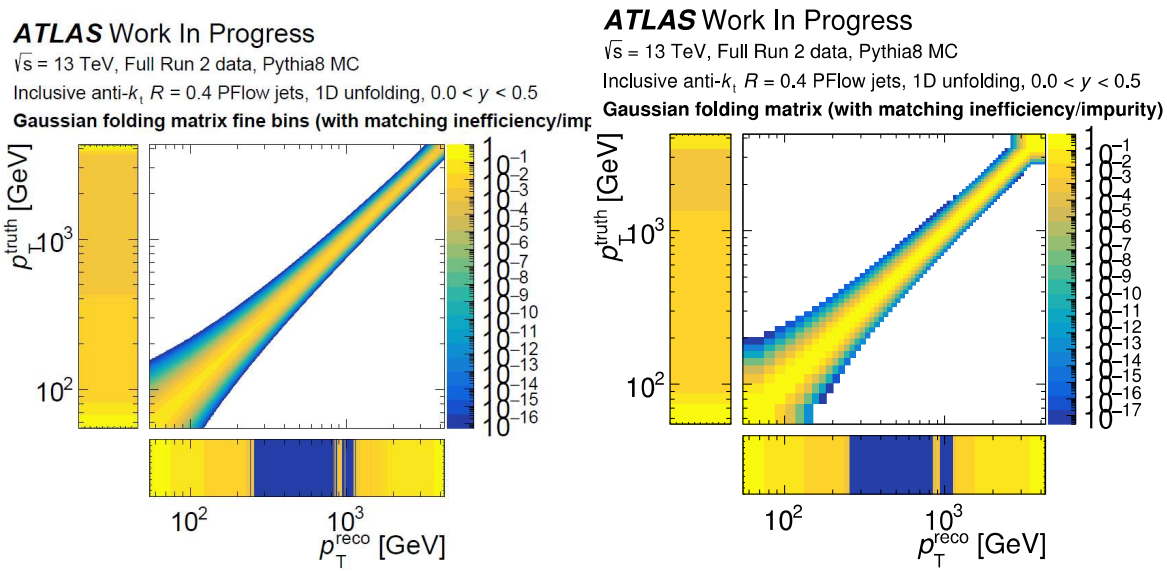
The computation is as follows. The content of the bin $\left[p_T^{\text{truth},1}, p_T^{\text{truth},2}\right] \times \left[p_T^{\text{reco},1}, p_T^{\text{reco},2}\right]$ of the Gaussian transfer matrix is expressed:

$$\mathcal{A} = N\left(p_T^{\text{truth},1}, p_T^{\text{truth},2}\right) \int_{p_T^{\text{reco},1}}^{p_T^{\text{reco},2}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{(u-\mu)^2}{\sigma^2}} \, \mathrm{d}u \tag{5.21}$$

with $N\left(p_T^{\text{truth},1}, p_T^{\text{truth},2}\right)$ the number of events at truth level in the bin $\left[p_T^{\text{truth},1}, p_T^{\text{truth},2}\right]$. Using the error function (defined as $\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} \, \mathrm{d}t$), we end up with:

$$\mathcal{A} = \frac{N\left(p_T^{\text{truth},1}, p_T^{\text{truth},2}\right)}{2} \left[\text{erf}\left(\frac{p_T^{\text{reco},2} - \mu}{\sqrt{2}\sigma}\right) - \text{erf}\left(\frac{p_T^{\text{reco},1} - \mu}{\sqrt{2}\sigma}\right)\right] \tag{5.22}$$

The Gaussian folding matrix is shown in Fig. 5.20, in fine bins and then in the normal "large" bins. The non-Gaussianity uncertainty is shown in the lower panels of Fig. 5.25. While never being dominant compared to other uncertainties, it is observed to be a non-negligible effect, which is added as another uncertainty on the measurement.



(a) (b)

Figure 5.20: Gaussian folding matrices for the estimation of the non-Gaussianity bias for the inclusive jet cross-section measurement in the $|y| < 0.5$ bin, with **(a)** fine bins or **(b)** normal "large" bins.

5.3.3 Other uncertainties

There are two categories of remaining uncertainties: the uncertainties due to the jet selection and the uncertainties due to the knowledge of the beam energy and the recorded luminosity.

The geometrical matching used to build the transfer matrix uses the $\Delta R < 0.3$ selection criterion to decide whether jets are matched or not. Variations of this selection criterion by ± 0.1 are used to evaluate uncertainties.

The jet timing selection of ± 10 ns introduces an inefficiency that has been shown in [103] to be negligible, of the permille level, that does not need to be corrected for but that can be accounted for as an uncertainty. The inefficiency is evaluated by fitting the $[-10\text{ ns}, +10\text{ ns}]$ core of the jet timing distributions with a Crystal Ball function and to compare the integral of this function in this interval to the integral of this function in $]-\infty, +\infty[$.

The trigger selection induces small inefficiencies, at the permille level, that are taken into account in an uncertainty. The trigger efficiency is obtained with methods described in Chapter 4, Section 4.2.1.2: in a given phase space region, events are selected by a "reference" trigger that is fully efficient in this phase space region, and the trigger decision of the "probe" trigger is emulated. The efficiency of the probe trigger is the ratio between the number of jets that pass the emulated reference trigger and the number of jets that pass the reference trigger.

The per-jet cleaning induces an inefficiency at the percent level and therefore requires a correction. The efficiency of the per-jet cleaning is evaluated *in situ* with a tag-and-probe method in dijet events, both in data and in MC, with the idea that if one passes the cleaning, the other one should pass it too: one of the two jets passes the cleaning and is the reference jet, and the second jet is the probe jet. The efficiency is the ratio between the number of probe jets that pass the cleaning to the number of probe jets.

The impact of dead calorimeter tile modules has never been evaluated yet, and it is still unclear whether an inefficiency correction has to be applied or if an uncertainty is enough, and how to evaluate it.

The total luminosity recorded during Run 2 by ATLAS has been measured to be 140.1 fb^{-1} with a relative uncertainty of 0.83 % [62], and directly translates to a 0.83 % fully-correlated uncertainty over the whole phase-space (as shown by Eq. (5.2)).

An uncertainty in the beam energy of 0.1 % [159] is considered when comparing data to the theory prediction at fixed beam energy. To that extent, the theory predictions are compared at the nominal and shifted values of the beam energy.

5.4 Technical optimizations

The technical implementation of these analyses is non trivial, due to its demand in computational resources: storage, CPU time, RAM, read/write operations. A few technical optimizations over the years make all this possible and are listed here. My most notable contribution to this long-standing effort is the optimization of the bootstrap library: I made the filling of histograms of replicas 3 times faster, an improvement that will benefit to all future users of this library. Together with simple code optimizations, I made the execution time of the histogram creating code more than 10 times faster.

The resource-intensity of this analysis arises from the very high number of systematic uncertainties that have to be propagated through the analysis, which causes a number of issues. More precisely, there are 2432 systematic uncertainties that need to be propagated through the unfolding, including 2128 coming from the statistical uncertainties of the η-intercalibration, and 68 from the JER. One first difficulty is to store the events with the varied energy and p_T for all these 2432 systematic variations, before histogram filling. A dedicated file format has been invented by Matt Leblanc to store the events and their 2432 variations: since only the energy and the p_T change for each jet, the jet's angular information does not need to be stored 2432 times.

The execution is parallelized as much as possible: considering the machines we have access to, it is reasonable to send approximately 5000 jobs that each take almost a week to run (thanks to the execution time improvements mentioned hereafter). The events are therefore split into 5000 input files, then the histograms are simply added after execution.

Another problem is the execution time of the histogram-filling code and the RAM required, which are adjusted by reducing the number of bootstrap replicas where possible. Let's focus on the leading jet analysis. There are 9 rapidity bins, with respectively the following numbers of p_T bins: 20, 20, 19, 17, 16, 12, 8, 5, 3; and there are three jet orders considered, hence a total of $120 \times 3 = 360$ bins. The size of a `double` is 8 bytes, and is used for the nominal histograms; the size of a `float` is 4 bytes and is used for the bootstrap replicas (this is a first optimization; although it could be optimized further by only using `floats`, since the numerical precision is not expected to play any role here).

The statistical uncertainty on the nominal spectrum is computed with 1000 bootstrap replicas, the statistical uncertainty on each systematic uncertainty uses 100 replicas, except the systematic uncertainties that come from the statistical uncertainties of the η-intercalibration, that use 0 replica, as already mentioned (which drastically reduces the CPU calls and hence the execution time). The following RAM requirements are estimated for the bins of histograms alone:

- Nominal MC reco and truth spectra: 3 MB;
- Nominal transfer matrix: 500 MB;
- 2128 JES systematic uncertainties from statistical uncertainties of the η-intercalibration, reco spectra: 12 MB;
- 304 other JES/JER systematic uncertainties, reco spectra: 45 MB;
- 68 JER systematic uncertainties, transfer matrices: 3.6 GB.

The total estimated RAM required for the histogram creation code is 4.2 GB, driven by the number of replicas of the JER systematic uncertainties due to the transfer matrices; when testing, I observed 4.7 GB. While still large, this is acceptable and does not motivate further reductions of the number of bootstrap replicas.

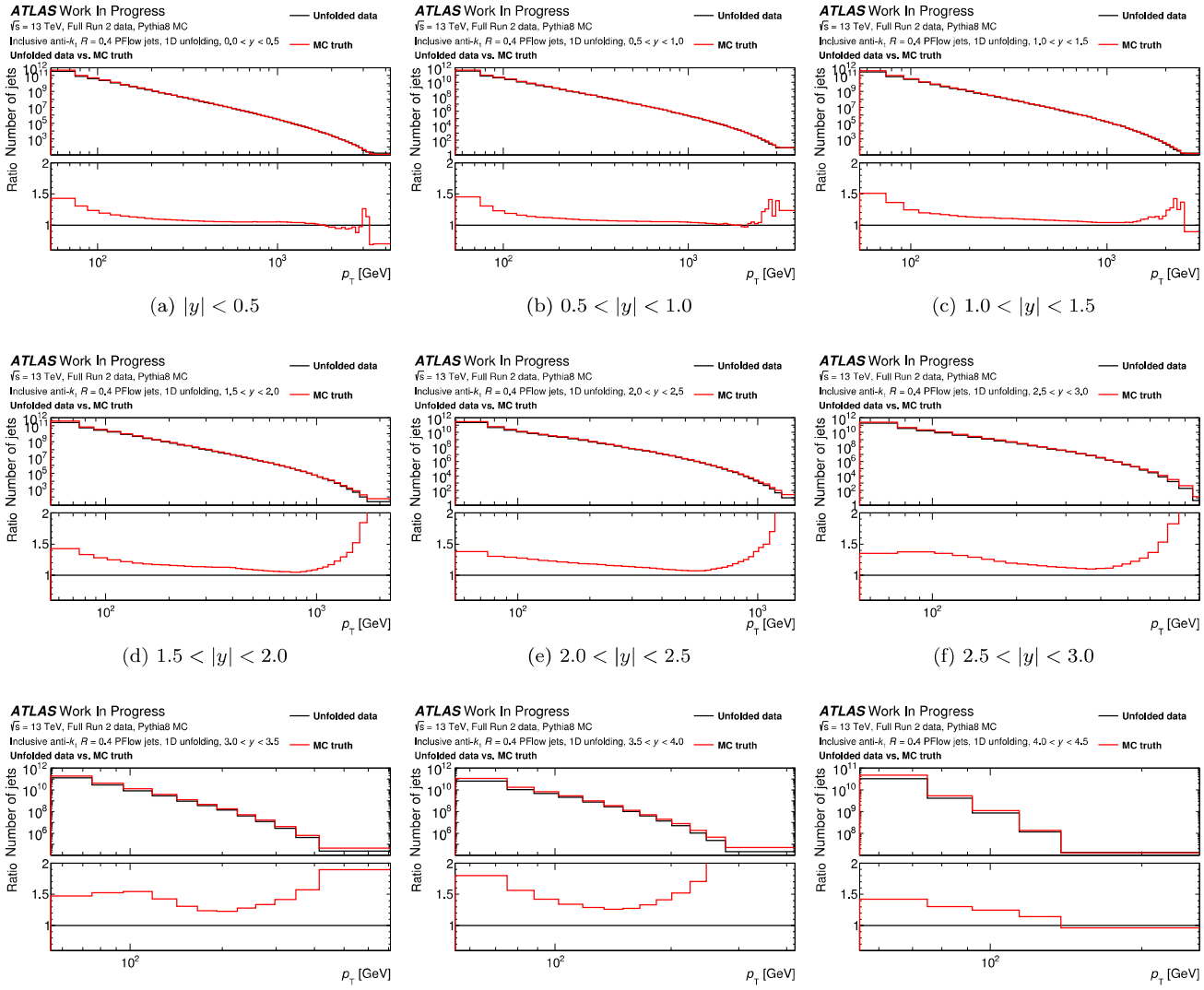
The size of the files has also shrunk a lot when moving from 1000 replicas to 100 replicas for the JER systematic uncertainties, which also helped the distributed computing resources with the automated file transfer operations at the end of the code execution.

Finally, I have drastically improved the code execution speed, by improving the main bottleneck: the histogram filling operation for the bootstrap replicas.

I have improved the `Fill()` operation of the bootstrap library by a factor of 3, an improvement that is useful to all users of the library. Internally, the bootstrap library stores the replicas of a one-dimensional histogram in a two-dimensional histogram, and similarly the replicas of a two-dimensional histogram are stored in a three-dimensional histogram. However, when filling an entry, each call to the `Fill()` method of `ROOT` searches the bin number to add the entry's weight to. Since the bin number is the same for all the replicas, it does not have to be determined every single time the replicas are filled, only once. Combined with the removal of useless loops, with the instantiation of some variables at initialization time, and with the exclusion of some frequent cases that take time for no reason, I managed to make it 3 times faster.

Also, another simple code optimization reduced the bin finding time of the histogram filling operations by a factor 3 to 5.

All the aforementioned improvements make these two analyses manageable from the computational power point of view, on the available machines, within a few weeks.



(a) $|y| < 0.5$ (b) $0.5 < |y| < 1.0$ (c) $1.0 < |y| < 1.5$

(d) $1.5 < |y| < 2.0$ (e) $2.0 < |y| < 2.5$ (f) $2.5 < |y| < 3.0$

(g) $3.0 < |y| < 3.5$ (h) $3.5 < |y| < 4.0$ (i) $4.0 < |y| < 4.5$

Figure 5.21: Inclusive jet cross-section measurement, unfolded using PYTHIA8, compared to the truth-level prediction of PYTHIA8 in the different rapidity bins of the analysis.

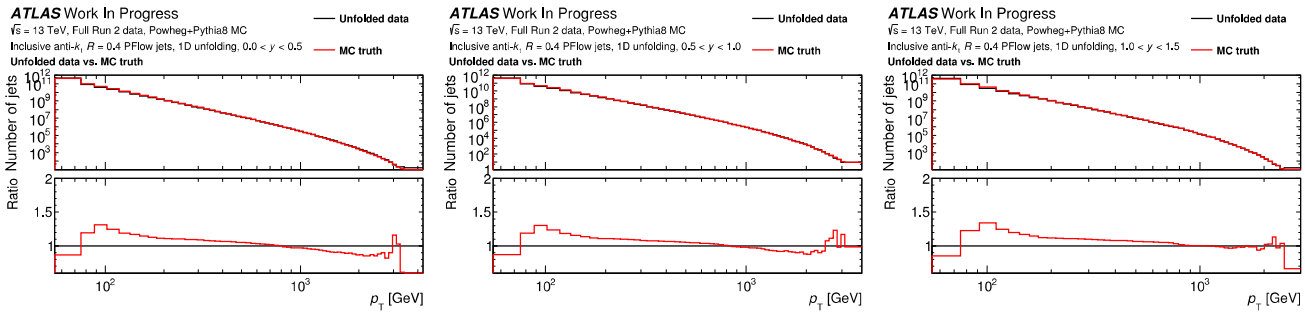
5.5 Results

5.5.1 Unfolded data and uncertainties

The unfolded measurement for the inclusive jet production cross-section measurement has been obtained with PYTHIA8 and with POWHEG+PYTHIA8, and compared to the truth-level predictions of these generators, as can be observed in Figs. 5.21 and 5.22. Since unfolded data depend very little on the choice of MC generator, the observed difference above 800 GeV is mainly the difference between the truth shapes in the MCs (which are expected since PYTHIA8 is a leading order generator while POWHEG+PYTHIA8 is a next-to-leading order one).

As a reminder, since those plots are not final, the p_T bins below 100 GeV have not been removed, and should not be considered as part of the final results.

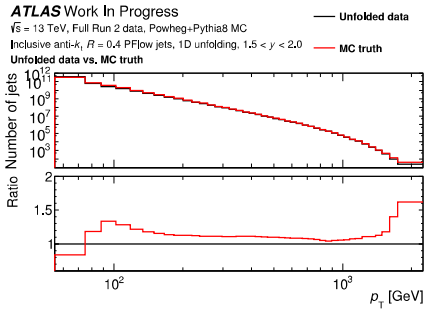
The statistical uncertainties on the measurement have been obtained with the bootstrap method, see Section 5.2.3.2, and an estimation for the bias related to the sensitivity of the IDS unfolding method to the data/MC shape differences has been obtained with a data-driven test, see Section 5.2.3.3. The sum in quadrature of them both is shown in Fig. 5.23. In the most central region, $|y| < 0.5$, the sum in quadrature of them is of the order of 1 %, due to the shape bias, and quickly falls to 0.1 % or even lower around 200 GeV, than starts rising around 700 GeV due to the data statistical uncertainties, reaches 1 % at 1.5 TeV, and goes as high as 30 % in the highest p_T region above 3 TeV. The trend is similar for the other rapidity bins, though the lowest uncertainty value reached is higher when $|y|$ increases, and the raise due to the data statistical uncertainty starts at lower p_T values. In



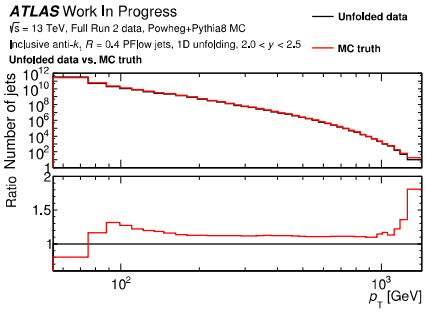
(a) $|y| < 0.5$

(b) $0.5 < |y| < 1.0$

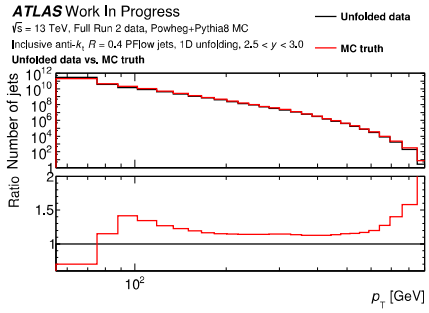
(c) $1.0 < |y| < 1.5$



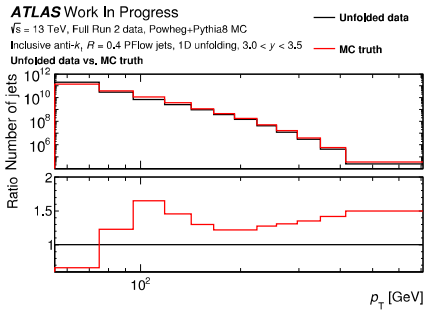
(d) $1.5 < |y| < 2.0$



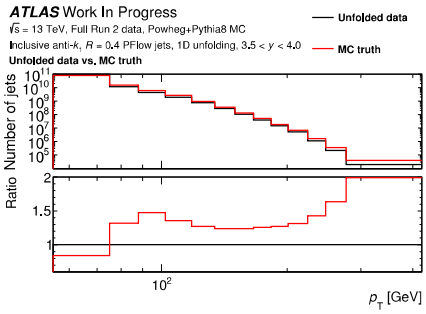
(e) $2.0 < |y| < 2.5$



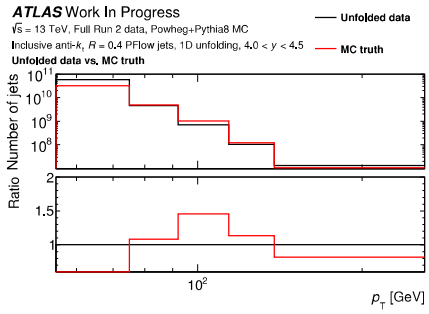
(f) $2.5 < |y| < 3.0$



(g) $3.0 < |y| < 3.5$

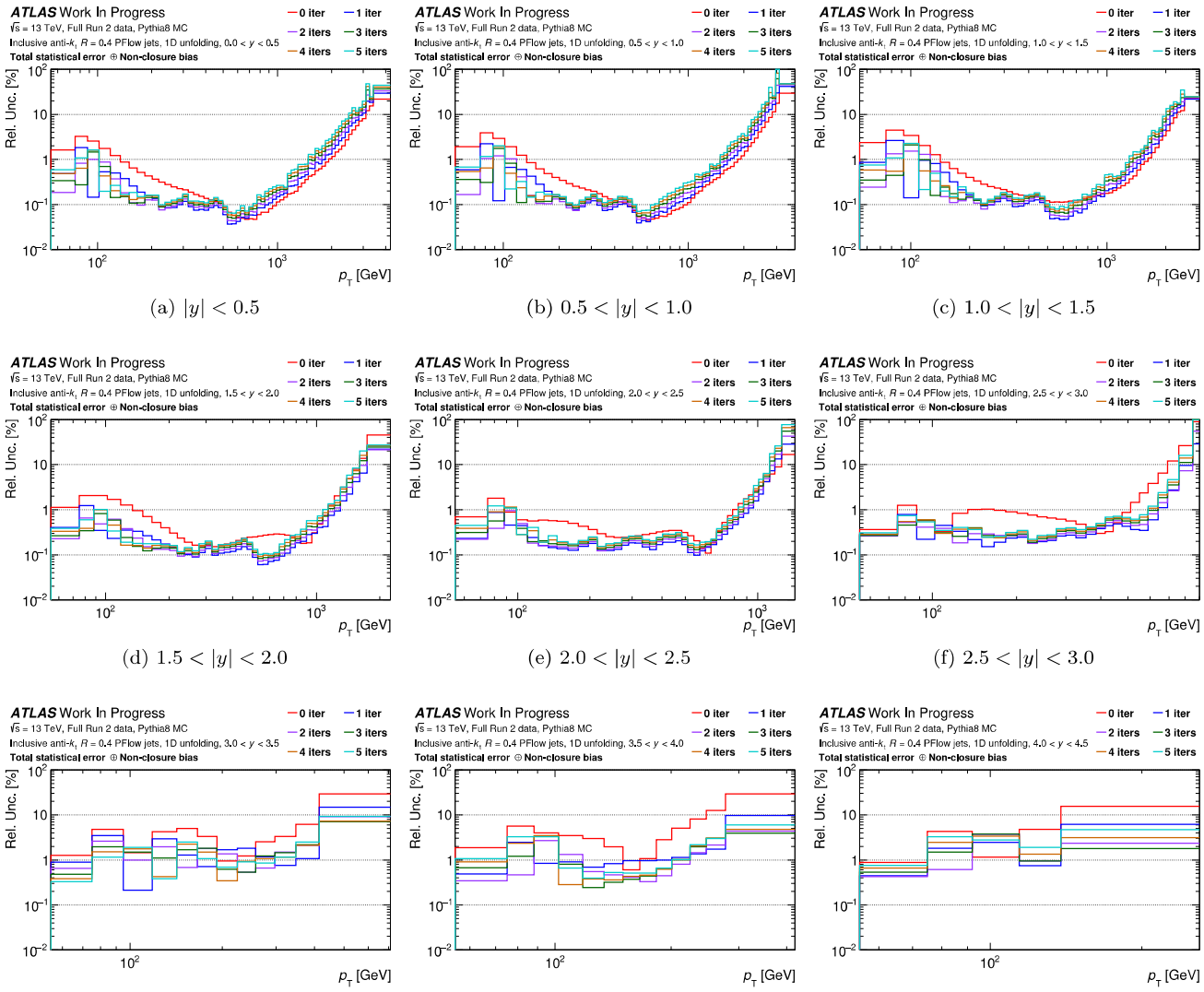


(h) $3.5 < |y| < 4.0$



(i) $4.0 < |y| < 4.5$

Figure 5.22: Inclusive jet cross-section measurement, unfolded using POWHEG+PYTHIA8, compared to the truth-level prediction of POWHEG+PYTHIA8 in the different rapidity bins of the analysis.

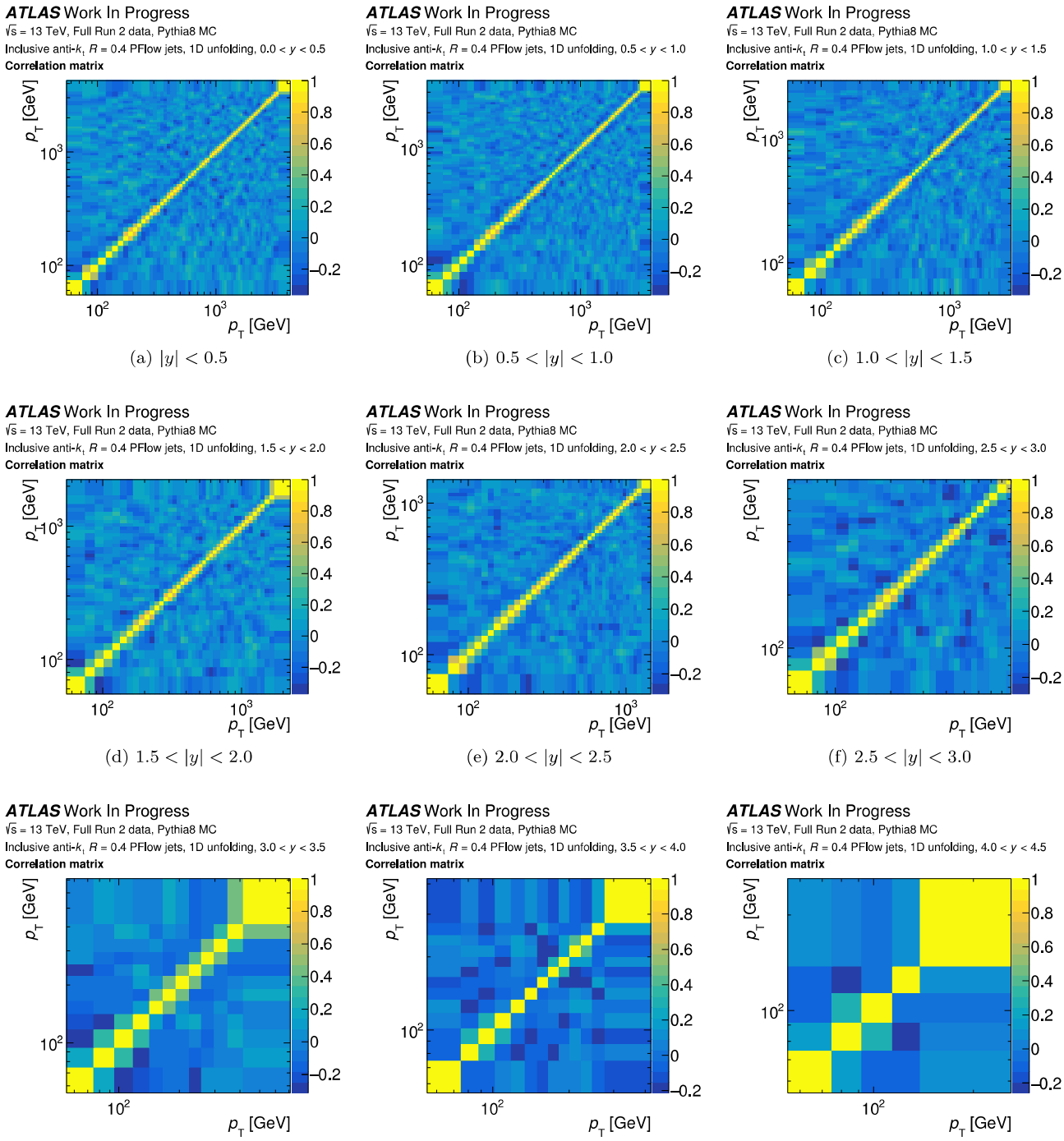


(a) $|y| < 0.5$ (b) $0.5 < |y| < 1.0$ (c) $1.0 < |y| < 1.5$

(d) $1.5 < |y| < 2.0$ (e) $2.0 < |y| < 2.5$ (f) $2.5 < |y| < 3.0$

(g) $3.0 < |y| < 3.5$ (h) $3.5 < |y| < 4.0$ (i) $4.0 < |y| < 4.5$

Figure 5.23: Sum in quadrature of the statistical uncertainty and the bias of sensitivity of the unfolding method to the data/MC shape differences, for the inclusive jet cross-section measurement, unfolded using PYTHIA8, in the different rapidity bins of the analysis. The number of iterations chosen for the analysis is 1.



(a) $|y| < 0.5$ (b) $0.5 < |y| < 1.0$ (c) $1.0 < |y| < 1.5$

(d) $1.5 < |y| < 2.0$ (e) $2.0 < |y| < 2.5$ (f) $2.5 < |y| < 3.0$

(g) $3.0 < |y| < 3.5$ (h) $3.5 < |y| < 4.0$ (i) $4.0 < |y| < 4.5$

Figure 5.24: Total statistical correlations, for the inclusive jet cross-section measurement, unfolded using PYTHIA8, in the different rapidity bins of the analysis. (Note that **(a)** is identical to Fig. 5.12d, to ease the side-by-side comparison of Fig. 5.12).

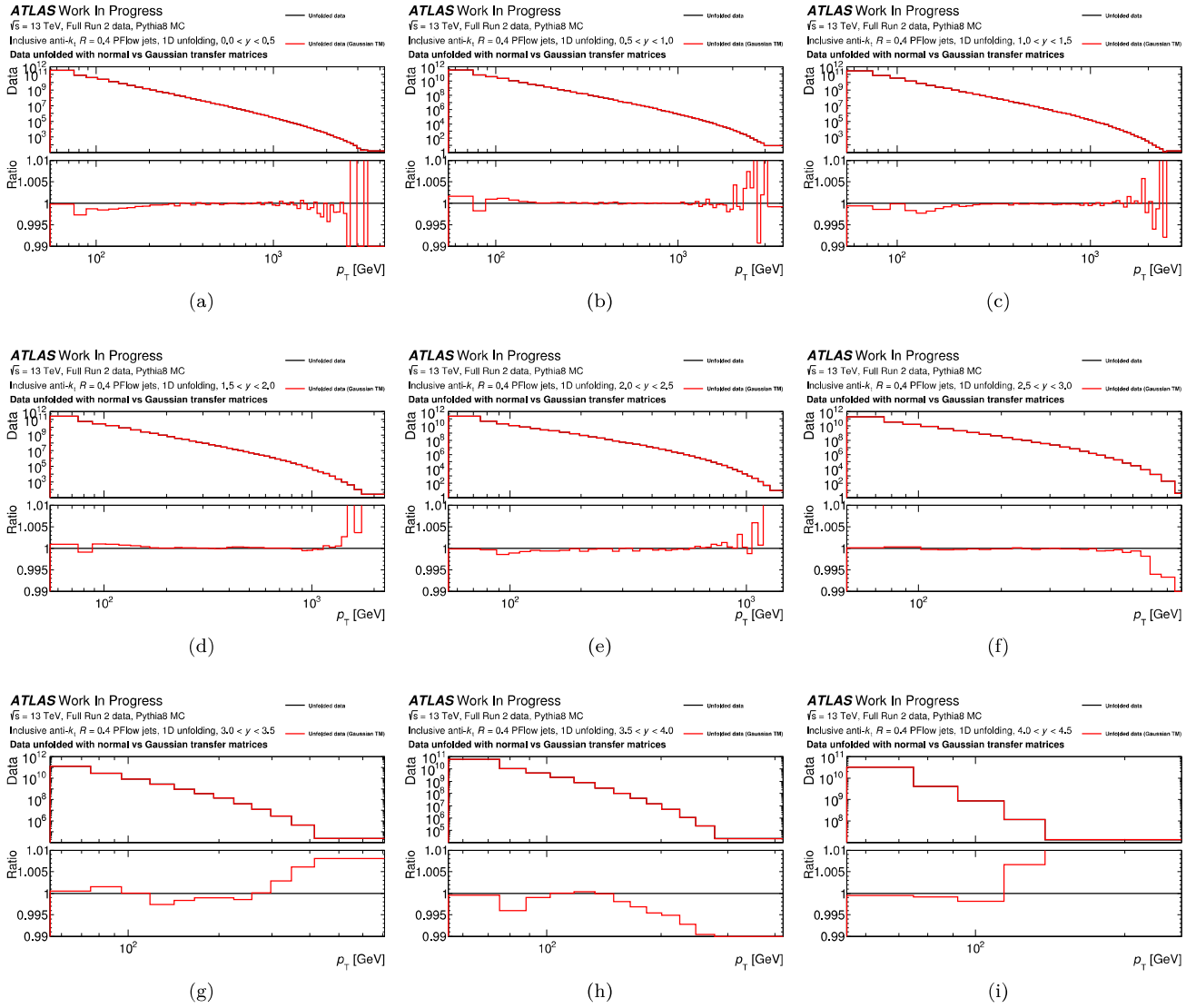


Figure 5.25: Non-Gaussianity uncertainty for the inclusive jet cross-section measurement, in the different rapidity bins of the analysis. The upper panels show the cross-section, unfolded either with the normal transfer matrix or with the Gaussian one, and the lower panel shows the ratio of the two, which is the non-Gaussianity uncertainty (after subtraction of 1).

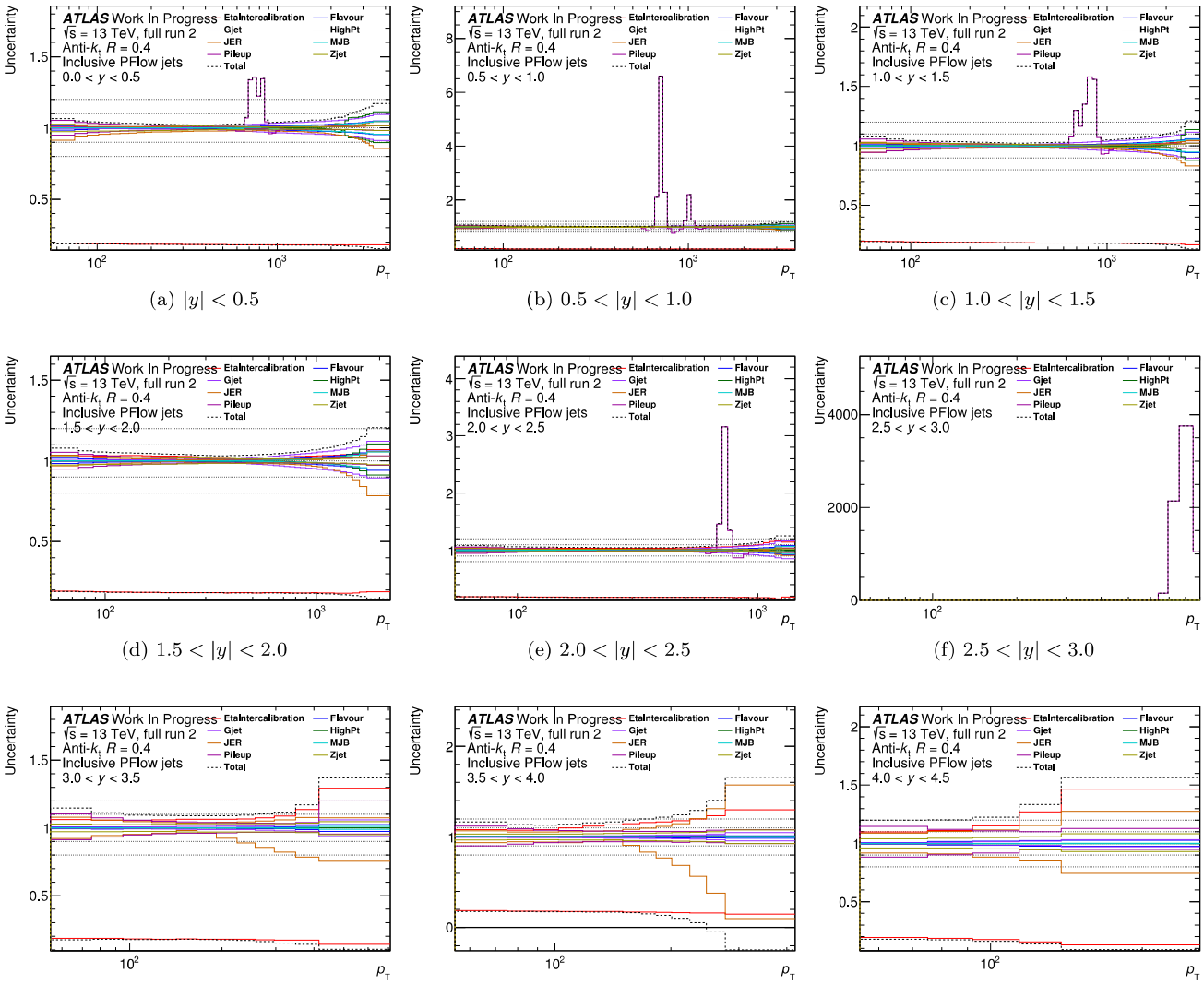
the forward rapidity regions, at $3.0 < |y| < 4.5$, the uncertainty oscillates around 1 %, and reaches 10 % in the respective highest-p_T regions.

The statistical correlations have been obtained with the bootstrap method as well, shown in Fig. 5.24, and have already been discussed in Section 5.2.3.2. The (anti-)correlations are mostly located near the diagonal, as the effect of unfolding is to migrate entries to the neighboring bins: correlations of 30 % to 70 % are observed in the bins adjacent to the diagonal, while anti-correlations up to 40 % are observed in the 1 or 2 bins adjacent to those ones.

The new resolution non-Gaussianity uncertainty has been evaluated by comparing the unfolding with the normal transfer matrix, constructed from events matched between truth and reco levels, and the unfolding with a "Gaussian" transfer matrix, built from the convolution of the truth distribution with a Gaussian JER. This uncertainty is shown in Fig. 5.25, and is below the permille level in the central rapidity bins from 200 GeV to 1 TeV. It reaches 0.2 % below 200 GeV, and it is dominated by statistical fluctuations above 1 TeV. For $3.0 < |y| < 4.5$, this uncertainty reaches the percent level, especially in the highest-p_T regions.

The systematic uncertainties of JES and JER have been propagated through the unfolding, yielding the relative uncertainties shown in Fig. 5.26:

- In the central rapidity bins $|y| < 1.5$, the total systematic uncertainty is within 1 %, except in the high-p_T region above 2.5 TeV where it goes up to almost 2 %. The total systematic uncertainty of JES and JER goes as



(a) $|y| < 0.5$ (b) $0.5 < |y| < 1.0$ (c) $1.0 < |y| < 1.5$

(d) $1.5 < |y| < 2.0$ (e) $2.0 < |y| < 2.5$ (f) $2.5 < |y| < 3.0$

(g) $3.0 < |y| < 3.5$ (h) $3.5 < |y| < 4.0$ (i) $4.0 < |y| < 4.5$

Figure 5.26: Systematic uncertainties of JES and JER propagated through the unfolding, summed in quadrature per category, for the inclusive jet cross-section measurement, unfolded using PYTHIA8, in the different rapidity bins of the analysis. Note that two uncertainties are incorrect and will be produced again: one pile-up uncertainty and one η-intercalibration uncertainty.

low as 0.3 % between 200 GeV and 800 GeV, dominated by the pile-up, Z+jet, γ+jet and JER uncertainties. In this p_T region, the statistical and shape bias uncertainties are also very low, making this region dominated by the 0.83 % luminosity uncertainty, which is an impressive feat for this long-standing inclusive jet cross-section analysis! Below 200 GeV, the dominant uncertainties are the pile-up and the JER uncertainties. Above 800 GeV, the dominant uncertainties are the γ+jet, single particle and JER uncertainties. The down part of the total JES and JER systematic uncertainty is almost always dominated by the JER, that provides the most asymmetrical uncertainties.

- In the intermediate rapidity bins $1.5 < |y| < 3.0$, the conclusions are similar, except that the η-intercalibration uncertainties become dominant in the 100 GeV to 500 GeV range, and the JER uncertainties are only dominant in the highest p_T bins.
- In the forward rapidity bins $3.0 < |y| < 4.5$, the total JES and JER uncertainty is of the order of 1 % to 2 %, increasing in the highest-p_T bins. The total JES+JER uncertainty is dominated by the η-intercalibration and JER uncertainties only.

5.5.2 Comparison to theory predictions

The theory predictions have only been done for the inclusive jet cross-section measurement, by Tancredi Carli, who also did the data/theory comparisons with my unfolding results. This work is not complete, as the theory predictions used here only use NLO matrix elements and not NNLO ones, without the electro-weak corrections applied, and some work remains to be done on the experimental uncertainties as mentioned in Section 5.3.3.

Theory predictions

The framework used to provide the theory predictions at fixed order has been described in Chapter 1, Section 1.3.1.7: NLOJET++ (or NNLOJET++) is used to provide NLO partonic predictions (or NNLO ones respectively), then non-perturbative and electro-weak corrections are applied.

Multiple choices are made at this stage: choice of the set of PDFs, of the regularization and factorization scales, of α_S, of the MC tune for the non-perturbative corrections, *etc.* Furthermore, uncertainties on these inputs have to be propagated through the computation. However, the theory predictions are very computationally heavy, and it would be very impractical to redo them for every possible choice. To make all this possible, the Applgrid project was developed [160]. It allows to convolve the PDFs or renormalization scale *a posteriori*. To do so, the perturbative coefficients (matrix elements) of the fixed-order QCD calculations are stored in 3-dimensional grids depending on the scale choice and two Bjorken-x variables of the scattering partons, for every single bin of the observable and for every possible partonic sub-process (which are gg, qg, gq, qq, $q\bar{q}$, qq' and $q\bar{q}'$ where g is a gluon, q a quark and q' a quark with a different flavor). The PDFs used are CT14 [161], CT18 [162] and NNPDF4.0NLO [163]. The convolution with the PDFs is done with the LHAPDF package [164], along with the HOPPET package [165] for the evolution of the PDFs to the required scale.

Multiple sources of uncertainties affect the theory predictions:

- The uncertainties of each PDF set are propagated through the calculations;
- Different choices of the renormalization and factorization scales can be made: the p_T of each jet p_T^{jet}, or the p_T of the leading jet p_T^{max}, where the former is chosen as the nominal one, and the comparison between both provides an uncertainty (note that other choices are possible, like H_T the scalar sum of the p_T of the jets, or $\hat{H}_T$ the scalar sum of the p_T of the outgoing partons);
- The nominal renormalization and factorization scales are varied up or down by a factor of 2 in both directions, excluding opposite variations of μ_R and μ_F, to estimate the uncertainty due to missing higher-order terms in the perturbative QCD expansion;
- The value of α_S is the one from world average done by the PDG [3]: $\alpha_S(M_Z^2) = 0.1179 \pm 0.0009$. The uncertainty on α_S is evaluated by doing the calculation with two PDF sets that only differ by the value of α_S used and then scaling the cross-section difference corresponding to an α_S uncertainty $\Delta\alpha_S = 0.0009$;
- The number of Monte Carlo parton-level events generated also introduces a statistical uncertainty.

Results

The comparison between the theory predictions and the unfolded experimental data has been made for the inclusive jet cross-section analysis, although as said this work is incomplete: the theory predictions used here only use NLO matrix elements and not NNLO ones, without the electro-weak corrections applied, and some work remains to be done on the experimental uncertainties as mentioned in Section 5.3.3. I show the result for the rapidity bin $|y| < 0.5$ in Fig. 5.27, for the three PDF sets.

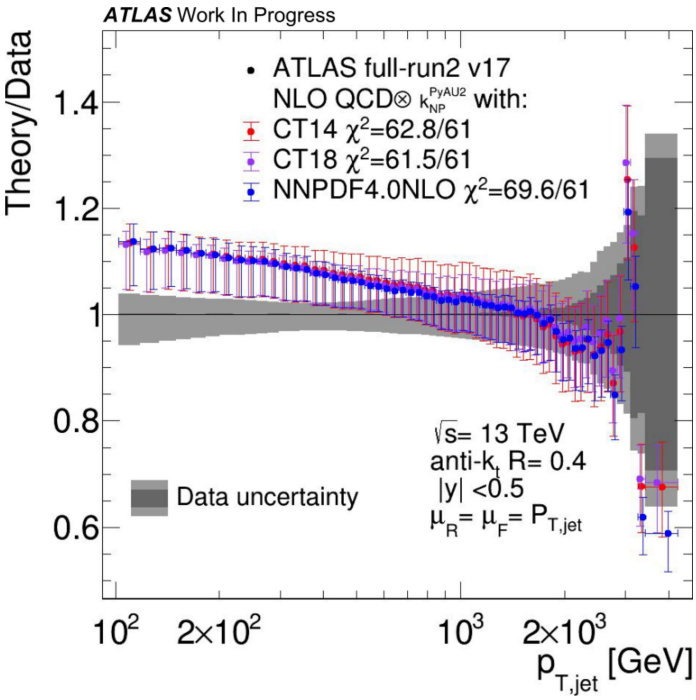


Figure 5.27: Comparison between unfolded data and theory predictions for the inclusive jet cross-section in the $|y| < 0.5$ rapidity bin. The statistical uncertainty of data is shown in the dark grey band while the total uncertainty including the systematic uncertainties is shown in the light grey band. The non-smoothed η-intercalibration statistical uncertainties are used. Multiple predictions made with three separate PDF sets are shown. The agreement between data and theory is quantified with the χ^2/N_{dof}, provided for each PDF set. Plot made by Tancredi Carli.

The bin-by-bin agreement between data and theory is very good, and the global agreement in one single rapidity bin is also rather good: it is quantified as $\chi/N_{\text{dof}} = 69.6/61$, using NNPDF4.0NLO as the default PDF set (note that this χ^2 computation takes the statistical correlations into account). But when taking all the rapidity bins into account, the global χ^2/N_{dof} is 496/297. This tension is far too strong, as was already observed in the previous 8 TeV and 13 TeV measurements [31, 136], pointing at plausible issues, whether in the experimental measurement or in the theoretical prediction. The χ^2 is sensitive to global trends, local fluctuations and correlations:

- The local fluctuations are being addressed with the statistical uncertainties: the ones from data, from the *in situ* analyses, and the ones from theory. Using individual η-intercalibration statistical uncertainties that are local in $(p_T, |y|)$, instead of one single fully-correlated uncertainty, helped reducing the χ^2 by 200 units in the 8 TeV analysis of [136], that had 159 bins. Using non-smoothed η-intercalibration statistical uncertainties was also shown to help in the past month: it reduced the χ^2 by 150 units in the current inclusive jet cross-section analysis, that has 297 bins.
- Other possibilities are currently being investigated, mostly on the correlation side, as explained in [31, 136, 166]. Alternative correlation schemes of the JES and JER uncertainties have been tested, reducing the χ^2 by up to 12 units in some phase-space regions. The two-point experimental and theoretical uncertainties have been decorrelated, reducing the χ^2 by up to 96 units with the most optimistic approach. All of these improvements were still not enough to make the $\frac{\chi^2}{N_{\text{dof}}}$ acceptable: it dropped down to 256/159, the associated p value remaining smaller than 10^{-3}.

Conclusion and prospects

I worked on two jet production cross-section analyses, the inclusive jet production cross-sections jointly with Tancredi Carli, and the leading jet production cross-section based on the work of Robert Hankache in [103]. In both cases, progress has been made, although work remains to be done.

For the inclusive jet cross-section analysis, I did the unfolding of the detector effects with the IDS method, evaluated the statistical uncertainties and correlations as well as the bias related to the sensitivity of the method to shape differences between data and MC. I measured the impact of the non-Gaussianity of the JER for the first time in ATLAS cross-section measurements by comparing the spectrum unfolded either with the nominal transfer matrix, containing non-Gaussianities, and with a Gaussian transfer matrix. I propagated the numerous systematic uncertainties of the JES and the JER through the analysis.

The comparison of the experimental result after unfolding with the theory predictions showed a strong disagreement, already observed in some recent previous inclusive jet cross-section measurements in ATLAS and CMS. Compared to previous work on the subject, I have worked on a crucial strategy that have reduced this tension by more than 150 χ^2 units: the use of non-smoothed η-intercalibration statistical uncertainties.

Some work remains to be done on this analysis, both on the experimental and theory sides, though it is approaching the end. The selection uncertainties have still not been evaluated, and the efficiency corrections are currently not applied. The NNLO prediction is currently being computed, along with the electro-weak corrections.

For the leading jet cross-section analysis, I started at a lower-level as I produced the nominal histograms myself: spectra, transfer matrix, migration efficiencies and purities. I extended the measurement to the forward rapidity region $3.0 < |y| < 4.5$ to take potential migrations with this region into account.

On the technical side, I improved the code of the bootstrap library to make it three times faster. Along with other simple but crucial improvements, this will make the histogram production approximately 10 times faster, making the propagation of the thousands of systematic uncertainties feasible.

Some work is still ongoing for this analysis, for which most key ingredients are already in place: the unfolding, and the theory predictions.

Conclusion and Perspectives

This thesis was about testing the Standard Model of particle physics at the energy frontier, using percent-level precision measurements. Experimental data was recorded by the ATLAS experiment from proton-proton collisions produced by the LHC during its Run 2 operation period at a center-of-mass energy of 13 TeV, in high pile-up conditions. The focus of this thesis is jets: hadronic sprays of particles stemming from the hadronization of an outgoing quark or gluon; more specifically on anti-k_t jets with a small radius parameter of $R = 0.4$. Hopefully, this manuscript reflects the level of attention that is put in understanding the finest effects, in view of obtaining the most precise results possible. It is the interplay between all the elements of jet reconstruction and calibration that allows for the greatest precision, and the highest confidence in the results.

After the presentation of the theoretical and experimental contexts, the η-intercalibration for the *in situ* JES is presented in detail. The calibration step aims at making the detector response homogeneous across the whole latitudinal range of the detector, expressed by the pseudo-rapidity η. I obtained the precision calibration for the whole Run 2 period for the first time, and performed all the needed parameter-tweaking and cross-checks to ensure its quality. This led me to a deeper understanding of the method, justifying its consistency mathematically for the first time by studying the relationship between the mean asymmetry and the detector response.

I worked on improving the η-intercalibration with the "absolute implementation" of the method, that deconvolves physics effects from detector effects. It consists in using the truth-level Monte Carlo (MC) intercalibration coefficient at the denominator of the calibration, as this one only contains physics effects, so that the cancellation happens with the physics effects of the reconstruction level intercalibration coefficients, to only leave detector effects in the calibration. This separation of physics and detector effects allowed me to study them independently. This led to the reduction by a factor 2 of the dominant uncertainty in the transverse momentum range $p_T < 40$ GeV. It also allowed me to check that the selection uncertainties do cover the physics effects that they are supposed to, and are therefore not underestimated.

As an outlook, I think that there is more to learn from the absolute implementation, especially tests for the robustness of the matrix method. The next step for the absolute implementation would be a test in a physics analysis to ensure that the uncertainties improve when using this method, with an adequate treatment of the physics correlations.

Furthermore, I contributed to two precision measurements: the inclusive and leading jet analyses cross-sections, where the former counts all jets and the latter counts only one jet per event, the one with the highest p_T. Both analyses are double differential in p_T and $|y|$.

For the inclusive jet analysis, I performed the unfolding with an Iterative Dynamically Stabilized (IDS) method. The statistical uncertainties were evaluated with the Bootstrap method, along with statistical correlations. The bias related to the sensitivity of the unfolding method to shape differences between data and MC simulation have been evaluated, with special care as it was evaluated in the $3.0 < |y| < 4.5$ region for the first time. To that extent, the truth-level MC distribution is reweighted such that its associated reco distribution has a similar shape as data; then the reweighted reco distribution is unfolded and compared to the reweighted truth distribution, to evaluate the bias. This required the use of folding matrices with fine bins, a tweaking of the reweighting functions, and led me to develop a more quantitative evaluation of the criterion that is used to deem the reweighting functions satisfactory. A full propagation of the systematic uncertainties from the Jet Energy Scale (JES) and the Jet Energy Resolution (JER) was performed, posing significant computing performance issues, leading me to improve the Bootstrap library by making it three times faster. For the first time in ATLAS cross-section measurements, I evaluated the impact of the non-Gaussianity of the JER by comparing the spectrum unfolded either with the nominal transfer matrix, containing non-Gaussianities, and with a Gaussian transfer matrix. This JER non-Gaussianity is below the percent level, and even below the permille level in the central rapidity bins $|y| < 3.0$ from 200 GeV to 1 TeV. The final unfolded spectra have been compared to theory predictions, showing similar disagreement as have been observed by previous analyses. A significant improvement was made by using the non-smoothed η-intercalibration, reducing the χ^2 by more than 150 units for this analysis that has 297 bins. More hypotheses are being tested both on the

experimental and theoretical sides to reduce this disagreement, for instance by decorrelating two-point systematic uncertainties.

For the leading jet analysis, I produced all the nominal histograms, including the spectra, efficiencies and purities, and the transfer matrix that allows to account for three types of migrations: in p_T, $|y|$ and jet order. I extended the measurement to the forward rapidity region $3.0 < |y| < 4.5$. Many ingredients are already in place for the unfolding, to enable a data to theory comparison as for the inclusive jet analysis.

Both analyses will hopefully reach completion in the following months, as most issues are in good shape to be addressed.

Bibliography

[1] François Vannucci. *Les techniques de détection en physique de l'infiniment petit*. Ellipses, 2018.

[2] Matthew D Schwartz. *Quantum field theory and the standard model*. Cambridge university press, 2014.

[3] Particle Data Group. "Review of Particle Physics". In: *Progress of Theoretical and Experimental Physics* 2022.8 (Aug. 8, 2022). ISSN: 2050-3911. DOI: 10.1093/ptep/ptac097.

[4] CERN. "Le premier webfest du CERN se lance à la conquête des particules". Aug. 27, 2012. URL: https://cds.cern.ch/journal/CERNBulletin/2012/35/News%20Articles/1473657 (visited on 07/21/2023).

[5] ATLAS Collaboration. "Standard Model Summary Plots February 2022". Pub. Note. Mar. 11, 2022. CERN: ATL-PHYS-PUB-2022-009. URL: http://cds.cern.ch/record/2804061 (visited on 06/14/2023).

[6] Bryan Webber. "Parton shower Monte Carlo event generators". Dec. 1, 2011. URL: http://www.scholarpedia.org/article/Parton_shower_Monte_Carlo_event_generators (visited on 06/13/2023).

[7] Christian Bierlich et al. "A Comprehensive Guide to the Physics and Usage of PYTHIA 8.3". Mar. 22, 2022. DOI: 10.48550/arXiv.2203.11601. arXiv: 2203.11601 [hep-ex, hep-ph].

[8] Michael Peskin and Daniel Schroeder. *An introduction to quantum field theory*. CRC press, 2018.

[9] Torbjörn Sjöstrand et al. "An Introduction to PYTHIA 8.2". In: *Computer Physics Communications* 191 (June 2015). ISSN: 00104655. DOI: 10.1016/j.cpc.2015.01.024. arXiv: 1410.3012 [hep-ph].

[10] ATLAS Collaboration. "ATLAS Pythia 8 tunes to 7 TeV datas". Pub. Note. Nov. 2, 2014. CERN: ATL-PHYS-PUB-2014-021. URL: https://cds.cern.ch/record/1966419 (visited on 06/09/2020).

[11] Richard D. Ball et al. "Parton Distributions with LHC Data". In: *Nuclear Physics B* 867.2 (Feb. 11, 2013). ISSN: 05503213. DOI: 10.1016/j.nuclphysb.2012.10.003. arXiv: 1207.1303 [hep-ph].

[12] M. Bahr et al. "Herwig++ Physics and Manual". In: *The European Physical Journal C* 58.4 (Nov. 20, 2008). ISSN: 1434-6044, 1434-6052. DOI: 10.1140/epjc/s10052-008-0798-9. arXiv: 0803.0883 [hep-ph].

[13] Johannes Bellm et al. "Herwig 7.0 / Herwig++ 3.0 Release Note". In: *The European Physical Journal C* 76.4 (Apr. 11, 2016). ISSN: 1434-6044, 1434-6052. DOI: 10.1140/epjc/s10052-016-4018-8. arXiv: 1512.01178 [hep-ph].

[14] Johannes Bellm et al. "Herwig 7.1 Release Note". May 19, 2017. arXiv: 1705.06919 [hep-ph].

[15] Stefan Gieseke, Philip Stephens, and Bryan Webber. "New Formalism for QCD Parton Showers". In: *Journal of High Energy Physics* 2003.12 (Jan. 12, 2004). ISSN: 1029-8479. DOI: 10.1088/1126-6708/2003/12/045. arXiv: hep-ph/0310083 [hep-ph].

[16] Simon Platzer and Stefan Gieseke. "Coherent Parton Showers with Local Recoils". In: *Journal of High Energy Physics* 2011.1 (Jan. 10, 2011). ISSN: 1029-8479. DOI: 10.1007/JHEP01(2011)024. arXiv: 0909.5593 [hep-ph].

[17] Simon Platzer and Stefan Gieseke. "Dipole Showers and Automated NLO Matching in Herwig++". In: *The European Physical Journal C* 72.11 (Nov. 14, 2012). ISSN: 1434-6044, 1434-6052. DOI: 10.1140/epjc/s10052-012-2187-7. arXiv: 1109.6256 [hep-ph].

[18] L. A. Harland-Lang et al. "Parton Distributions in the LHC Era: MMHT 2014 PDFs". In: *The European Physical Journal C* 75.5 (May 9, 2015). ISSN: 1434-6044, 1434-6052. DOI: 10.1140/epjc/s10052-015-3397-6. arXiv: 1412.3989 [hep-ex, hep-ph].

[19] Enrico Bothmann et al. "Event Generation with Sherpa 2.2". In: *SciPost Physics* 7.3 (Sept. 18, 2019). ISSN: 2542-4653. DOI: 10.21468/SciPostPhys.7.3.034. arXiv: 1905.09127 [hep-ex, hep-ph].

[20] Steffen Schumann and Frank Krauss. "A Parton Shower Algorithm Based on Catani-Seymour Dipole Factorisation". In: *Journal of High Energy Physics* 2008.03 (Mar. 13, 2008). ISSN: 1029-8479. DOI: 10.1088/1126-6708/2008/03/038. arXiv: 0709.1027 [hep-ph].

[21] Jan-Christopher Winter, Frank Krauss, and Gerhard Soff. "A Modified Cluster-Hadronization Model". In: *The European Physical Journal C* 36.3 (Aug. 2004). ISSN: 1434-6044, 1434-6052. DOI: 10.1140/epjc/s2004-01960-8. arXiv: hep-ph/0311085 [hep-ph].

[22] Sayipjamal Dulat et al. "New Parton Distribution Functions from a Global Analysis of Quantum Chromodynamics". In: *Physical Review D* 93.3 (Feb. 16, 2016). ISSN: 2470-0010, 2470-0029. DOI: 10.1103/PhysRevD.93.033006. arXiv: 1506.07443 [hep-ex, hep-ph, nucl-ex].

[23] Paolo Nason. "A New Method for Combining NLO QCD with Shower Monte Carlo Algorithms". In: *Journal of High Energy Physics* 2004.11 (Dec. 10, 2004). ISSN: 1029-8479. DOI: 10.1088/1126-6708/2004/11/040. arXiv: hep-ph/0409146 [hep-ph].

[24] Stefano Frixione, Paolo Nason, and Carlo Oleari. "Matching NLO QCD Computations with Parton Shower Simulations: The POWHEG Method". In: *Journal of High Energy Physics* 2007.11 (Nov. 23, 2007). ISSN: 1029-8479. DOI: 10.1088/1126-6708/2007/11/070. arXiv: 0709.2092 [hep-ph].

[25] Simone Alioli et al. "A General Framework for Implementing NLO Calculations in Shower Monte Carlo Programs: The POWHEG BOX". In: *Journal of High Energy Physics* 2010.6 (June 9, 2010). ISSN: 1029-8479. DOI: 10.1007/JHEP06(2010)043. arXiv: 1002.2581 [hep-ph].

[26] The NNPDF Collaboration et al. "Parton Distributions for the LHC Run II". In: *Journal of High Energy Physics* 2015.4 (Apr. 8, 2015). ISSN: 1029-8479. DOI: 10.1007/JHEP04(2015)040. arXiv: 1410.8849 [hep-ex, hep-ph].

[27] Zoltan Nagy. "Three-Jet Cross Sections in Hadron-Hadron Collisions at next-to-Leading Order". In: *Physical Review Letters* 88.12 (Mar. 12, 2002). ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.88.122003. arXiv: hep-ph/0110315 [hep-ph].

[28] Zoltan Nagy. "Next-to-Leading Order Calculation of Three-Jet Observables in Hadron-Hadron Collision". In: *Physical Review D* 68.9 (Nov. 5, 2003). ISSN: 0556-2821, 1089-4918. DOI: 10.1103/PhysRevD.68.094002. arXiv: hep-ph/0307268 [hep-ph].

[29] A. Gehrmann-De Ridder et al. "Precise QCD Predictions for the Production of a Z Boson in Association with a Hadronic Jet". In: *Physical Review Letters* 117.2 (July 6, 2016). ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.117.022001. arXiv: 1507.02850 [hep-ex, hep-ph].

[30] James Currie et al. "Single Jet Inclusive Production for the Individual Jet p_T Scale Choice at the LHC". In: *Acta Physica Polonica B* 48.6 (Apr. 4, 2017). ISSN: 0587-4254, 1509-5770. DOI: 10.5506/APhysPolB.48.955. arXiv: 1704.00923 [hep-ph].

[31] ATLAS Collaboration. "Measurement of Inclusive Jet and Dijet Cross-Sections in Proton-Proton Collisions at $\sqrt{s} = 13$ TeV with the ATLAS Detector". In: *Journal of High Energy Physics* 2018.5 (May 30, 2018). ISSN: 1029-8479. DOI: 10.1007/JHEP05(2018)195. arXiv: 1711.02692 [hep-ex].

[32] Rikkert Frederix et al. "The Complete NLO Corrections to Dijet Hadroproduction". In: *Journal of High Energy Physics* 2017.4 (Apr. 12, 2017). ISSN: 1029-8479. DOI: 10.1007/JHEP04(2017)076. arXiv: 1612.06548 [hep-ph].

[33] Gavin Salam. "Basics of QCD: jets & jet substructure". Lecture at ICTPSAIFR school on QCD and LHC physics. CERN, July 2015. URL: https://gsalam.web.cern.ch/gsalam/repository/talks/2015-SaoPaulo-lecture4.pdf (visited on 06/09/2020).

[34] Gavin P Salam. "Towards jetography". In: *The European Physical Journal C* 67.3-4 (May 8, 2020), pp. 637–686. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-010-1314-6. arXiv: 0906.1833 [hep-ph].

[35] John E. Huth et al. "Toward a standardization of jet definitions". In: *Proceedings of Summer Study on High Energy Physics. Research Directions for the Decade.* 1990 Summer Study on High Energy Physics. FNAL-C-90-249-E. FERMILAB. Snowmass, Colorado, 1990, pp. 134–136.

[36] ATLAS Collaboration. "The highest jet multiplicity event collected by the end of October 2010". Oct. 5, 2010. URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/EventDisplayStandAlone (visited on 06/09/2020).

[37] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "FastJet User Manual". In: *The European Physical Journal C* 72.3 (Mar. 16, 2012). ISSN: 1434-6044, 1434-6052. DOI: 10.1140/epjc/s10052-012-1896-2. arXiv: 1111.6097 [hep-ex, hep-ph].

[38] Matteo Cacciari et al. "Quantifying the Performance of Jet Definitions for Kinematic Reconstruction at the LHC". In: *Journal of High Energy Physics* 2008.12 (Dec. 9, 2008). ISSN: 1029-8479. DOI: 10.1088/1126-6708/2008/12/032. arXiv: 0810.1304 [hep-ph].

[39] Johannes Haller et al. "Status of the Global Electroweak Fit with Gfitter in the Light of New Precision Measurements". Nov. 14, 2022. arXiv: 2211.07665 [hep-ph].

[40] Brian Greene. *The elegant universe: Superstrings, hidden dimensions, and the quest for the ultimate theory.* W. W. Norton & Company, 2000.

[41] Carlo Rovelli. *Quantum gravity.* Cambridge university press, 2004.

[42] Jean-Philippe Uzan and Nathalie Deruelle. *Théories de la Relativité.* Belin, 2014.

[43] Jean-Philippe Uzan and Patrick Peter. *Cosmologie primordiale.* Belin, 2015.

[44] Steven Weinberg. *The quantum theory of fields: Volume 2, modern applications.* Cambridge university press, 2005.

[45] Mikio Nakahara. *Geometry, topology and physics.* CRC press, 2018.

[46] Arthur Jaffe and Edward Witten. "Quantum Yang-Mills Theory". June 2022. URL: https://www.arthurjaffe.com/Assets/pdf/QuantumYangMillsWebRevised.pdf (visited on 07/19/2023).

[47] Jean Bricmont. *Making sense of quantum mechanics*. Springer, 2016.

[48] Lev Davidovich Landau and Evgeni Mikhalovich Lifshitz. *Mechanics*. Vol. 1. MIR, 1960.

[49] Andreas Crivellin. "Anomalies in Particle Physics". Apr. 4, 2023. DOI: 10.48550/arXiv.2304.01694. arXiv: 2304.01694 [hep-ex].

[50] Louis Lyons. "Discovering the Significance of 5 σ". Oct. 4, 2013. arXiv: 1310.1284 [hep-ex, hep-ph].

[51] B. Abi et al. "Measurement of the Positive Muon Anomalous Magnetic Moment to 0.46 Ppm". In: *Physical Review Letters* 126.14 (Apr. 7, 2021). ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.126.141801. arXiv: 2104.03281 [hep-ex, nucl-ex].

[52] P. Gambino et al. "Challenges in Semileptonic B Decays". In: *The European Physical Journal C* 80.10 (Oct. 19, 2020). ISSN: 1434-6044, 1434-6052. DOI: 10.1140/epjc/s10052-020-08490-x. arXiv: 2006.07287 [hep-ex, hep-lat, hep-ph].

[53] CMS Collaboration. "Search for a Standard Model-like Higgs Boson in the Mass Range between 70 and 110 GeV in the Diphoton Final State in Proton-Proton Collisions at $\sqrt{s} = 13$ TeV". Physics Analysis Summary. Mar. 20, 2023. CERN: CMS-PAS-HIG-20-002. URL: https://cds.cern.ch/record/2852907 (visited on 07/20/2023).

[54] CDF Collaboration. "High-Precision Measurement of the W Boson Mass with the CDF II Detector". In: *Science* 376.6589 (Apr. 8, 2022), pp. 170–176. DOI: 10.1126/science.abk1781.

[55] ATLAS Collaboration. "Improved W Boson Mass Measurement Using $\sqrt{s} = 7$ TeV Proton-Proton Collisions with the ATLAS Detector". Conf. Note. Apr. 2, 2023. CERN: ATLAS-CONF-2023-004. URL: https://cds.cern.ch/record/2853290 (visited on 07/20/2023).

[56] Barton Zwiebach. *A first course in string theory*. Cambridge university press, 2009.

[57] CERN. "CERN website". 2023. URL: https://home.cern/ (visited on 07/24/2023).

[58] CERN. "The CERN Experimental Program: Experiments and Projects". 2023. URL: https://greybook.cern.ch/experiment/list (visited on 07/24/2023).

[59] Nicolas Delerue. "Understanding the basic principles of particle accelerators". PHENIICS Doctoral School, 2021. URL: http://lal.delerue.org/teaching/201906_PHENIICS/ (visited on 06/25/2021).

[60] Lyndon Evans and Philip Bryant. "LHC Machine". In: *Journal of Instrumentation* 3.08 (Aug. 2008). ISSN: 1748-0221. DOI: 10.1088/1748-0221/3/08/S08001.

[61] CERN. "LHC sets world record beam intensity". Apr. 22, 2011. URL: https://press.cern/news/press-release/cern/lhc-sets-world-record-beam-intensity (visited on 07/24/2023).

[62] ATLAS Collaboration. "Luminosity Determination in *pp* Collisions at $\sqrt{s} = 13$ TeV Using the ATLAS Detector at the LHC". In: *The European Physical Journal C* 83.10 (Oct. 31, 2023). ISSN: 1434-6052. DOI: 10.48550/arXiv.2212.09379. arXiv: 2212.09379 [hep-ex].

[63] ATLAS Collaboration. "Public ATLAS Luminosity Results for Run-2 of the LHC". 2023. URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2 (visited on 07/27/2023).

[64] R Bailey and Paul Collier. "Standard Filling Schemes for Various LHC Operation Modes". Project Note. Sept. 19, 2003. CERN: LHC-PROJECT-NOTE-323. URL: https://cds.cern.ch/record/691782 (visited on 07/28/2023).

[65] ATLAS Collaboration. "The ATLAS Experiment at the CERN Large Hadron Collider". In: *Journal of Instrumentation* 3.08 (Aug. 2008). Ed. by IOP Publishing. ISSN: 1748-0221. DOI: 10.1088/1748-0221/3/08/S08003.

[66] ATLAS Collaboration, ed. *ATLAS: A 25-Year Insider Story of the LHC Experiment*. Advanced Series on Directions in High Energy Physics vol. 30. New Jersey: World Scientific, 2019. 359 pp. ISBN: 978-981-327-179-1. DOI: 10.1142/11030.

[67] ATLAS Collaboration. *The Performance of the ATLAS Detector*. Springer, 2010. ISBN: 978-3-642-22115-6. DOI: 10.1007/978-3-642-22116-3.

[68] ATLAS Collaboration, ed. *ATLAS Calorimeter Performance: Technical Design Report*. Technical design report. CERN, Jan. 13, 1997. 178 pp. ISBN: 92-9083-089-1. URL: https://cds.cern.ch/record/331059.

[69] Monica L. Dunford and Peter Jenni. "The ATLAS experiment". 2014. URL: http://www.scholarpedia.org/article/The_ATLAS_experiment (visited on 07/28/2023).

[70] Izaak Neutelings. URL: https://tikz.net/author/izaak/ (visited on 07/30/2023).

[71] A Lechner. "Particle Interactions with Matter". Mar. 11, 2017. URL: https://cas.web.cern.ch/sites/default/files/lectures/erice-2017/lechner.pdf (visited on 07/30/2023).

[72] ATLAS Collaboration. "Jet Energy Scale and Resolution Measured in Proton-Proton Collisions at $\sqrt{s} = 13$ TeV with the ATLAS Detector". In: *The European Physical Journal C* 81.8 (Aug. 3, 2021). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-021-09402-3. arXiv: 2007.02645 [hep-ex].

[73] Richard Wigmans. "On the Energy Resolution of Uranium and Other Hadron Calorimeters". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 259.3 (Sept. 15, 1987), pp. 389–429. ISSN: 0168-9002. DOI: 10.1016/0168-9002(87)90823-0.

[74] ATLAS Collaboration. "Approved plots Data Acquisition". 2023. URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/ApprovedPlotsDAQ (visited on 07/31/2023).

[75] ATLAS Collaboration. "Trigger Menu in 2018". Pub. Note. Oct. 14, 2019. CERN: ATL-DAQ-PUB-2019-001. URL: https://cds.cern.ch/record/2693402 (visited on 08/01/2023).

[76] The ATLAS Collaboration. "The ATLAS Simulation Infrastructure". In: *The European Physical Journal C* 70.3 (Sept. 25, 2010). ISSN: 1434-6044, 1434-6052. DOI: 10.1140/epjc/s10052-010-1429-9. arXiv: 1005.4568 [hep-ex].

[77] S. Agostinelli et al. "Geant4 - A Simulation Toolkit". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (July 1, 2003). DOI: 10.1016/S0168-9002(03)01368-8. URL: https://hal.in2p3.fr/in2p3-00020246.

[78] T Cornelissen et al. "Updates of the ATLAS Tracking Event Data Model (Release 13)". Pub. Note. June 14, 2007. CERN: ATL-SOFT-PUB-2007-003,ATL-COM-SOFT-2007-008. URL: https://cds.cern.ch/record/1038095 (visited on 07/31/2023).

[79] G. Borissov et al. "ATLAS Strategy for Primary Vertex Reconstruction during Run-2 of the LHC". In: *Journal of Physics: Conference Series* 664.7 (Dec. 2015). ISSN: 1742-6596. DOI: 10.1088/1742-6596/664/7/072041.

[80] J. T. Boyd. "LHC Run-2 and Future Prospects". Jan. 13, 2020. arXiv: 2001.04370 [hep-ex].

[81] ATLAS Collaboration. "Electron Reconstruction and Identification in the ATLAS Experiment Using the 2015 and 2016 LHC Proton-Proton Collision Data at $\sqrt{s} = 13$ TeV". In: *The European Physical Journal C* 79.8 (Aug. 3, 2019). ISSN: 1434-6044, 1434-6052. DOI: 10.1140/epjc/s10052-019-7140-6. arXiv: 1902.04655 [hep-ex].

[82] ATLAS Collaboration. "Electron and Photon Reconstruction and Performance in ATLAS Using a Dynamical, Topological Cell Clustering-Based Approach". Pub. Note. Dec. 22, 2017. CERN: ATL-PHYS-PUB-2017-022. URL: https://cds.cern.ch/record/2298955 (visited on 08/01/2023).

[83] ATLAS Collaboration. "Electron and Photon Performance Measurements with the ATLAS Detector Using the 2015-2017 LHC Proton-Proton Collision Data". In: *Journal of Instrumentation* 14.12 (Dec. 10, 2019). ISSN: 1748-0221. DOI: 10.1088/1748-0221/14/12/P12006. arXiv: 1908.00005 [hep-ex].

[84] ATLAS Collaboration. "Muon Reconstruction Performance of the ATLAS Detector in Proton–Proton Collision Data at $\sqrt{s}$=13 TeV". In: *The European Physical Journal C* 76.5 (May 23, 2016). ISSN: 1434-6044, 1434-6052. DOI: 10.1140/epjc/s10052-016-4120-y. arXiv: 1603.05598 [hep-ex].

[85] ATLAS Collaboration. "Muon Reconstruction and Identification Efficiency in ATLAS Using the Full Run 2 *pp* Collision Data Set at $\sqrt{s} = 13$ TeV". In: *The European Physical Journal C* 81.7 (July 5, 2021). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-021-09233-2. arXiv: 2012.00578 [hep-ex].

[86] ATLAS Collaboration. "Measurement of the Tau Lepton Reconstruction and Identification Performance in the ATLAS Experiment Using pp Collisions at $\sqrt{s} = 13$ TeV". Conf. Note. May 1, 2017. CERN: ATLAS-CONF-2017-029. URL: https://cds.cern.ch/record/2261772 (visited on 08/01/2023).

[87] ATLAS Collaboration. "Performance of Missing Transverse Momentum Reconstruction with the ATLAS Detector Using Proton-Proton Collisions at $\sqrt{s} = 13$ TeV". In: *The European Physical Journal C* 78.11 (Nov. 8, 2018). ISSN: 1434-6044, 1434-6052. DOI: 10.1140/epjc/s10052-018-6288-9. arXiv: 1802.08168 [hep-ex].

[88] ATLAS Collaboration. "In Situ Calibration of Large-Radius Jet Energy and Mass in 13 TeV ProtonProton Collisions with the ATLAS Detector". In: *The European Physical Journal C* 79.2 (Feb. 13, 2019). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-019-6632-8. arXiv: 1807.09477 [hep-ex].

[89] ATLAS Collaboration. "New Techniques for Jet Calibration with the ATLAS Detector". In: *The European Physical Journal C* 83.8 (Aug. 29, 2023). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-023-11837-9. arXiv: 2303.17312 [hep-ex].

[90] ATLAS Collaboration. "Energy scale calibration of b-tagged jets with ATLAS Run 2 data using $t\bar{t}$ lepton+jets events". Conf. note. Mar. 10, 2022. CERN: ATLAS-CONF-2022-004. URL: http://cds.cern.ch/record/2803523 (visited on 02/06/2023).

[91] ATLAS Collaboration. "Measurement of the ATLAS Detector Jet Mass Response using Forward Folding with 80 fb^{-1} of $\sqrt{13}$ TeV pp data". Conf. note. July 20, 2022. CERN: ATLAS-CONF-2020-022. URL: http://cds.cern.ch/record/2724442 (visited on 02/06/2023).

[92] ATLAS Collaboration. "Topological Cell Clustering in the ATLAS Calorimeters and Its Performance in LHC Run 1". In: *The European Physical Journal C* 77.7 (July 24, 2017). ISSN: 1434-6044, 1434-6052. DOI: 10.1140/epjc/s10052-017-5004-5. arXiv: 1603.02934 [hep-ex].

[93] Deepak Kar. *Experimental Particle Physics*. 2053-2563. IOP Publishing, 2019. ISBN: 978-0-7503-2112-9. DOI: 10.1088/2053-2563/ab1be6.

[94] ATLAS Collaboration. "Jet Reconstruction and Performance Using Particle Flow with the ATLAS Detector". In: *The European Physical Journal C* 77.7 (July 13, 2017). ISSN: 1434-6044, 1434-6052. DOI: 10.1140/epjc/s10052-017-5031-2. arXiv: 1703.10485 [hep-ex].

[95] Francesco Armando Di Bello et al. "Towards a Computer Vision Particle Flow". In: *The European Physical Journal C* 81.2 (Feb. 2, 2021). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-021-08897-0. arXiv: 2003.08863 [hep-ex].

[96] ATLAS Collaboration. "$R = 0.4$ jets input comparison and Monte Carlo calibration with the ATLAS Detector". Pub. note. Aug. 12, 2022. CERN: ATL-PHYS-PUB-2022-038. URL: https://cds.cern.ch/record/2824558 (visited on 02/06/2023).

[97] ATLAS Collaboration. "Constituent-level pileup mitigation performance using 2015 data". Conference note. July 20, 2017. CERN: ATLAS-CONF-2017-065. URL: https://cds.cern.ch/record/2281055 (visited on 02/06/2023).

[98] ATLAS Collaboration. "Optimisation of Large-Radius Jet Reconstruction for the ATLAS Detector in 13 TeV Proton-Proton Collisions". In: *The European Physical Journal C* 81.4 (Apr. 19, 2021). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-021-09054-3. arXiv: 2009.04986 [hep-ex].

[99] ATLAS Collaboration. "Improving jet substructure performance in ATLAS using Track-CaloClusters". Pub. note. July 10, 2017. CERN: ATL-PHYS-PUB-2017-015. URL: https://cds.cern.ch/record/2275636 (visited on 02/06/2023).

[100] ATLAS Collaboration. "Performance of pile-up mitigation techniques for jets in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector". In: *The European Physical Journal C* 76.11 (Oct. 27, 2016). ISSN: 1434-6044, 1434-6052. DOI: 10.1140/epjc/s10052-016-4395-z. arXiv: 1510.03823 [hep-ex].

[101] ATLAS Collaboration. "Identification and rejection of pile-up jets at high pseudorapidity with the ATLAS detector". In: *The European Physical Journal C* 77.9 (Sept. 2, 2017). ISSN: 1434-6044, 1434-6052. DOI: 10.1140/epjc/s10052-017-5081-5. arXiv: 1705.02211 [hep-ex].

[102] ATLAS Collaboration. "Selection of Jets Produced in 13 TeV Proton-Proton Collisions with the ATLAS Detector". Conf. Note. July 23, 2015. CERN: ATLAS-CONF-2015-029. URL: https://cds.cern.ch/record/2037702/files/ATLAS-CONF-2015-029.pdf (visited on 04/24/2023).

[103] Robert Hankache. "Jet calibration, cross-section measurements and New Physics searches with the ATLAS experiment within the Run 2 data". Ph.D. thesis in physics. LPNHE Paris, Sorbonne Université, CNRS, CERN, Sept. 24, 2019. URL: https://cds.cern.ch/record/2699000/files/CERN-THESIS-2019-193.pdf.

[104] ATLAS Collaboration. "Jet Cleaning in the Full Run-2 Data". Tech. Note. Jan. 16, 2023. CERN: ATL-COM-PHYS-2023-012. URL: https://cds.cern.ch/record/2846229/files/ATL-COM-PHYS-2023-012.pdf (visited on 05/31/2023).

[105] ATLAS Collaboration. "Jet Energy Measurement with the ATLAS Detector in Proton-Proton Collisions at $\sqrt{s} = 7$ TeV". In: *The European Physical Journal C* 73.3 (Mar. 2, 2013). ISSN: 1434-6044, 1434-6052. DOI: 10.1140/epjc/s10052-013-2304-2. arXiv: 1112.6426 [hep-ex].

[106] ATLAS Collaboration. "Jet Energy Measurement and Its Systematic Uncertainty in Proton-Proton Collisions at $\sqrt{s} = 7$ TeV with the ATLAS Detector". In: *The European Physical Journal C* 75.1 (Jan. 15, 2015). ISSN: 1434-6044, 1434-6052. DOI: 10.1140/epjc/s10052-014-3190-y. arXiv: 1406.0076 [hep-ex].

[107] ATLAS Collaboration. "Determination of Jet Calibration and Energy Resolution in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV Using the ATLAS Detector". In: *The European Physical Journal C* 80.12 (Dec. 1, 2020). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-020-08477-8. arXiv: 1910.04482 [hep-ex].

[108] Joint Committee for Guides in Metrology. "Evaluation of Measurement Data Guide to the Expression of Uncertainty in Measurement". Guide. Sept. 2008. BIPM: JCGM100:2008. URL: https://www.bipm.org/documents/20126/2071204/JCGM_100_2008_F.pdf (visited on 07/25/2023).

[109] ATLAS Collaboration. "Dependence of the Jet Energy Scale on the Particle Content of Hadronic Jets in the ATLAS Detector Simulation". Pub. Note. Apr. 29, 2022. CERN: ATL-PHYS-PUB-2022-021. URL: http://cds.cern.ch/record/2808016 (visited on 04/09/2023).

[110] ATLAS Collaboration. "Jet Energy Scale uncertainties ($R = 0.4$ PFlow) with updated flavour treatment". Tech. rep. Aug. 15, 2022. CDS: JETM-2022-005. URL: https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/JETM-2022-005/ (visited on 02/06/2023).

[111] ATLAS Collaboration. "Single Hadron Response Measurement and Calorimeter Jet Energy Scale Uncertainty with the ATLAS Detector at the LHC". In: *The European Physical Journal C* 73.3 (Mar. 2, 2013). ISSN: 1434-6044, 1434-6052. DOI: 10.1140/epjc/s10052-013-2305-1. arXiv: 1203.1302 [hep-ex].

[112] ATLAS Collaboration. "A Measurement of the Calorimeter Response to Single Hadrons and Determination of the Jet Energy Scale Uncertainty Using LHC Run-1 pp-Collision Data with the ATLAS Detector". In: *The European Physical Journal C* 77.1 (Jan. 13, 2017). ISSN: 1434-6044, 1434-6052. DOI: 10.1140/epjc/s10052-016-4580-0. arXiv: 1607.08842 [hep-ex].

[113] ATLAS Collaboration. "Measurement of the Energy Response of the ATLAS Calorimeter to Charged Pions from $W^{\pm} \to \tau^{\pm}(\to \pi^{\pm}\nu_{\tau})\nu_{\tau}$ Events in Run 2 Data". In: *The European Physical Journal C* 82.3 (Mar. 14, 2022). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-022-10117-2. arXiv: 2108.09043 [hep-ex].

[114] ATLAS Collaboration. "Electron and Photon Energy Calibration with the ATLAS Detector Using 2015-2016 LHC Proton-Proton Collision Data". In: *Journal of Instrumentation* 14.03 (Mar. 18, 2019). ISSN: 1748-0221. DOI: 10.1088/1748-0221/14/03/P03017. arXiv: 1812.03848 [hep-ex].

[115] Steven Schramm. "Pre-HCW, In Situ Measurements for Jets". Aug. 24, 2022. URL: https://indico.cern.ch/event/1188525/contributions/4994856/attachments/2496571/4287876/StevenSchramm-InSituMeasurements.pdf (visited on 08/24/2022).

[116] ATLAS Collaboration. "Search for New Resonances in Mass Distributions of Jet Pairs Using 139 fb^{-1} of pp Collisions at $\sqrt{s} = 13$ TeV with the ATLAS Detector". In: *Journal of High Energy Physics* 2020.3 (Mar. 25, 2020). ISSN: 1029-8479. DOI: 10.1007/JHEP03(2020)145. arXiv: 1910.08447 [hep-ex].

[117] CMS Collaboration. "Jet Energy Scale and Resolution in the CMS Experiment in pp Collisions at 8 TeV". In: *Journal of Instrumentation* 12.02 (Feb. 22, 2017). ISSN: 1748-0221. DOI: 10.1088/1748-0221/12/02/P02014. arXiv: 1607.03663 [hep-ex].

[118] CMS Collaboration. "Jet Energy Scale and Resolution Performance with 13 TeV Data Collected by CMS in 2016-2018". Tech. note. Apr. 9, 2020. CERN: CMS-DP2020_019. URL: https://cds.cern.ch/record/2715872/files/DP2020_019.pdf (visited on 02/07/2023).

[119] CMS Collaboration. "Jet Energy Scale and Resolution Measurement with Run 2 Legacy Data Collected by CMS at 13 TeV". Performance note. Sept. 27, 2021. CERN: CMS-DP2021_033. URL: http://cds.cern.ch/record/2792322/files/DP2021_033.pdf (visited on 02/07/2023).

[120] Chris Young. "The Story of Batman Ears - a Short Summary". Aug. 28, 2017. URL: https://indico.cern.ch/event/642438/contributions/2696751/attachments/1517034/2367905/Presentation.pdf (visited on 04/24/2023).

[121] James Lacey. "MC Weights and Quality Cuts in 2012". July 31, 2012. URL: https://indico.cern.ch/event/201937/contributions/1487715/attachments/303276/423637/JLacey_MC12_JZXW_weights_and_CleaningCuts.pdf (visited on 04/18/2023).

[122] Tancredi Carli and Tigran Mkrtchyan. "ATLAS Jet Trigger Performance in Run-2". Tech. Note. Dec. 16, 2019. CERN: ATL-COM-DAQ-2019-209. URL: https://cds.cern.ch/record/2704391/files/ATL-COM-DAQ-2019-209.pdf (visited on 01/28/2022).

[123] Z Marshall. "Re-Defining the Standard QCD Di-Jet Samples: Event Weights Make Me Grumpy". Aug. 2, 2011. (Visited on 04/25/2023).

[124] Z Marshall. "Re-Re-Defining the Standard QCD Di-Jet Samples: Beginning to like Event Weights". May 17, 2015. (Visited on 04/25/2023).

[125] Victor Lendermann et al. "Combining Triggers in HEP Data Analysis". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 604.3 (June 11, 2009), pp. 707–718. ISSN: 01689002. DOI: 10.1016/j.nima.2009.03.173. arXiv: 0901.4118 [hep-ex].

[126] Leslie Kish. *Survey Sampling*. Ed. by Inc. New York: John Wiley & Sons. 1965. ISBN: 0-471-10949-5.

[127] Kevin H. Knuth. "Optimal Data-Based Binning for Histograms". Sept. 16, 2013. DOI: 10.48550/arXiv.physics/0605197. arXiv: physics/0605197 [physics.data-an].

[128] ATLAS Collaboration. "Determination of the Parton Distribution Functions of the Proton Using Diverse ATLAS Data from pp Collisions at $\sqrt{s} = 7$, 8 and 13 TeV". In: *The European Physical Journal C* 82.5 (May 13, 2022). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-022-10217-z. arXiv: 2112.11266 [hep-ex, hep-ph].

[129] ATLAS Collaboration. "Determination of the Strong Coupling Constant from Transverse Energy−Energy Correlations in Multijet Events at $\sqrt{s} = 13$ TeV with the ATLAS Detector". In: *Journal of High Energy Physics* 2023.7 (July 11, 2023). ISSN: 1029-8479. DOI: 10.48550/arXiv.2301.09351. arXiv: 2301.09351 [hep-ex].

[130] CMS Collaboration. "Constraints on Parton Distribution Functions and Extraction of the Strong Coupling Constant from the Inclusive Jet Cross Section in Pp Collisions at Sqrt(s) = 7 TeV". In: *The European Physical Journal C* 75.6 (June 26, 2015). ISSN: 1434-6044, 1434-6052. DOI: 10.1140/epjc/s10052-015-3499-1. arXiv: 1410.6765 [hep-ex].

[131] CMS Collaboration. "Measurement and QCD Analysis of Double-Differential Inclusive Jet Cross Sections in Proton-Proton Collisions at $\sqrt{s} = 13$ TeV". In: *Journal of High Energy Physics* 2022.12 (Dec. 7, 2022). ISSN: 1029-8479. DOI: 10.1007/JHEP12(2022)035. arXiv: 2111.10431 [hep-ex].

[132] ATLAS Collaboration. "Measurement of Dijet Cross Sections in pp Collisions at 7 TeV Centre-of-Mass Energy Using the ATLAS Detector". In: *Journal of High Energy Physics* 2014.5 (May 14, 2014). ISSN: 1029-8479. DOI: 10.1007/JHEP05(2014)059. arXiv: 1312.3524 [hep-ex].

[133] ATLAS Collaboration. "Measurement of Inclusive Jet and Dijet Cross Sections in Proton-Proton Collisions at 7 TeV Centre-of-Mass Energy with the ATLAS Detector". In: *The European Physical Journal C* 71.2

(Feb. 3, 2011). ISSN: 1434-6044, 1434-6052. DOI: 10.1140/epjc/s10052-010-1512-2. arXiv: 1009.5908 [hep-ex].

[134] ATLAS Collaboration. "Measurement of Inclusive Jet and Dijet Production in *pp* Collisions at $\sqrt{s} = 7$ TeV Using the ATLAS Detector". In: *Physical Review D* 86.1 (July 24, 2012). ISSN: 1550-7998, 1550-2368. DOI: 10.1103/PhysRevD.86.014022. arXiv: 1112.6297 [hep-ex].

[135] ATLAS Collaboration. "Measurement of the Inclusive Jet Cross-Section in Proton-Proton Collisions at $\sqrt{s} =$ 7 TeV Using 4.5 fb^{-1} of Data with the ATLAS Detector". In: *Journal of High Energy Physics* 2015.2 (Feb. 24, 2015). ISSN: 1029-8479. DOI: 10.1007/JHEP02(2015)153. arXiv: 1410.8857 [hep-ex].

[136] ATLAS Collaboration. "Measurement of the Inclusive Jet Cross-Sections in Proton–Proton Collisions at $\sqrt{s} = 8$ TeV with the ATLAS Detector". In: *Journal of High Energy Physics* 2017.9 (Sept. 5, 2017). ISSN: 1029-8479. DOI: 10.1007/JHEP09(2017)020. arXiv: 1706.03192 [hep-ex].

[137] CMS Collaboration. "Measurement of the Inclusive Jet Cross Section in *pp* Collisions at $\sqrt{s} = 2.76$ TeV". In: *The European Physical Journal C* 76.5 (May 12, 2016). ISSN: 1434-6044, 1434-6052. DOI: 10.1140/epjc/s10052-016-4083-z. arXiv: 1512.06212 [hep-ex].

[138] CMS Collaboration. "Measurement of the Inclusive Jet Cross Section in *pp* Collisions at $\sqrt{s} = 7$ TeV". In: *Physical Review Letters* 107.13 (Sept. 19, 2011). ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.107.132001. arXiv: 1106.0208 [hep-ex].

[139] CMS Collaboration. "Measurements of Differential Jet Cross Sections in Proton-Proton Collisions at Sqrt(s) = 7 TeV with the CMS Detector". In: *Physical Review D* 87.11 (June 3, 2013). ISSN: 1550-7998, 1550-2368. DOI: 10.1103/PhysRevD.87.112002. arXiv: 1212.6660 [hep-ex].

[140] CMS Collaboration. "Measurement and QCD Analysis of Double-Differential Inclusive Jet Cross-Sections in *pp* Collisions at $\sqrt{s} = 8$ TeV and Ratios to 2.76 and 7 TeV". In: *Journal of High Energy Physics* 2017.3 (Mar. 29, 2017). ISSN: 1029-8479. DOI: 10.1007/JHEP03(2017)156. arXiv: 1609.05331 [hep-ex].

[141] CMS Collaboration. "Measurement of the Double-Differential Inclusive Jet Cross Section in Proton-Proton Collisions at $\sqrt{s} = 13$ TeV". In: *The European Physical Journal C* 76.8 (Aug. 11, 2016). ISSN: 1434-6044, 1434-6052. DOI: 10.1140/epjc/s10052-016-4286-3. arXiv: 1605.04436 [hep-ex].

[142] CMS Collaboration. "Measurement of the Differential Dijet Production Cross Section in Proton-Proton Collisions at $\sqrt{s}$ =7 TeV". In: *Physics Letters B* 700.3-4 (June 13, 2011). ISSN: 03702693. DOI: 10.1016/j.physletb.2011.05.027. arXiv: 1104.1693 [hep-ex].

[143] CMS Collaboration. "Measurement of the Triple-Differential Dijet Cross Section in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV and Constraints on Parton Distribution Functions". In: *The European Physical Journal C* 77.11 (Nov. 7, 2017). ISSN: 1434-6044, 1434-6052. DOI: 10.1140/epjc/s10052-017-5286-7. arXiv: 1705.02628 [hep-ex].

[144] CMS Collaboration. "Measurements of the Differential Jet Cross Section as a Function of the Jet Mass in Dijet Events from Proton-Proton Collisions at $\sqrt{s} = 13$ TeV". In: *Journal of High Energy Physics* 2018.11 (Nov. 20, 2018). ISSN: 1029-8479. DOI: 10.1007/JHEP11(2018)113. arXiv: 1807.05974 [hep-ex].

[145] ATLAS Collaboration. "Measurement of Three-Jet Production Cross-Sections in *pp* Collisions at 7 TeV Centre-of-Mass Energy Using the ATLAS Detector". In: *The European Physical Journal C* 75.5 (May 27, 2015). ISSN: 1434-6044, 1434-6052. DOI: 10.1140/epjc/s10052-015-3363-3. arXiv: 1411.1855 [hep-ex].

[146] CMS Collaboration. "Measurement of the Inclusive 3-Jet Production Differential Cross Section in Proton-Proton Collisions at 7 TeV and Determination of the Strong Coupling Constant in the TeV Range". In: *The European Physical Journal C* 75.5 (May 1, 2015). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-015-3376-y. arXiv: 1412.1633 [hep-ex].

[147] CMS Collaboration. "Distributions of Topological Observables in Inclusive Three- and Four-Jet Events in *pp* Collisions at $\sqrt{s} = 7$ TeV". In: *The European Physical Journal C* 75.7 (July 1, 2015). ISSN: 1434-6044, 1434-6052. DOI: 10.1140/epjc/s10052-015-3491-9. arXiv: 1502.04785 [hep-ex].

[148] ATLAS Collaboration. "Measurement of Four-Jet Differential Cross Sections in $\sqrt{s} = 8$ TeV Proton-Proton Collisions Using the ATLAS Detector". In: *Journal of High Energy Physics* 2015.12 (Dec. 16, 2015). ISSN: 1029-8479. DOI: 10.1007/JHEP12(2015)105. arXiv: 1509.07335 [hep-ex].

[149] ATLAS Collaboration. "Measurements of Multijet Event Isotropies Using Optimal Transport with the ATLAS Detector". In: *Journal of High Energy Physics* 2023.10 (Oct. 10, 2023). ISSN: 1029-8479. DOI: 10.1007/jhep10(2023)060. arXiv: 2305.16930 [hep-ex].

[150] Sonia Leach. "Singular Value Decomposition, A Primer". URL: http://people.csail.mit.edu/hasinoff/320/SingularValueDecomposition.pdf (visited on 06/22/2023).

[151] L. B. Lucy. "An Iterative Technique for the Rectification of Observed Distributions". In: *The Astronomical Journal* 79 (June 1974), p. 745. ISSN: 00046256. DOI: 10.1086/111605.

[152] Bogdan Malaescu. "An Iterative, Dynamically Stabilized Method of Data Unfolding". July 22, 2009. DOI: 10.48550/arXiv.0907.3791. arXiv: 0907.3791 [hep-ex, physics.data-an].

[153] Bogdan Malaescu. "An Iterative, Dynamically Stabilized(IDS) Method of Data Unfolding". June 15, 2011. DOI: 10.5170/CERN-2011-006.271. arXiv: 1106.3107 [hep-ex].

[154] G. D'Agostini. "A Multidimensional Unfolding Method Based on Bayes' Theorem". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 362.2 (Aug. 15, 1995), pp. 487–498. ISSN: 0168-9002. DOI: 10.1016/0168-9002(95)00274-X.

[155] G. D'Agostini. "Improved Iterative Bayesian Unfolding". Oct. 4, 2010. arXiv: 1010.0632 [physics.data-an].

[156] Jason Gallicchio and Yang-Ting Chien. "Quit Using Pseudorapidity, Transverse Energy, and Massless Constituents". Feb. 14, 2018. arXiv: 1802.05356 [hep-ph].

[157] ATLAS Collaboration. "Evaluating Statistical Uncertainties and Correlations Using the Bootstrap Method". Pub. Note. Sept. 6, 2021. CERN: ATL-PHYS-PUB-2021-011. URL: https://cds.cern.ch/record/2759945/files/ATL-PHYS-PUB-2021-011.pdf (visited on 07/29/2022).

[158] Michaela Queitsch-Maitland. "JER Recommendation". Aug. 29, 2018. URL: https://indico.cern.ch/event/752759/contributions/3119136/attachments/1706979/2750689/JetETmiss_JERRecommendation_20180829.pdf (visited on 06/26/2023).

[159] E. Todesco and J. Wenninger. "Large Hadron Collider Momentum Calibration and Accuracy". In: *Physical Review Accelerators and Beams* 20.8 (Aug. 21, 2017). DOI: 10.1103/PhysRevAccelBeams.20.081003.

[160] Tancredi Carli, Gavin P. Salam, and Frank Siegert. "A Posteriori Inclusion of PDFs in NLO QCD Final-State Calculations". Oct. 25, 2005. arXiv: hep-ph/0510324 [hep-ph].

[161] Sayipjamal Dulat et al. "New Parton Distribution Functions from a Global Analysis of Quantum Chromodynamics". In: *Physical Review D* 93.3 (Feb. 16, 2016). ISSN: 2470-0010, 2470-0029. DOI: 10.1103/PhysRevD.93.033006. arXiv: 1506.07443 [hep-ex, hep-ph, nucl-ex].

[162] Mengshi Yan et al. "A CT18 Global PDF Fit at the Leading Order in QCD". In: *Physical Review D* 107.11 (June 1, 2023). ISSN: 2470-0010, 2470-0029. DOI: 10.1103/PhysRevD.107.116001. arXiv: 2205.00137 [hep-ph].

[163] Richard D. Ball et al. "The Path to Proton Structure at One-Percent Accuracy". In: *The European Physical Journal C* 82.5 (May 11, 2022). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-022-10328-7. arXiv: 2109.02653 [hep-ex, hep-ph].

[164] Andy Buckley et al. "LHAPDF6: Parton Density Access in the LHC Precision Era". In: *The European Physical Journal C* 75.3 (Mar. 20, 2015). ISSN: 1434-6044, 1434-6052. DOI: 10.1140/epjc/s10052-015-3318-8. arXiv: 1412.7420 [hep-ph].

[165] Gavin P. Salam and Juan Rojo. "A Higher Order Perturbative Parton Evolution Toolkit (HOPPET)". In: *Computer Physics Communications* 180.1 (Jan. 2009), pp. 120–156. ISSN: 00104655. DOI: 10.1016/j.cpc.2008.08.010. arXiv: 0804.3755 [hep-ph].

[166] Simone Amoroso et al. "Recommendations on the treatment of theoretical systematic uncertainties in statistical analysis of ATLAS data". Tech. Note. Apr. 2020. CERN: ATL-COM-PHYS-2020-272. URL: https://cds.cern.ch/record/2715689 (visited on 08/02/2023).

Appendix A

Appendix for the η-intercalibration

A.1 Brief description of the code for the η-intercalibration

The code for the η-intercalibration can be found in the following gitlab repository: https://gitlab.cern.ch/atlas-jetetmiss-jesjer/Insitu/Dijet-Insitu-Calibration-and-Resolution/.

The first sub-package, "`EtaInterCalxAODAnalysis`", runs on the computing grid of CERN on the dAODs (derivations of Analysis Object Datasets, *i.e.* on the files containing reconstructed events) provided by the JetETmiss group of ATLAS, more specifically on the dAODs that only contain dijet events (DAOD_JETM1). It produces smaller files usually called "NTuples" (sometimes also referred to as "miniAODs", "microAODs", "tiny trees", because they actually do not contain ROOT NTuples but TTrees). This piece of code calibrates the jets, applies the selection criteria (either the nominal selection or the systematic variations of the selection), and only keeps relevant quantities.

The second sub-package, "`GenerateHistograms`", generates 3D histograms of the asymmetry $\mathcal{A}$ *vs.* η_{left} and η_{right}, in each p_T^{avg} bin. It chooses either the nominal selection or a systematic variation of the selection, applies the slice normalization for MC or the trigger normalization for data. In case the statistical uncertainties on the systematic uncertainties are propagated with the bootstrap method, it also produces the bootstrap replicas of the histograms.

The third sub-package, "`MinimizeResponseMatrix`"[1], finds the mean asymmetry $\langle\mathcal{A}\rangle$ in each $(p_T^{\text{avg}}, \eta_{\text{left}}, \eta_{\text{right}})$ bin, by a Gaussian fit when possible. An automatic rebinning in asymmetry is performed in case the effective statistics in this bin is too low. Then it computes the relative response $\mathcal{R}$ from the mean asymmetry and performs a constraint combination to obtain the intercalibration factors c in the (p_T^{avg}, η) binning by a minimization of a χ^2-like function. Finally, the intercalibration coefficients are rescaled such that the average of the intercalibration factors in the central region $|\eta| < 0.8$ is equal to 1 on average. The code is able to perform simple η bin combinations to combine the constraints of two adjacent bins in case the effective statistics of two adjacent bins is lower than a threshold and if the effective statistics of the combined bin is higher than that threshold. This allows to get some constraints that would otherwise be lost, and to improve the overall statistical uncertainties.

The final sub-package, "`CalibrationAndUncertainties`", computes the calibration factor $\mathcal{C}$ as ratios of intercalibration factors, combines the nominal calibration, the systematic and the statistical uncertainties, and performs a smoothing and interpolation into the final binning with a Gaussian kernel.

A.2 Differences for the large-R jets

Large-R jets are reconstructed from Unified Flow Objects (UFOs, see Section 3.1.1.4), with Constituent Subtraction and Soft-Killer for pile-up mitigation (CS+SK, see Section 3.1.1.3), with the anti-k_t algorithm with a radius of $R = 1.0$ (see Section 3.1.2).

In the context of the η-intercalibration for large-R jets:

- The two leading jets are the two large-R jets with the highest p_T;
- The third jet is the highest-p_T small-R jet (PFlow, $R = 0.4$) that is not geometrically matched to one of the two leading large-R jets, with the geometrical condition $\Delta R > 0.75$.[2]

[1] From a technical point of view, the code in `MinimiseResponseMatrix` would benefit from being re-written from scratch. The code has become so unusable that it can only serve one purpose: being a proof-of-concept that the method works. Now that we understand how and why it works, it is essential to rewrite it using good programming practices, using the detailed description presented in this chapter as guidelines.

[2] $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$

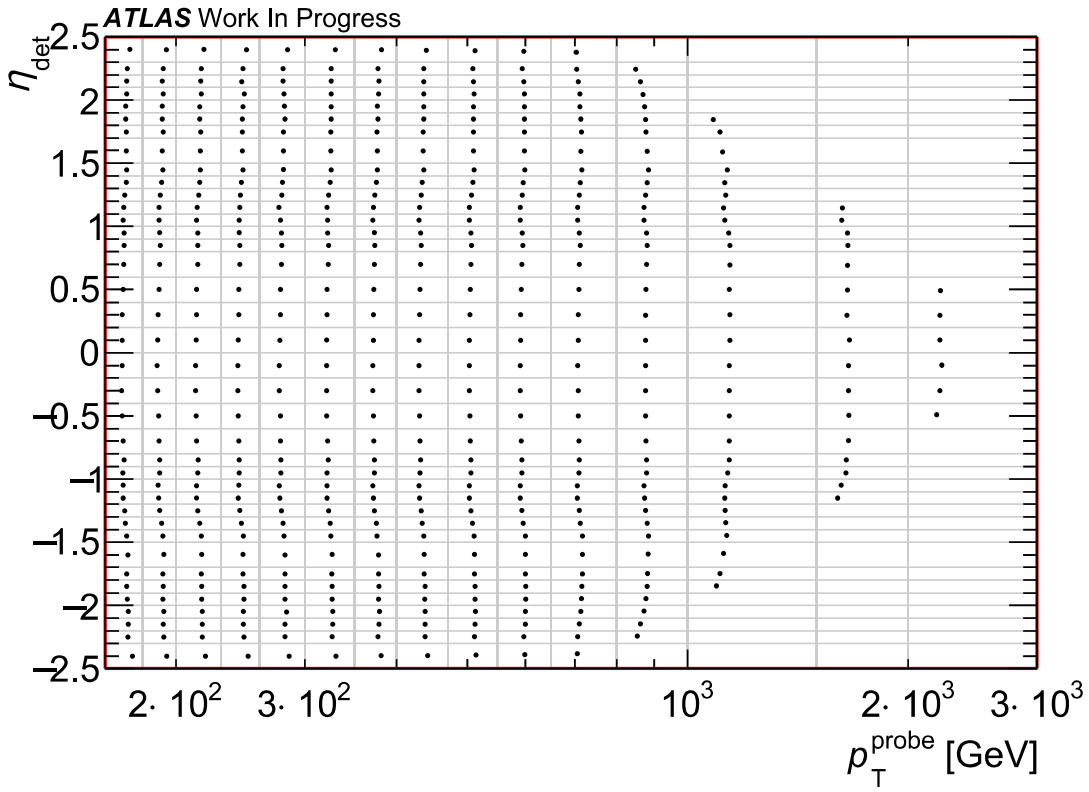


Figure A.1: (p_T^{avg}, η) binning for the large-R η-intercalibration. Some areas of phase space are not covered, either because they are unphysical (the $\sqrt{s} = 13$ TeV energy in the center of mass imposes constraints: $p_T < \frac{13\text{ TeV}}{2\cosh\eta}$) or because there is not enough statistics in those regions. The dots show the barycenter of the bins that are used. Plot made by Oleksandr Burlayenko.

The (p_T^{avg}, η) binning is shown in Fig. A.1, and different triggers are used. The association between p_T^{avg} bins and triggers is listed in table Table A.1. The η range being reduced to $-2.5 < \eta < 2.5$, only central triggers are used for large-R jets.

p_T^{avg} bin (in GeV)	Trigger used for 2015 and 2016	Trigger used for 2017 and 2018
$[160, 180[$	HLT_j85	HLT_j85
$[180, 200[$		
$[200, 230[$	HLT_j110	HLT_j110
$[230, 260[$		
$[260, 300[$		
$[300, 350[$	HLT_j175	HLT_j175
$[350, 400[$	HLT_j300_a10_lcw_L1J75	
$[400, 470[$	HLT_j360_a10_lcw_L1J100	
$[470, 550[$	HLT_j420_a10_lcw_L1J100	HLT_j390_a10t_lcw_jes_L1J100
$[550, 650[$		HLT_j460_a10_lcw_subjes_L1J100
$[650, 800[$		
$[800, 1000[$		
$[1000, 1500[$		
$[1500, 2000[$		
$[2000, 3000[$		

Table A.1: p_T^{avg} binning and trigger combination used in each bin for large-R jets.

Large-R jets start at a higher p_T^{avg} (160 GeV compared to 25 GeV for small-R jets), so there is no need for pile-up rejection using the JVT.

The truth level of MC is not used, because the studies shown in this thesis for small-R jets have not been conducted for large-R jets. Namely, the absolute implementation has not been studied for large-R jets, and the MC modeling uncertainty is still using the reco-level MC.

A.3 Internal tags of the datasets used, for users within ATLAS

The datasets are provided by the JetETmiss group of ATLAS. The derivations, called DAOD_JETM1, are the dijet ones.

For data, the internal production tag used is "p4061".

MC generator	Matrix elements	Parton shower	Hadronization	PDFs	Internal DataSet IDentifiers of ATLAS (mcChannelNumber)
Pythia8.230	$2 \to 2$ LO	Dipole-style p_T-ordered	Lund string	NNPDF 2.3	364700 - 364712
Powheg V2 + Pythia8.235	$2 \to 2$ NLO	Dipole-style p_T-ordered	Lund string	NNPDF30NLO	600624 - 600631
Powheg + Herwig7	$2 \to 2$ NLO	Angle-ordered	Cluster	NNPDF30NLO	600957 - 600964
Herwig 7.1.3	$2 \to 2$ NLO	Angle-ordered	Cluster	MMHT2014NLO	364934, 364922 - 364929
Herwig 7.1.3	$2 \to 2$ NLO	Dipole	Cluster	MMHT2014NLO	364933, 364902 - 364909
Sherpa 2.2.5	$2 \to 2$ LO	Dipole	Cluster	CT14NNLO	364677 - 364685
Sherpa 2.2.5	$2 \to 2$ LO	Dipole	Lund string	CT14NNLO	364686 - 364694

Table A.2: Characteristics of the different MC generators used here. See Section 1.3.1 for explanations.

The different MC generators used are listed in Table A.2 with their DataSet IDentifiers (DSIDs, also known as mcChannelNumber). The internal tags are the following ones: the event generation tag is "e7954", the simulation tag is "s3126", the reconstruction tag is either

- "r9364" for datasets called hereafter "mc16a", corresponding to the conditions present during data-taking periods 2015 and 2016
- "r10201" for datasets called hereafter "mc16d", corresponding to the conditions present during data-taking period 2017
- "r10724" for datasets called hereafter "mc16e", corresponding to the conditions present during data-taking period 2018

and the production tag is either "p4694" for Powheg+Pythia8, or "p4870" for all the others MCs.

Pile-up is added (as described in Section 2.2), and the detection is fully simulated with Geant4 (internal setting: s3126), providing the "reconstructed" ("reco") level.

A.4 Computation of the barycenters of the (p_T^{avg}, η) bins

The barycenter of one bin $\left(\left[{p_T^{\text{avg}}}_{\text{min}}, {p_T^{\text{avg}}}_{\text{max}}\right], [\eta_{\text{min}}, \eta_{\text{max}}]\right)$ could simply be computed as the weighted mean of the events entering that bin:

$$\overline{\eta} = \frac{\sum_{i \in \text{events in that bin}} w_i \eta_i}{\sum_{i \in \text{events}} w_i} \tag{A.1}$$

$$\overline{p_T^{\text{avg}}} = \frac{\sum_{i \in \text{events in that bin}} w_i {p_T^{\text{avg}}}_i}{\sum_{i \in \text{events}} w_i} \tag{A.2}$$

For a historical reason of which I am not aware, the barycenter of the bin is not computed this way. Probably technical reasons: at the time of filling the histograms, the information of which JZ slice (for MC) or which trigger (for data) is used in not known.

Some histograms with very fine (η, p_T^{avg}) bins are filled for each JZ slice (for MC) or for each trigger combination (for data). The barycenter of one normal (η, p_T^{avg}) is computed as the barycenter of all the fine (η, p_T^{avg}) bins contained in the normal bin:

$$\overline{\eta} = \frac{\sum_{e=e_{\text{min}}}^{e_{\text{max}}} \sum_{p=p_{\text{min}}}^{p_{\text{max}}} \sum_{h \in \text{histograms}} w_{eph} \eta_{\text{center of bin } (h,e)}}{\sum_e \sum_p \sum_h w_{eph}} \tag{A.3}$$

$$\overline{p_T^{\text{avg}}} = \frac{\sum_{e=e_{\text{min}}}^{e_{\text{max}}} \sum_{p=p_{\text{min}}}^{p_{\text{max}}} \sum_{h \in \text{histograms}} w_{eph} p_{T\ \text{center of bin } (h,p)}^{\text{avg}}}{\sum_e \sum_p \sum_h w_{eph}} \tag{A.4}$$

Then the uncertainty on those barycenters is evaluated as:

$$N_{\text{eff}} = \frac{\left(\sum_e \sum_p \sum_h w_{eph}\right)^2}{\sum_e \sum_p \sum_h w_{eph}^2} \tag{A.5}$$

$$\Delta\overline{\eta} = \sqrt{\frac{\frac{\sum_e \sum_p \sum_h w_{eph} \eta^2_{\text{center of bin } (h,e)}}{\sum_e \sum_p \sum_h w_{eph}} - \overline{\eta}^2}{N_{\text{eff}}}} \tag{A.6}$$

$$\Delta\overline{p_T^{\text{avg}}} = \sqrt{\frac{\frac{\sum_e \sum_p \sum_h w_{eph} p_{T\ \text{center of bin } (h,p)}^{\text{avg}2}}{\sum_e \sum_p \sum_h w_{eph}} - \overline{p_T^{\text{avg}}}^2}{N_{\text{eff}}}} \tag{A.7}$$

A.5 Analytic solution for the matrix method when combining η bins

Two η^{left} bins can be combined, or two η^{right} bins.

Linear combination of constraints
One constraint is obtained for the combined bin, and then the intercalibration coefficients corresponding to those two bins are combined linearly into one single constraint, according to their relative statistical weight.

When two left bins i and $i+1$ are combined (with $1 \leq i$ and $i+1 < j$), the constraint is expressed:

$$X_{(i,i+1),j}\left(c_i, c_{i+1}, c_j\right) = \left(\frac{\frac{N_i c_i + N_{i+1} c_{i+1}}{N_i + N_{i+1}} \mathcal{R}_{(i,i+1),j} - c_j}{\Delta\mathcal{R}_{(i,i+1),j}}\right)^2 \tag{A.8}$$

instead of

$$X_{i,j} + X_{i+1,j} = \left(\frac{c_i \mathcal{R}_{ij} - c_j}{\Delta\mathcal{R}_{ij}}\right)^2 + \left(\frac{c_{i+1} \mathcal{R}_{i+1,j} - c_j}{\Delta\mathcal{R}_{i+1,j}}\right)^2 \tag{A.9}$$

and similarly for the combination of two right bins j and $j+1$ (with $i < j$ and $j+1 \leq N$):

$$X_{i,(j,j+1)}\left(c_i, c_j, c_{j+1}\right) = \left(\frac{c_i \mathcal{R}_{i,(j,j+1)} - \frac{N_j c_j + N_{j+1} c_{j+1}}{N_j + N_{j+1}}}{\Delta\mathcal{R}_{i,(j,j+1)}}\right)^2 \tag{A.10}$$

Subsets of bin combinations
The set of possible combinations of left bins I is expressed:

$$I = \left\{\left(\left(i, i+1\right), j\right) \in [\![1, N]\!]^2 \times [\![1, N]\!] \middle| i+1 < j\right\} \tag{A.11}$$

and the set of possible combinations of right bins J is expressed:

$$J = \left\{\left(i, \left(j, j+1\right)\right) \in [\![1, N]\!] \times [\![1, N]\!]^2 \middle| i < j \wedge j+1 \leq N\right\} \tag{A.12}$$

The sets of actual left or right bin combinations are only subsets of the possible left or right bin combinations, respectively:

$$I' \subset I \tag{A.13}$$

$$J' \subset J \tag{A.14}$$

Terms to add to S
The terms to add to S, the function to be minimized, are:

$$S^{I'} = \sum_{((i,i+1),j) \in I'} X_{(i,i+1),j}\left(c_i, c_{i+1}, c_j\right) \tag{A.15}$$

$$= \sum_{((i,i+1),j) \in I'} \left(\frac{\frac{N_i c_i + N_{i+1} c_{i+1}}{N_i + N_{i+1}} \mathcal{R}_{(i,i+1),j} - c_j}{\Delta\mathcal{R}_{(i,i+1),j}}\right)^2 \tag{A.16}$$

$$S^{J'} = \sum_{(i,(j,j+1)) \in J'} X_{i,(j,j+1)}\left(c_i, c_j, c_{j+1}\right) \tag{A.17}$$

$$= \sum_{(i,(j,j+1)) \in J'} \left(\frac{c_i \mathcal{R}_{i,(j,j+1)} - \frac{N_j c_j + N_{j+1} c_{j+1}}{N_j + N_{j+1}}}{\Delta\mathcal{R}_{i,(j,j+1)}}\right)^2 \tag{A.18}$$

Computing U

In the $U = (u_{kl})_{(k,l)\in[\![1,N]\!]^2}$ matrix, u_{kl} contains the terms of the form $\frac{\partial S}{\partial c_k} = u_{kl}c_l$. More precisely, $\forall k \in [\![1, N]\!]$, $\frac{\partial S}{\partial c_k} = \frac{\partial S^F}{\partial c_k} + \frac{\partial S^{I'}}{\partial c_k} + \frac{\partial S^{J'}}{\partial c_k} + \frac{\partial S_\lambda}{\partial c_k} = 2\sum_{l=1}^{N} u_{kl}c_l - \frac{2\lambda}{N}$.

Let's differentiate a left bin combination $X_{(i,i+1),j}(c_i, c_{i+1}, c_j)$ with respect to c_i, c_{i+1} and c_j:

$$\frac{1}{2}\frac{\partial X_{(i,i+1),j}}{\partial c_i}(c_i, c_{i+1}, c_j) = \frac{1}{\Delta\mathcal{R}_{(i,i+1),j}{}^2}\left(R_{(i,i+1),j}{}^2\frac{N_i^2 c_i + N_i N_{i+1} c_{i+1}}{(N_i + N_{i+1})^2} - \mathcal{R}_{(i,i+1),j}\frac{N_i}{N_i + N_{i+1}}c_j\right) \tag{A.19}$$

$$\frac{1}{2}\frac{\partial X_{(i,i+1),j}}{\partial c_{i+1}}(c_i, c_{i+1}, c_j) = \frac{1}{\Delta\mathcal{R}_{(i,i+1),j}{}^2}\left(R_{(i,i+1),j}{}^2\frac{N_{i+1}^2 c_{i+1} + N_i N_{i+1} c_i}{(N_i + N_{i+1})^2} - \mathcal{R}_{(i,i+1),j}\frac{N_{i+1}}{N_i + N_{i+1}}c_j\right) \tag{A.20}$$

$$\frac{1}{2}\frac{\partial X_{(i,i+1),j}}{\partial c_j}(c_i, c_{i+1}, c_j) = \frac{1}{\Delta\mathcal{R}_{(i,i+1),j}{}^2}\left(-\mathcal{R}_{(i,i+1),j}\frac{N_i c_i + N_{i+1} c_{i+1}}{N_i + N_{i+1}} + c_j\right) \tag{A.21}$$

Therefore, the constraint $X_{(i,i+1),j}(c_i, c_{i+1}, c_j)$ contributes to 9 terms in the U matrix:

$$u_{i,i} += \frac{\mathcal{R}_{(i,i+1),j}{}^2}{\Delta\mathcal{R}_{(i,i+1),j}{}^2}\frac{N_i^2}{(N_i + N_{i+1})^2} \tag{A.22}$$

$$u_{i,i+1} += \frac{\mathcal{R}_{(i,i+1),j}{}^2}{\Delta\mathcal{R}_{(i,i+1),j}{}^2}\frac{N_i N_{i+1}}{(N_i + N_{i+1})^2} \tag{A.23}$$

$$u_{i,j} += -\frac{\mathcal{R}_{(i,i+1),j}}{\Delta\mathcal{R}_{(i,i+1),j}{}^2}\frac{N_i}{N_i + N_{i+1}} \tag{A.24}$$

$$u_{i+1,i} += \frac{\mathcal{R}_{(i,i+1),j}{}^2}{\Delta\mathcal{R}_{(i,i+1),j}{}^2}\frac{N_i N_{i+1}}{(N_i + N_{i+1})^2} \tag{A.25}$$

$$u_{i+1,i+1} += \frac{\mathcal{R}_{(i,i+1),j}{}^2}{\Delta\mathcal{R}_{(i,i+1),j}{}^2}\frac{N_{i+1}^2}{(N_i + N_{i+1})^2} \tag{A.26}$$

$$u_{i+1,j} += -\frac{\mathcal{R}_{(i,i+1),j}}{\Delta\mathcal{R}_{(i,i+1),j}{}^2}\frac{N_{i+1}}{N_i + N_{i+1}} \tag{A.27}$$

$$u_{j,i} += -\frac{\mathcal{R}_{(i,i+1),j}}{\Delta\mathcal{R}_{(i,i+1),j}{}^2}\frac{N_i}{N_i + N_{i+1}} \tag{A.28}$$

$$u_{j,i+1} += -\frac{\mathcal{R}_{(i,i+1),j}}{\Delta\mathcal{R}_{(i,i+1),j}{}^2}\frac{N_{i+1}}{N_i + N_{i+1}} \tag{A.29}$$

$$u_{j,j} += \frac{1}{\Delta\mathcal{R}_{(i,i+1),j}{}^2} \tag{A.30}$$

Let's differentiate a right bin combination $X_{i,(j,j+1)}(c_i, c_j, c_{j+1})$ with respect to c_i, c_j and c_{j+1}:

$$\frac{1}{2}\frac{\partial X_{i,(j,j+1)}}{\partial c_i}(c_i, c_j, c_{j+1}) = \frac{1}{\Delta\mathcal{R}_{i,(j,j+1)}{}^2}\left(\mathcal{R}_{i,(j,j+1)}{}^2 c_i - \mathcal{R}_{i,(j,j+1)}\frac{N_j c_j + N_{j+1}c_{j+1}}{N_j + N_{j+1}}\right) \tag{A.31}$$

$$\frac{1}{2}\frac{\partial X_{i,(j,j+1)}}{\partial c_j}(c_i, c_j, c_{j+1}) = \frac{1}{\Delta\mathcal{R}_{i,(j,j+1)}{}^2}\left(-\mathcal{R}_{i,(j,j+1)}\frac{N_j}{N_j + N_{j+1}}c_i + \frac{N_j^2 c_j + N_j N_{j+1} c_{j+1}}{(N_j + N_{j+1})^2}\right) \tag{A.32}$$

$$\frac{1}{2}\frac{\partial X_{i,(j,j+1)}}{\partial c_{j+1}}(c_i, c_j, c_{j+1}) = \frac{1}{\Delta\mathcal{R}_{i,(j,j+1)}{}^2}\left(-\mathcal{R}_{i,(j,j+1)}\frac{N_{j+1}}{N_j + N_{j+1}}c_i + \frac{N_{j+1}^2 c_{j+1} + N_j N_{j+1} c_j}{(N_j + N_{j+1})^2}\right) \tag{A.33}$$

Therefore, the constraint $X_{(i,i+1),j}(c_i, c_{i+1}, c_j)$ contributes to 9 terms in the U matrix:

$$u_{i,i} += \frac{\mathcal{R}_{i,(j,j+1)}{}^2}{\Delta\mathcal{R}_{i,(j,j+1)}{}^2} \tag{A.34}$$

$$u_{i,j} += -\frac{\mathcal{R}_{i,(j,j+1)}}{\Delta\mathcal{R}_{i,(j,j+1)}{}^2}\frac{N_j}{N_j + N_{j+1}} \tag{A.35}$$

$$u_{i,j+1} += -\frac{\mathcal{R}_{i,(j,j+1)}}{\Delta\mathcal{R}_{i,(j,j+1)}{}^2}\frac{N_{j+1}}{N_j + N_{j+1}} \tag{A.36}$$

$$u_{j,i} += -\frac{\mathcal{R}_{i,(j,j+1)}}{\Delta\mathcal{R}_{i,(j,j+1)}{}^2}\frac{N_j}{N_j + N_{j+1}} \tag{A.37}$$

$$u_{j,j} += \frac{1}{\Delta\mathcal{R}_{i,(j,j+1)}{}^2}\frac{N_j^2}{(N_j + N_{j+1})^2} \tag{A.38}$$

$$u_{j,j+1} += \frac{1}{\Delta\mathcal{R}_{i,(j,j+1)}{}^2}\frac{N_j N_{j+1}}{N_j + N_{j+1}} \tag{A.39}$$

$$u_{j+1,i} += -\frac{\mathcal{R}_{i,(j,j+1)}}{\Delta\mathcal{R}_{i,(j,j+1)}{}^2}\frac{N_{j+1}}{N_j + N_{j+1}} \tag{A.40}$$

$$u_{j+1,j} += -\frac{1}{\Delta\mathcal{R}_{i,(j,j+1)}{}^2}\frac{N_j N_{j+1}}{N_j + N_{j+1}} \tag{A.41}$$

$$u_{j+1,j+1} += \frac{1}{\Delta\mathcal{R}_{i,(j,j+1)}{}^2}\frac{N_{j+1}^2}{(N_j + N_{j+1})^2} \tag{A.42}$$

Therefore, the concrete way to compute the U matrix is, after the filling with the terms from F, to loop over each left or right constraint in I' and J' and to add contributions to the appropriate element of U. Thanks to similarities in the terms, the implementation and the book-keeping of all the contributions are not as intricate as what the numerous indexes suggest.

For completion sake[3], it is possible to express the general term of the U matrix.

$$\frac{1}{2}\frac{\partial S^{I'}}{\partial c_k} = \left(\sum_{\substack{(m,m')\in[\![1,N]\!]^2 \\ ((k,m),m')\in I'}} \frac{\mathcal{R}_{(k,m),m'}{}^2}{\Delta\mathcal{R}_{(k,m),m'}{}^2} \frac{N_k^2}{(N_k+N_m)^2} + \sum_{\substack{(m,m')\in[\![1,N]\!]^2 \\ ((m,k),m')\in I'}} \frac{\mathcal{R}_{(m,k),m'}{}^2}{\Delta\mathcal{R}_{(m,k),m'}{}^2} \frac{N_k^2}{(N_m+N_k)^2} \right.$$
$$\left. + \sum_{\substack{(m,m')\in[\![1,N]\!]^2 \\ ((m,m'),k)\in I'}} \frac{1}{\Delta\mathcal{R}_{(m,m'),k}{}^2} \right) c_k$$
$$+ \sum_{l=1}^{N} \left(\sum_{\substack{m\in[\![1,N]\!] \\ ((k,l),m)\in I'}} \frac{\mathcal{R}_{(k,l),m}{}^2}{\Delta\mathcal{R}_{(k,l),m}{}^2} \frac{N_k N_l}{(N_k+N_l)^2} + \sum_{\substack{m\in[\![1,N]\!] \\ ((l,k),m)\in I'}} \frac{\mathcal{R}_{(l,k),m}{}^2}{\Delta\mathcal{R}_{(l,k),m}{}^2} \frac{N_k N_l}{(N_k+N_l)^2} \right.$$
$$- \sum_{\substack{m\in[\![1,N]\!] \\ ((k,m),l)\in I'}} \frac{\mathcal{R}_{(k,m),l}}{\Delta\mathcal{R}_{(k,m),l}{}^2} \frac{N_k}{N_k+N_m} - \sum_{\substack{m\in[\![1,N]\!] \\ ((m,k),l)\in I'}} \frac{\mathcal{R}_{(m,k),l}}{\Delta\mathcal{R}_{(m,k),l}{}^2} \frac{N_k}{N_m+N_k}$$
$$\left. - \sum_{\substack{m\in[\![1,N]\!] \\ ((l,m),k)\in I'}} \frac{\mathcal{R}_{(l,m),k}}{\Delta\mathcal{R}_{(l,m),k}{}^2} \frac{N_l}{N_l+N_m} - \sum_{\substack{m\in[\![1,N]\!] \\ ((m,l),k)\in I'}} \frac{\mathcal{R}_{(m,l),k}}{\Delta\mathcal{R}_{(m,l),k}{}^2} \frac{N_l}{N_m+N_l} \right) c_l \quad \text{(A.43)}$$

$$\frac{1}{2}\frac{\partial S^{J'}}{\partial c_k} = \left(\sum_{\substack{(m,m')\in[\![1,N]\!]^2 \\ (k,(m,m'))\in J'}} \frac{\mathcal{R}_{k,(m,m')}{}^2}{\Delta\mathcal{R}_{k,(m,m')}{}^2} + \sum_{\substack{(m,m')\in[\![1,N]\!]^2 \\ (m,(k,m'))\in J'}} \frac{1}{\Delta\mathcal{R}_{m,(k,m')}{}^2} \frac{N_k^2}{(N_k+N_m)^2} \right.$$
$$\left. + \sum_{\substack{(m,m')\in[\![1,N]\!]^2 \\ (m,(m',k))\in J'}} \frac{1}{\Delta\mathcal{R}_{m,(m',k)}{}^2} \frac{N_k^2}{(N_m+N_k)^2} \right) c_k$$
$$+ \sum_{l=1}^{N} \left(\sum_{\substack{m\in[\![1,N]\!] \\ (k,(l,m))\in J'}} \frac{\mathcal{R}_{k,(l,m)}}{\Delta\mathcal{R}_{k,(l,m)}{}^2} \frac{N_l}{N_l+N_m} - \sum_{\substack{m\in[\![1,N]\!] \\ (l,(k,m))\in J'}} \frac{\mathcal{R}_{l,(k,m)}}{\Delta\mathcal{R}_{l,(k,m)}{}^2} \frac{N_k}{N_k+N_m} \right.$$
$$- \sum_{\substack{m\in[\![1,N]\!] \\ (k,(m,l))\in J'}} \frac{\mathcal{R}_{k,(m,l)}}{\Delta\mathcal{R}_{k,(m,l)}{}^2} \frac{N_l}{N_m+N_l} - \sum_{\substack{m\in[\![1,N]\!] \\ (m,(k,l))\in J'}} \frac{1}{\Delta\mathcal{R}_{m,(k,l)}{}^2} \frac{N_k N_l}{(N_k+N_l)^2}$$
$$\left. - \sum_{\substack{m\in[\![1,N]\!] \\ (l,(m,k))\in J'}} \frac{\mathcal{R}_{l,(m,k)}}{\Delta\mathcal{R}_{l,(m,k)}{}^2} \frac{N_k}{N_m+N_k} + \sum_{\substack{m\in[\![1,N]\!] \\ (m,(l,k))\in J'}} \frac{1}{\Delta\mathcal{R}_{m,(l,k)}{}^2} \frac{N_l N_k}{(N_l+N_k)^2} \right) c_l \quad \text{(A.44)}$$

[3] *I.e.* for fun, because it is useless.

And finally, satisfying the constraints $\forall k \in [\![1, N]\!], \frac{\partial S}{\partial c_k} = 0$, the general term of the U matrix is obtained. It is expressed as follows:

$$
\begin{aligned}
\forall (k,l) \in [\![1,N]\!]^2, u_{kl} = &-\frac{\mathcal{R}_{lk}}{\Delta\mathcal{R}_{lk}{}^2}\mathbb{1}_{l<k}\mathbb{1}_{(l,k)\in F} - \frac{\mathcal{R}_{kl}}{\Delta\mathcal{R}_{kl}{}^2}\mathbb{1}_{l>k}\mathbb{1}_{(k,l)\in F} + \frac{\lambda}{N^2} \\
&+ \Bigg(\sum_{\substack{m\in[\![1,k-1]\!] \\ (m,k)\in F}} \frac{1}{\Delta\mathcal{R}_{mk}{}^2} + \sum_{\substack{m\in[\![k+1,N]\!] \\ (k,m)\in F}} \frac{\mathcal{R}_{km}^2}{\Delta\mathcal{R}_{km}{}^2} \\
&+ \sum_{\substack{(m,m')\in[\![1,N]\!]^2 \\ ((k,m),m')\in I'}} \frac{\mathcal{R}_{(k,m),m'}{}^2}{\Delta\mathcal{R}_{(k,m),m'}{}^2}\frac{N_k^2}{(N_k+N_m)^2} + \sum_{\substack{(m,m')\in[\![1,N]\!]^2 \\ ((m,k),m')\in I'}} \frac{\mathcal{R}_{(m,k),m'}{}^2}{\Delta\mathcal{R}_{(m,k),m'}{}^2}\frac{N_k^2}{(N_m+N_k)^2} \\
&+ \sum_{\substack{(m,m')\in[\![1,N]\!]^2 \\ ((m,m'),k)\in I'}} \frac{1}{\Delta\mathcal{R}_{(m,m'),k}{}^2} + \sum_{\substack{(m,m')\in[\![1,N]\!]^2 \\ (k,(m,m'))\in J'}} \frac{\mathcal{R}_{k,(m,m')}{}^2}{\Delta\mathcal{R}_{k,(m,m')}{}^2} \\
&+ \sum_{\substack{(m,m')\in[\![1,N]\!]^2 \\ (m,(k,m'))\in J'}} \frac{1}{\Delta\mathcal{R}_{m,(k,m')}{}^2}\frac{N_k^2}{(N_k+N_m)^2} + \sum_{\substack{(m,m')\in[\![1,N]\!]^2 \\ (m,(m',k))\in J'}} \frac{1}{\Delta\mathcal{R}_{m,(m',k)}{}^2}\frac{N_k^2}{(N_m+N_k)^2} \Bigg)\mathbb{1}_{l=k} \\
&+ \Bigg(\sum_{\substack{m\in[\![1,N]\!] \\ ((k,l),m)\in I'}} \frac{\mathcal{R}_{(k,l),m}{}^2}{\Delta\mathcal{R}_{(k,l),m}{}^2}\frac{N_kN_l}{(N_k+N_l)^2} + \sum_{\substack{m\in[\![1,N]\!] \\ ((l,k),m)\in I'}} \frac{\mathcal{R}_{(l,k),m}{}^2}{\Delta\mathcal{R}_{(l,k),m}{}^2}\frac{N_kN_l}{(N_k+N_l)^2} \\
&- \sum_{\substack{m\in[\![1,N]\!] \\ ((k,m),l)\in I'}} \frac{\mathcal{R}_{(k,m),l}}{\Delta\mathcal{R}_{(k,m),l}{}^2}\frac{N_k}{N_k+N_m} - \sum_{\substack{m\in[\![1,N]\!] \\ ((m,k),l)\in I'}} \frac{\mathcal{R}_{(m,k),l}}{\Delta\mathcal{R}_{(m,k),l}{}^2}\frac{N_k}{N_m+N_k} \\
&- \sum_{\substack{m\in[\![1,N]\!] \\ ((l,m),k)\in I'}} \frac{\mathcal{R}_{(l,m),k}}{\Delta\mathcal{R}_{(l,m),k}{}^2}\frac{N_l}{N_l+N_m} - \sum_{\substack{m\in[\![1,N]\!] \\ ((m,l),k)\in I'}} \frac{\mathcal{R}_{(m,l),k}}{\Delta\mathcal{R}_{(m,l),k}{}^2}\frac{N_l}{N_m+N_l} \\
&+ \sum_{\substack{m\in[\![1,N]\!] \\ (k,(l,m))\in J'}} \frac{\mathcal{R}_{k,(l,m)}}{\Delta\mathcal{R}_{k,(l,m)}{}^2}\frac{N_l}{N_l+N_m} - \sum_{\substack{m\in[\![1,N]\!] \\ (l,(k,m))\in J'}} \frac{\mathcal{R}_{l,(k,m)}}{\Delta\mathcal{R}_{l,(k,m)}{}^2}\frac{N_k}{N_k+N_m} \\
&- \sum_{\substack{m\in[\![1,N]\!] \\ (k,(m,l))\in J'}} \frac{\mathcal{R}_{k,(m,l)}}{\Delta\mathcal{R}_{k,(m,l)}{}^2}\frac{N_l}{N_m+N_l} - \sum_{\substack{m\in[\![1,N]\!] \\ (m,(k,l))\in J'}} \frac{1}{\Delta\mathcal{R}_{m,(k,l)}{}^2}\frac{N_kN_l}{(N_k+N_l)^2} \\
&- \sum_{\substack{m\in[\![1,N]\!] \\ (l,(m,k))\in J'}} \frac{\mathcal{R}_{l,(m,k)}}{\Delta\mathcal{R}_{l,(m,k)}{}^2}\frac{N_k}{N_m+N_k} + \sum_{\substack{m\in[\![1,N]\!] \\ (m,(l,k))\in J'}} \frac{1}{\Delta\mathcal{R}_{m,(l,k)}{}^2}\frac{N_lN_k}{(N_l+N_k)^2} \Bigg)\mathbb{1}_{l\neq k}
\end{aligned}
\tag{A.45}
$$

Generalizing to the combination of more bins

The generalization is not complicated in the sense that the process is always the same: choose a combination scheme for the constraints, differentiate with respect to the contributing intercalibration factors, and do the book-keeping of the contributions to the U matrix.

For the combination of the constraints, the generalization to a combination of 3 or more bins in the same row or in the same column is straightforward, as one would just add more terms in the sums. But for 3 bins that would not be aligned, or for 4 bins in a square, there are multiple choices already; and the combination of even more bins requires even more arbitrary choices. As already mentioned, combining three or more bins is not well-motivated, because the granularity of the method is crucial and would be lost, and because the linearity assumptions would require more cross-checks.

A.6 The whole calibration

The η-intercalibrations for the three data-taking periods 2015+2016, 2017 and 2018 are presented respectively in Figs. A.2 to A.4 *vs.* η and in Figs. A.5 to A.7 *vs.* p_T. Calibrations are shown for two representative p_T bins ($25 \leq p_T < 45$ GeV and $85 \leq p_T < 115$ GeV) and two representative η bins ($1.2 \leq \eta < 1.4$ and $2.4 \leq \eta < 2.5$), and the fractional uncertainties are presented for two p_T values (35 GeV and 100 GeV) and two η values ($\eta = 1.5$ and $\eta = 2.5$).

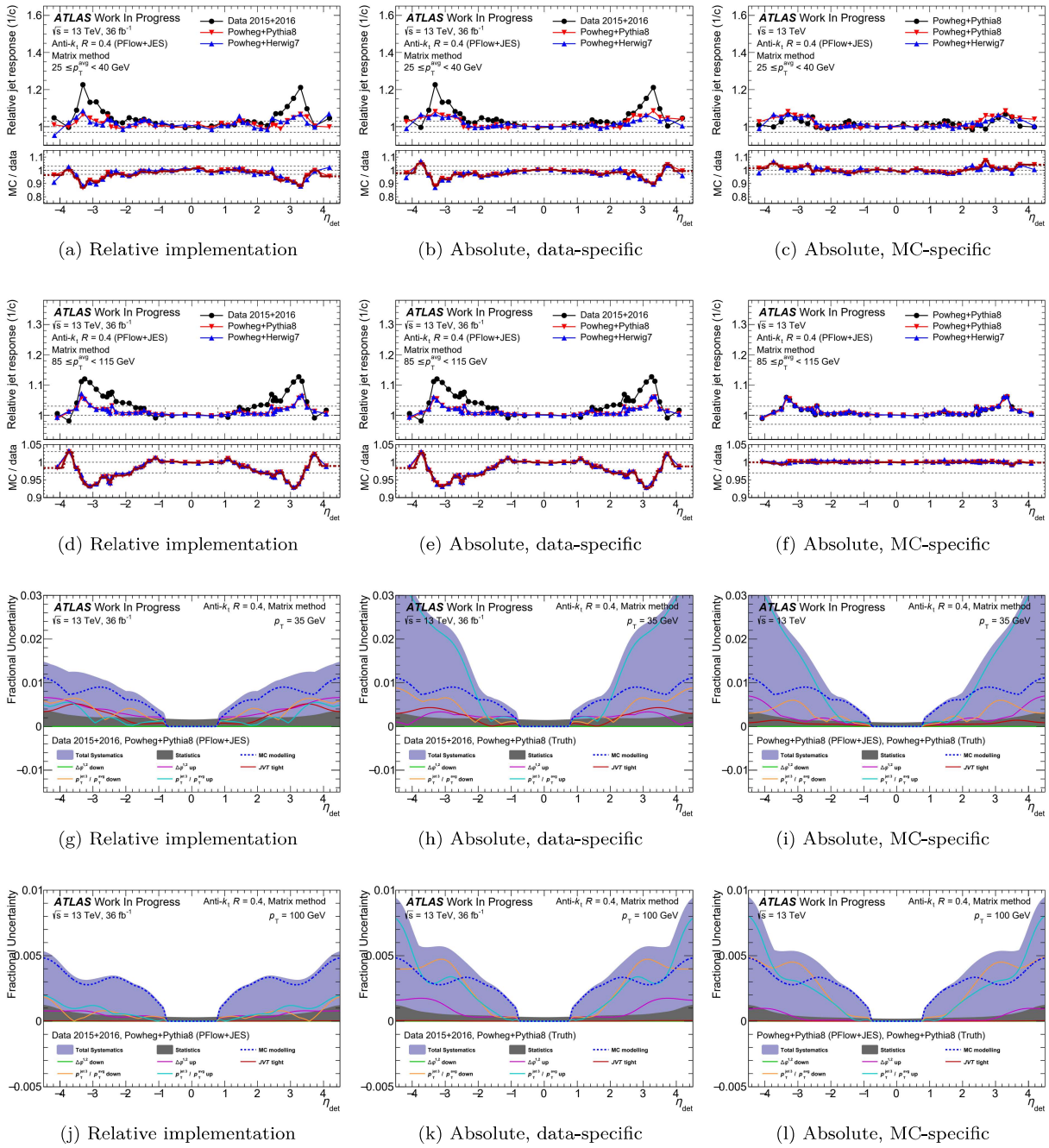
The calibrations use Powheg+Pythia8 as the nominal MC generator and Powheg+Herwig7 as the alternative MC generator to evaluate the MC modeling uncertainty. The matrix method is used to derive the intercalibration coefficients, along with the post-fit propagation method of the statistical uncertainties.

A.7 Precision calibration for large-R jets

I have contributed to the new precision calibration for large-R jets: I ran the first steps of the η-intercalibration framework: `EtaInterCalxAODAnalysis` and `GenerateHistograms`, that respectively apply the event selection and create the asymmetry histograms and the systematic variations. This was part of a team effort by Katherine Dunne, Oleksandr Burlayenko, Davide Melini, Jona Bossio and myself (and most probably other people before I arrived and after I left).

The final precision calibration is shown for the 2017 data-taking period, using Powheg+Pythia8 as the nominal MC generator and Powheg+Herwig7 as the alternative one for the MC modeling uncertainty. The details of the analysis are identical to those of the small-R analysis, with a few exceptions listed in Appendix A.2.

Similarly to the small-R η-intercalibration, some structure is observed, though it is smoother. The calibration increases when $|\eta|$ increases, starting around $|\eta| \approx 1.4$ which is the transition region between the barrel and the endcap of the calorimeter as shown in Fig. 2.22, and reaches a correction of 5 % at the limit of the η region, for $2 < |\eta| < 2.5$. The uncertainties are way below the percent level: outside the central region where they are set to 0 (as for the small-R calibration), they are of the order of 2.5 ‰ at $p_T = 160$ GeV, driven by the MC modeling uncertainty, and decrease when p_T increases, until they reach a sub-permille level around 2 TeV, where multiple uncertainties are of the same order of magnitude.



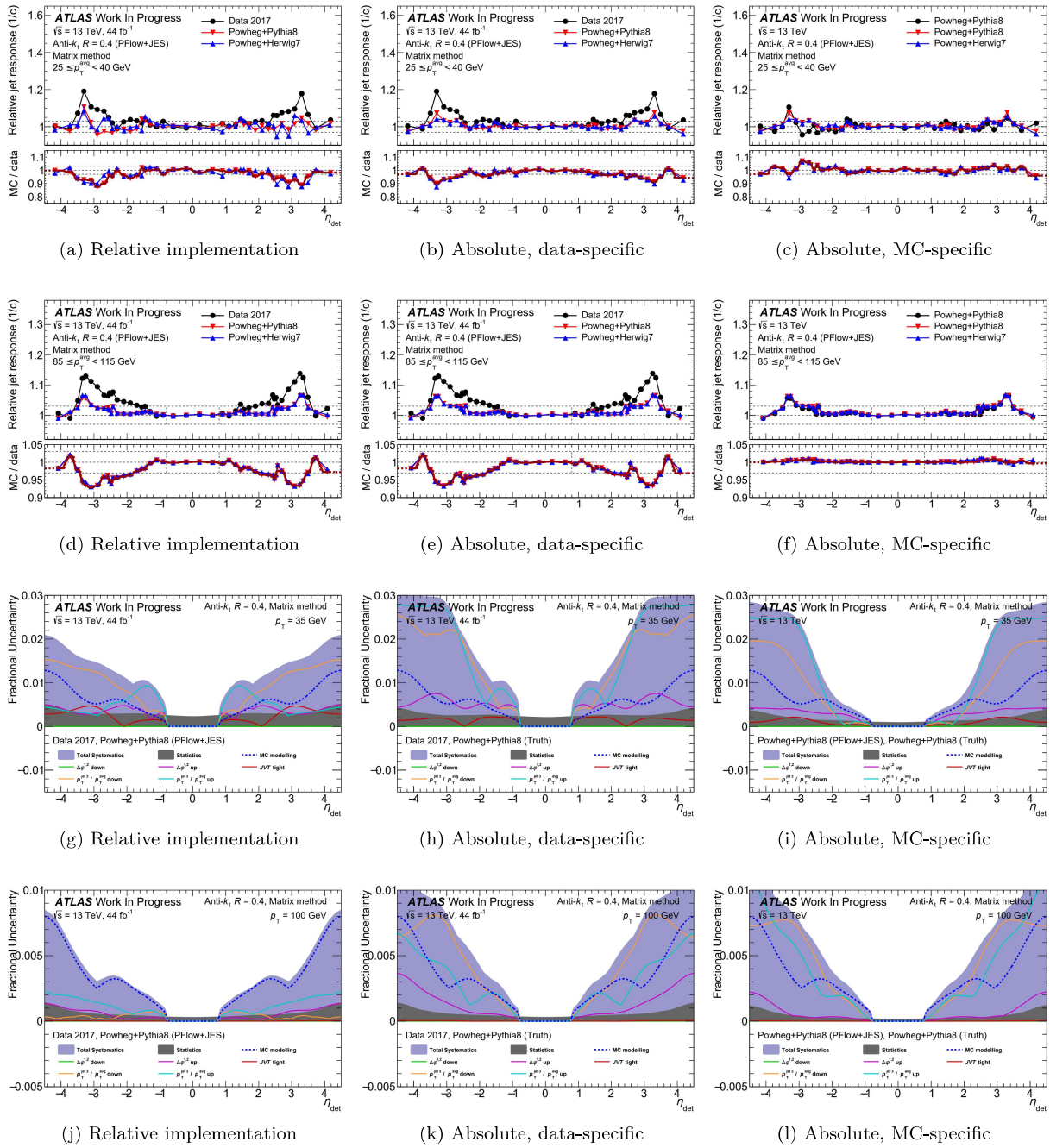
(a) Relative implementation (b) Absolute, data-specific (c) Absolute, MC-specific

(d) Relative implementation (e) Absolute, data-specific (f) Absolute, MC-specific

(g) Relative implementation (h) Absolute, data-specific (i) Absolute, MC-specific

(j) Relative implementation (k) Absolute, data-specific (l) Absolute, MC-specific

Figure A.2: η-intercalibration *vs.* η for the data-taking period 2015+2016. **(a-f)** The upper part of each plot shows the (inverse) intercalibration coefficients $1/c$ *vs.* η; the black dots are the data points (for the relative and absolute coefficients) or Powheg+Pythia8 (for the MC-specific coefficient), the red triangles are the Powheg+Pythia8 points, the blue triangles are the Powheg+Herwig7 points. The lower part of each plot shows the calibration $\mathcal{C}$ *vs.* η; the red points show the nominal calibration points with Powheg+Pythia8 at the denominator of $\mathcal{C}$, the blue points show the alternative calibration points with Powheg+Herwig7 at the denominator of $\mathcal{C}$, used to derive the MC modeling uncertainty; the red curve is the smoothed calibration. The intercalibrations and calibrations are obtained with the fine η binning (Section 4.2.2.2) that follows the structure of the calibration and with the weak smoothing (Section 4.2.7.3). **(g-l)** The systematic uncertainties on the η-intercalibration are obtained with the coarse η binning (Sections 4.2.2.2 and 4.2.8.2) and with the strong smoothing (sections Sections 4.2.7.3 and 4.2.8.2). **(a,b,c)** η-intercalibration *vs.* η in the $25 \leq p_T < 45$ GeV bin, **(d,e,f)** η-intercalibration *vs.* η in the $85 \leq p_T < 115$ GeV bin, **(g,h,i)** fractional uncertainties on the η-intercalibration *vs.* η at 35 GeV and **(j,k,l)** fractional uncertainties on the η-intercalibration *vs.* η at 100 GeV. **(a,d,g,j)** show the relative implementation, **(b,e,g,k)** show the data-specific calibration for the absolute implementation and **(c,f,i,l)** show the MC-specific calibration for the absolute implementation.



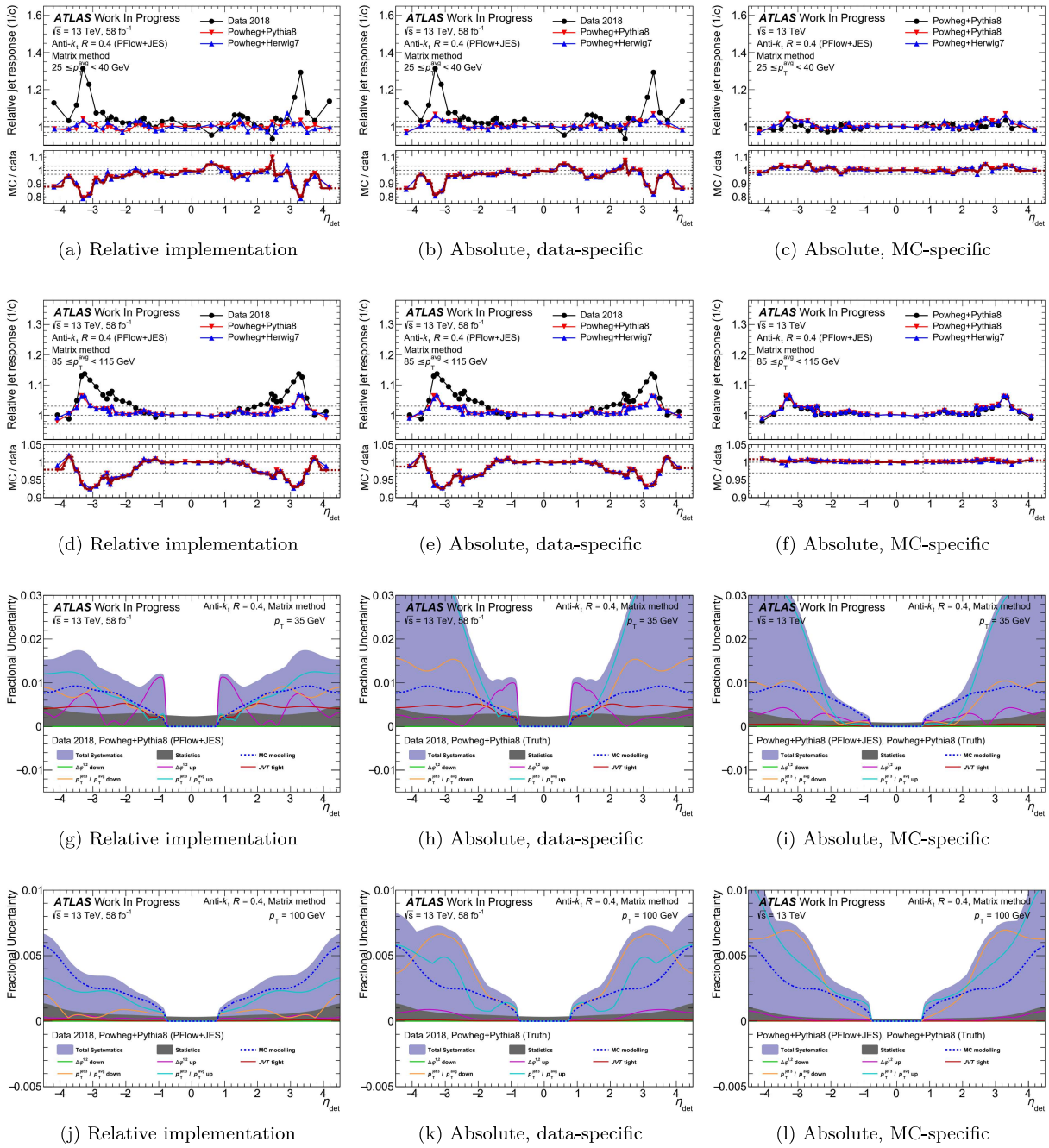
(a) Relative implementation (b) Absolute, data-specific (c) Absolute, MC-specific

(d) Relative implementation (e) Absolute, data-specific (f) Absolute, MC-specific

(g) Relative implementation (h) Absolute, data-specific (i) Absolute, MC-specific

(j) Relative implementation (k) Absolute, data-specific (l) Absolute, MC-specific

Figure A.3: η-intercalibration *vs.* η for the data-taking period 2017. **(a-f)** The upper part of each plot shows the (inverse) intercalibration coefficients $1/c$ *vs.* η; the black dots are the data points (for the relative and absolute coefficients) or Powheg+Pythia8 (for the MC-specific coefficient), the red triangles are the Powheg+Pythia8 points, the blue triangles are the Powheg+Herwig7 points. The lower part of each plot shows the calibration $\mathcal{C}$ *vs.* η; the red points show the nominal calibration points with Powheg+Pythia8 at the denominator of $\mathcal{C}$, the blue points show the alternative calibration points with Powheg+Herwig7 at the denominator of $\mathcal{C}$, used to derive the MC modeling uncertainty; the red curve is the smoothed calibration. The intercalibrations and calibrations are obtained with the fine η binning (Section 4.2.2.2) that follows the structure of the calibration and with the weak smoothing (Section 4.2.7.3). **(g-l)** The systematic uncertainties on the η-intercalibration are obtained with the coarse η binning (Sections 4.2.2.2 and 4.2.8.2) and with the strong smoothing (Sections 4.2.7.3 and 4.2.8.2). **(a,b,c)** η-intercalibration *vs.* η in the $25 \leq p_T < 45$ GeV bin, **(d,e,f)** η-intercalibration *vs.* η in the $85 \leq p_T < 115$ GeV bin, **(g,h,i)** fractional uncertainties on the η-intercalibration *vs.* η at 35 GeV and **(j,k,l)** fractional uncertainties on the η-intercalibration *vs.* η at 100 GeV. **(a,d,g,j)** show the relative implementation, **(b,e,g,k)** show the data-specific calibration for the absolute implementation and **(c,f,i,l)** show the MC-specific calibration for the absolute implementation.



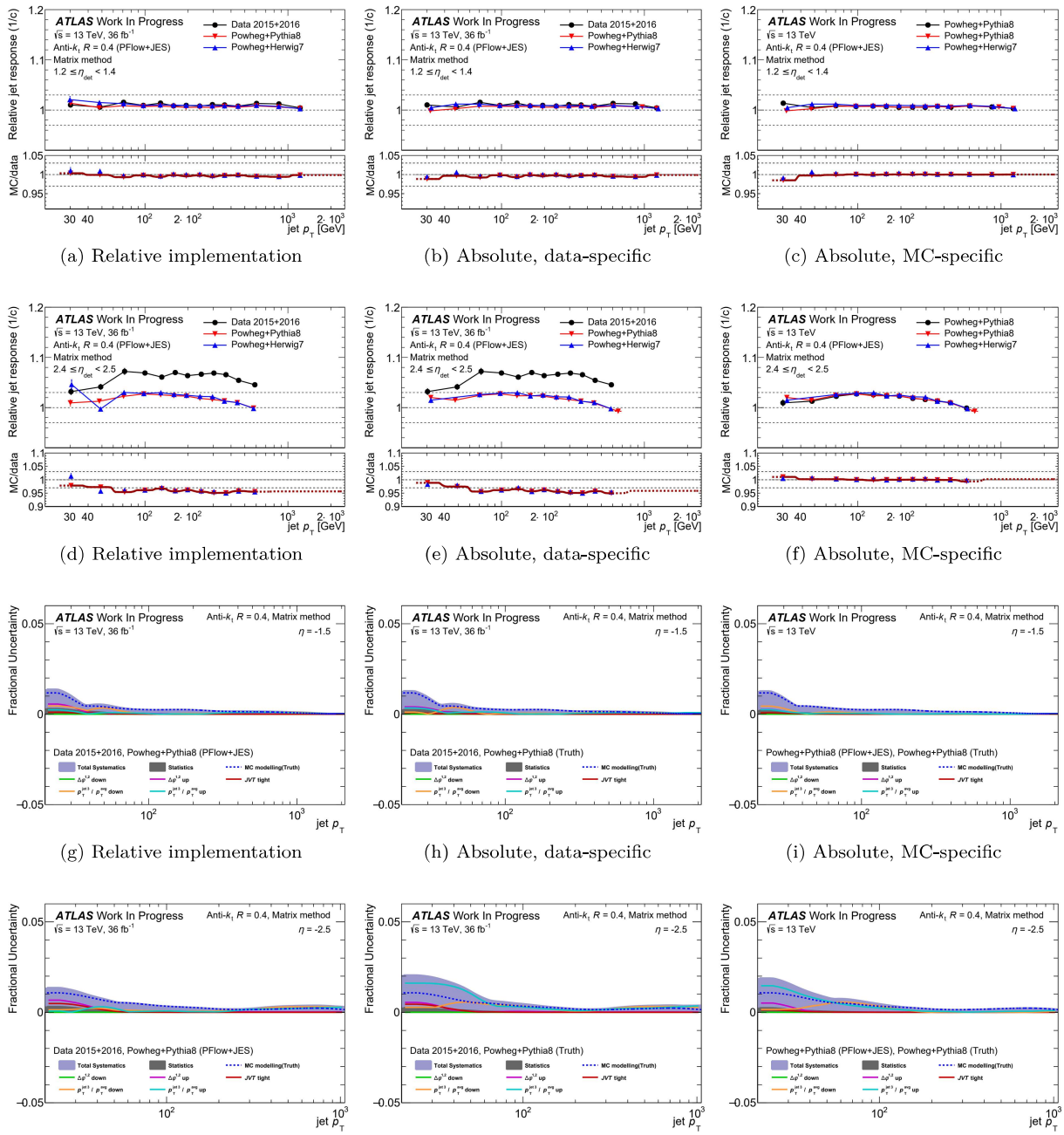
(a) Relative implementation (b) Absolute, data-specific (c) Absolute, MC-specific

(d) Relative implementation (e) Absolute, data-specific (f) Absolute, MC-specific

(g) Relative implementation (h) Absolute, data-specific (i) Absolute, MC-specific

(j) Relative implementation (k) Absolute, data-specific (l) Absolute, MC-specific

Figure A.4: η-intercalibration *vs.* η for the data-taking period 2018. **(a-f)** The upper part of each plot shows the (inverse) intercalibration coefficients $1/c$ *vs.* η; the black dots are the data points (for the relative and absolute coefficients) or Powheg+Pythia8 (for the MC-specific coefficient), the red triangles are the Powheg+Pythia8 points, the blue triangles are the Powheg+Herwig7 points. The lower part of each plot shows the calibration $\mathcal{C}$ *vs.* η; the red points show the nominal calibration points with Powheg+Pythia8 at the denominator of $\mathcal{C}$, the blue points show the alternative calibration points with Powheg+Herwig7 at the denominator of $\mathcal{C}$, used to derive the MC modeling uncertainty; the red curve is the smoothed calibration. The intercalibrations and calibrations are obtained with the fine η binning (Section 4.2.2.2) that follows the structure of the calibration and with the weak smoothing (Section 4.2.7.3). **(g-l)** The systematic uncertainties on the η-intercalibration are obtained with the coarse η binning (Sections 4.2.2.2 and 4.2.8.2) and with the strong smoothing (Sections 4.2.7.3 and 4.2.8.2). **(a,b,c)** η-intercalibration *vs.* η in the $25 \leq p_T < 45$ GeV bin, **(d,e,f)** η-intercalibration *vs.* η in the $85 \leq p_T < 115$ GeV bin, **(g,h,i)** fractional uncertainties on the η-intercalibration *vs.* η at 35 GeV and **(j,k,l)** fractional uncertainties on the η-intercalibration *vs.* η at 100 GeV. **(a,d,g,j)** show the relative implementation, **(b,e,g,k)** show the data-specific calibration for the absolute implementation and **(c,f,i,l)** show the MC-specific calibration for the absolute implementation.



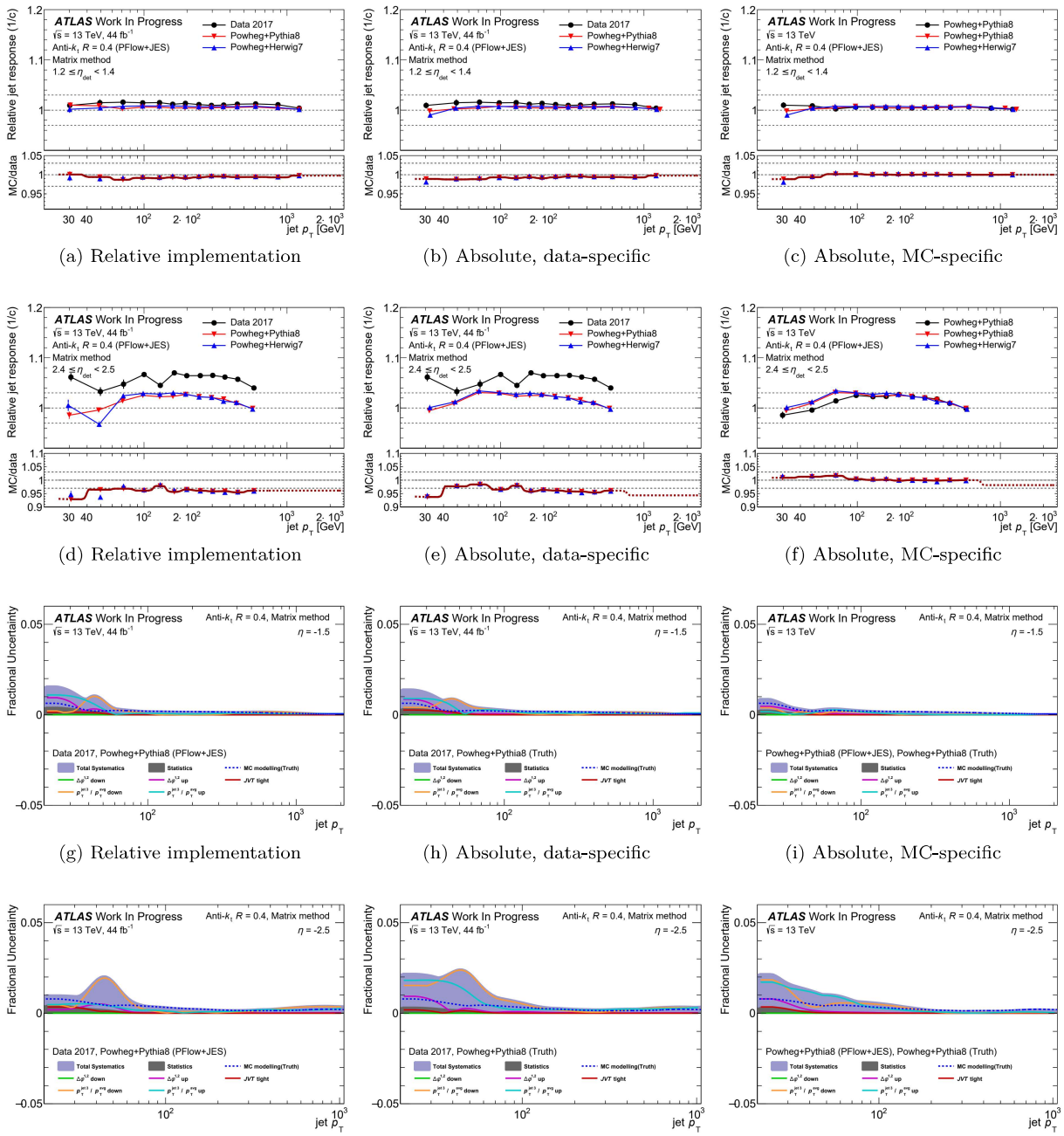
(a) Relative implementation (b) Absolute, data-specific (c) Absolute, MC-specific

(d) Relative implementation (e) Absolute, data-specific (f) Absolute, MC-specific

(g) Relative implementation (h) Absolute, data-specific (i) Absolute, MC-specific

(j) Relative implementation (k) Absolute, data-specific (l) Absolute, MC-specific

Figure A.5: η-intercalibration *vs.* p_T for the data-taking period 2015+2016. **(a-f)** The upper part of each plot shows the (inverse) intercalibration coefficients $1/c$ *vs.* p_T; the black dots are the data points (for the relative and absolute coefficients) or Powheg+Pythia8 (for the MC-specific coefficient), the red triangles are the Powheg+Pythia8 points, the blue triangles are the Powheg+Herwig7 points. The lower part of each plot shows the calibration $\mathcal{C}$ *vs.* p_T; the red points show the nominal calibration points with Powheg+Pythia8 at the denominator of $\mathcal{C}$, the blue points show the alternative calibration points with Powheg+Herwig7 at the denominator of $\mathcal{C}$, used to derive the MC modeling uncertainty; the red curve is the smoothed calibration; the dashed curve is the extrapolation. The intercalibrations and calibrations are obtained with the fine η binning (Section 4.2.2.2) that follows the structure of the calibration and with the weak smoothing (Section 4.2.7.3). **(g-l)** The systematic uncertainties on the η-intercalibration are obtained with the coarse η binning (Sections 4.2.2.2 and 4.2.8.2) and with the strong smoothing (Sections 4.2.7.3 and 4.2.8.2). **(a,b,c)** η-intercalibration *vs.* p_T in the $1.2 \leq \eta < 1.4$ bin, **(d,e,f)** η-intercalibration *vs.* p_T in the $2.4 \leq \eta < 2.5$ bin, **(g,h,i)** fractional uncertainties on the η-intercalibration *vs.* p_T at $\eta = 1.5$ and **(j,k,l)** fractional uncertainties on the η-intercalibration *vs.* p_T at $\eta = 1.5$. **(a,d,g,j)** show the relative implementation, **(b,e,g,k)** show the data-specific calibration for the absolute implementation and **(c,f,i,l)** show the MC-specific calibration for the absolute implementation.



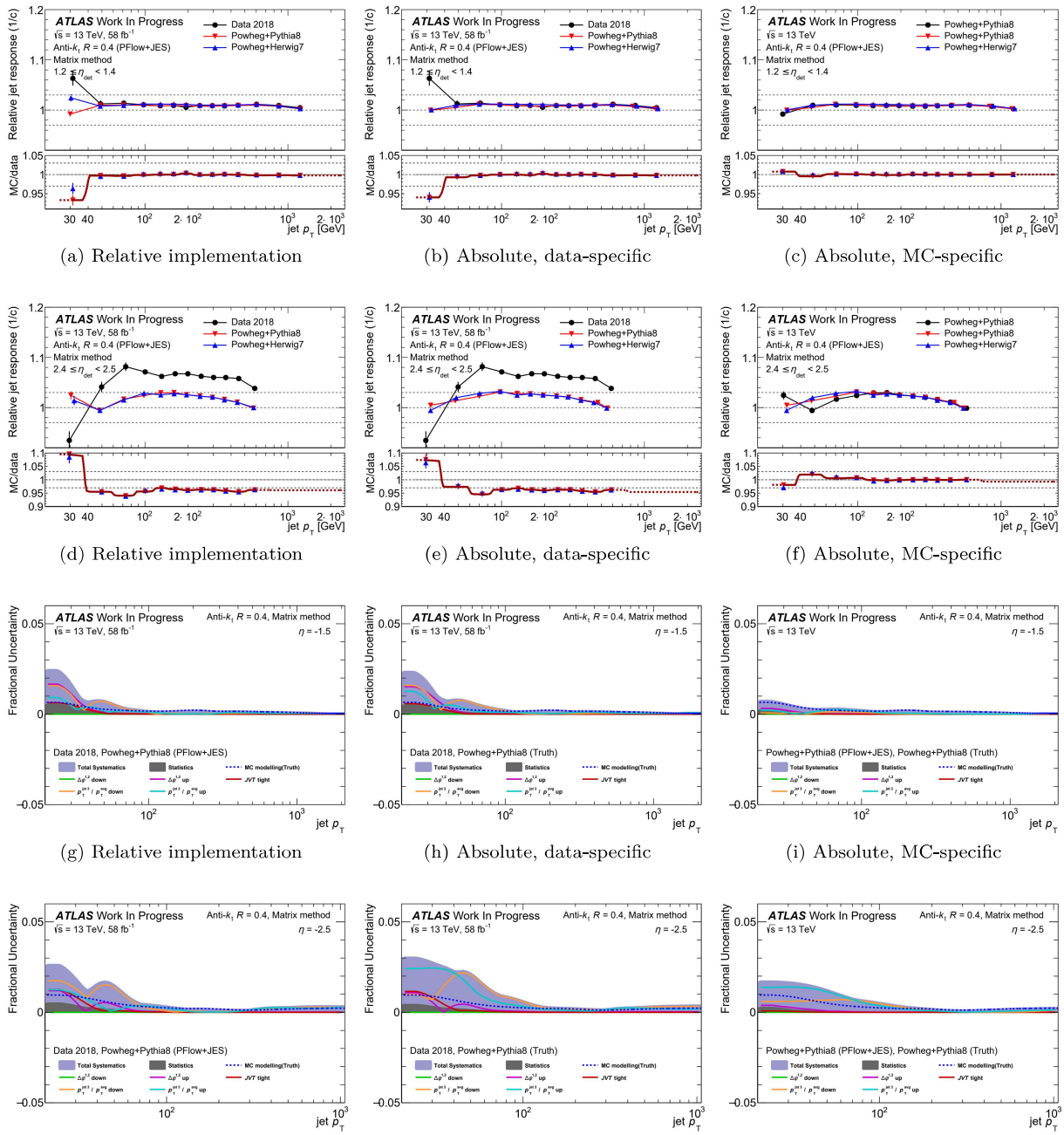
(a) Relative implementation (b) Absolute, data-specific (c) Absolute, MC-specific

(d) Relative implementation (e) Absolute, data-specific (f) Absolute, MC-specific

(g) Relative implementation (h) Absolute, data-specific (i) Absolute, MC-specific

(j) Relative implementation (k) Absolute, data-specific (l) Absolute, MC-specific

Figure A.6: η-intercalibration *vs.* p_T for the data-taking period 2017. **(a-f)** The upper part of each plot shows the (inverse) intercalibration coefficients $1/c$ *vs.* p_T; the black dots are the data points (for the relative and absolute coefficients) or Powheg+Pythia8 (for the MC-specific coefficient), the red triangles are the Powheg+Pythia8 points, the blue triangles are the Powheg+Herwig7 points. The lower part of each plot shows the calibration $\mathcal{C}$ *vs.* p_T; the red points show the nominal calibration points with Powheg+Pythia8 at the denominator of $\mathcal{C}$, the blue points show the alternative calibration points with Powheg+Herwig7 at the denominator of $\mathcal{C}$, used to derive the MC modeling uncertainty; the red curve is the smoothed calibration; the dashed curve is the extrapolation. The intercalibrations and calibrations are obtained with the fine η binning (Section 4.2.2.2) that follows the structure of the calibration and with the weak smoothing (Section 4.2.7.3). **(g-l)** The systematic uncertainties on the η-intercalibration are obtained with the coarse η binning (Sections 4.2.2.2 and 4.2.8.2) and with the strong smoothing (Sections 4.2.7.3 and 4.2.8.2). **(a,b,c)** η-intercalibration *vs.* p_T in the $1.2 \leq \eta < 1.4$ bin, **(d,e,f)** η-intercalibration *vs.* p_T in the $2.4 \leq \eta < 2.5$ bin, **(g,h,i)** fractional uncertainties on the η-intercalibration *vs.* p_T at $\eta = 1.5$ and **(j,k,l)** fractional uncertainties on the η-intercalibration *vs.* p_T at $\eta = 1.5$. **(a,d,g,j)** show the relative implementation, **(b,e,g,k)** show the data-specific calibration for the absolute implementation and **(c,f,i,l)** show the MC-specific calibration for the absolute implementation.



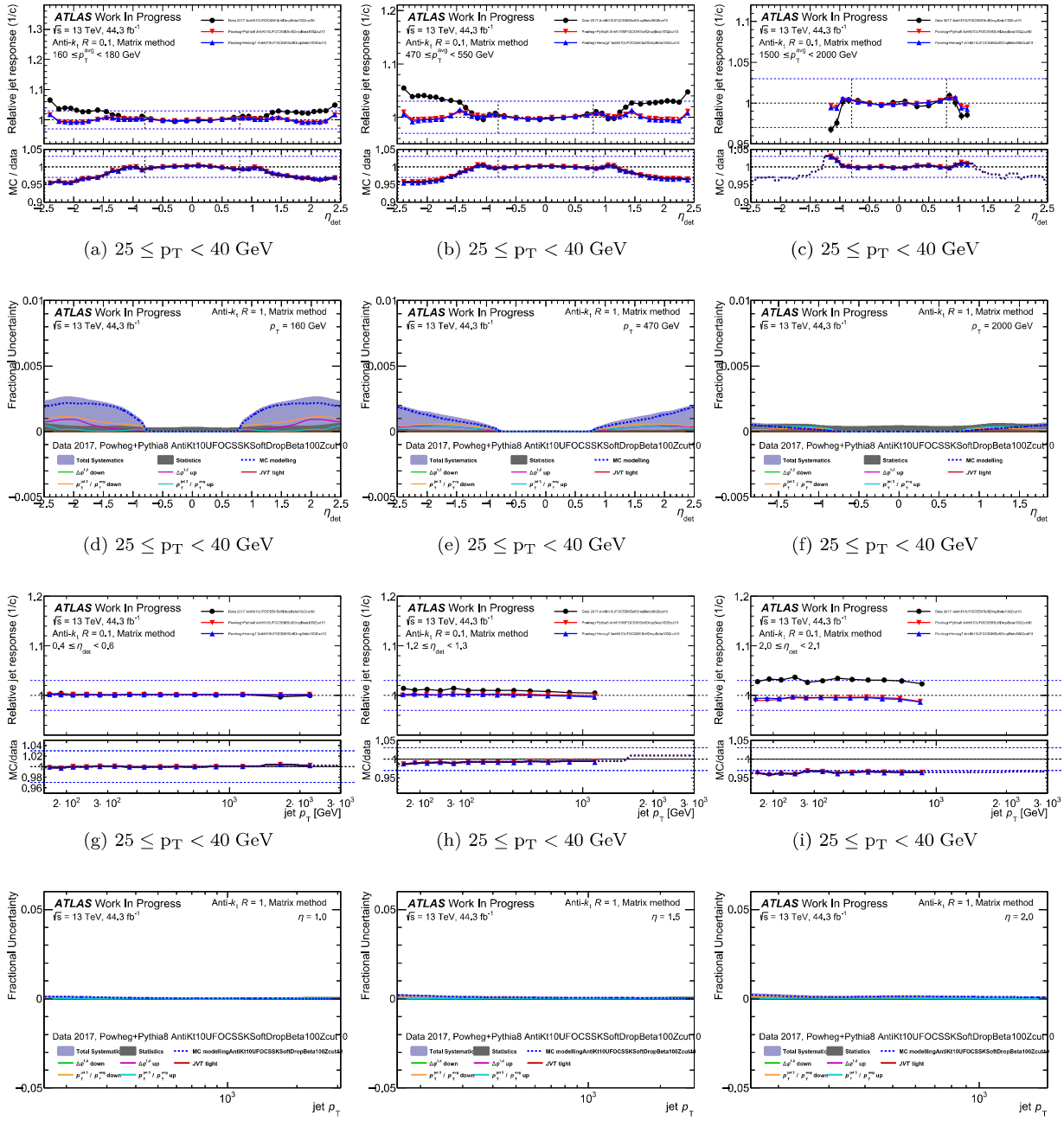
(a) Relative implementation (b) Absolute, data-specific (c) Absolute, MC-specific

(d) Relative implementation (e) Absolute, data-specific (f) Absolute, MC-specific

(g) Relative implementation (h) Absolute, data-specific (i) Absolute, MC-specific

(j) Relative implementation (k) Absolute, data-specific (l) Absolute, MC-specific

Figure A.7: η-intercalibration *vs.* p_T for the data-taking period 2018. **(a-f)** The upper part of each plot shows the (inverse) intercalibration coefficients $1/c$ *vs.* p_T; the black dots are the data points (for the relative and absolute coefficients) or POWHEG+PYTHIA8 (for the MC-specific coefficient), the red triangles are the POWHEG+PYTHIA8 points, the blue triangles are the POWHEG+HERWIG7 points. The lower part of each plot shows the calibration $\mathcal{C}$ *vs.* p_T; the red points show the nominal calibration points with POWHEG+PYTHIA8 at the denominator of $\mathcal{C}$, the blue points show the alternative calibration points with POWHEG+HERWIG7 at the denominator of $\mathcal{C}$, used to derive the MC modeling uncertainty; the red curve is the smoothed calibration; the dashed curve is the extrapolation. The intercalibrations and calibrations are obtained with the fine η binning (Section 4.2.2.2) that follows the structure of the calibration and with the weak smoothing (Section 4.2.7.3). **(g-l)** The systematic uncertainties on the η-intercalibration are obtained with the coarse η binning (Sections 4.2.2.2 and 4.2.8.2) and with the strong smoothing (Sections 4.2.7.3 and 4.2.8.2). **(a,b,c)** η-intercalibration *vs.* p_T in the $1.2 \leq \eta < 1.4$ bin, **(d,e,f)** η-intercalibration *vs.* p_T in the $2.4 \leq \eta < 2.5$ bin, **(g,h,i)** fractional uncertainties on the η-intercalibration *vs.* p_T at $\eta = 1.5$ and **(j,k,l)** fractional uncertainties on the η-intercalibration *vs.* p_T at $\eta = 1.5$. **(a,d,g,j)** show the relative implementation, **(b,e,g,k)** show the data-specific calibration for the absolute implementation and **(c,f,i,l)** show the MC-specific calibration for the absolute implementation.



(a) $25 \leq p_T < 40$ GeV (b) $25 \leq p_T < 40$ GeV (c) $25 \leq p_T < 40$ GeV

(d) $25 \leq p_T < 40$ GeV (e) $25 \leq p_T < 40$ GeV (f) $25 \leq p_T < 40$ GeV

(g) $25 \leq p_T < 40$ GeV (h) $25 \leq p_T < 40$ GeV (i) $25 \leq p_T < 40$ GeV

(j) $25 \leq p_T < 40$ GeV (k) $25 \leq p_T < 40$ GeV (l) $25 \leq p_T < 40$ GeV

Figure A.8: The η-intercalibration for large-R jets and the associated uncertainties for the data-taking period 2017, for the relative implementation, **(a-f)** *vs.* η or **(g-l)** *vs.* p_T. **(a-c,g-i)** Calibration. The upper part of each plot shows the (inverse) intercalibration coefficients $1/c$; the black dots are data, the red triangles are Powheg+Pythia8, the blue triangles are Powheg+Herwig7. The lower part of each plot shows the calibration $\mathcal{C}$; the red points show the nominal calibration obtained with Powheg+Pythia8, the blue points show the alternative calibration obtained with Powheg+Herwig7 used to derive the MC modeling uncertainty; the red curve is the smoothed calibration. **(d-f,j-l)** Uncertainties. Note that the MC modeling is not obtained at truth level, contrarily to the MC modeling for the small-R jets (see discussion in Section 4.2.8.7).

Appendix B

Appendix for the Cross-section Measurements

B.1 Detailed list of all the JES and JER uncertainties

List of all the JES and JER systematic uncertainties propagated through the unfolding, with the internal names from the software provider of JetETmiss, are given in Tables B.1 and B.2. Explanations for the origin of all the uncertainties are given in Section 3.3.

Category	Name	Number
pile-up	OffsetMu	2
	OffsetNPV	2
	PtTerm	2
	RhoTopology	2
Flavor	Composition_prop	2
	Response_prop	2
	PerJet_GenShower	2
	PerJet_GenShower_HF	2
	PerJet_Hadronization	2
	PerJet_Hadronization_HF	2
	PerJet_Shower	2
	PerJet_Shower_HF	2
HighPt	PunchThrough	2
	SingleParticle	2
EtaIntercalibration	201516_Stat	532×2
	2017_Stat	532×2
	NonClosure_highE	2
	NonClosure_negEta	2
	NonClosure_posEta	2
	TotalSyst	2
	NonClosure_2018data	2
	InSituProp_EtaIntercalStat	2
MJB	Stat	28×2
	Alpha	2
	Asym	2
	Beta	2
	Fragmentation	2
	JVT	2
	Threshold	2
Gjet	Stat	16×2
	GamESZee	14×2
	GamEsmear	2
	Generator	2
	Jvt	2
	Purity	2
	ShowerTopology	2
	Veto	2
	dPhi	2

Table B.1: Breakdown of the systematic uncertainties from pile-up, flavor, high-p_T, η-intercalibration, multijet balance, γ+jet balance. The $\times 2$ everywhere accounts for the up and down variations.

Category	Name	Number
Zjet	ElecStat	14×2
	MuStat	14×2
	ElecESEzz	2
	ElecEsmear	2
	Jvt	2
	MC	2
	MuSagittaRes	2
	MuSagittaRho	2
	MuScale	2
	MuSmearID	2
	MuSmearMS	2
	ShowerTopology	2
	Veto	2
	dPhi	2
JER	Stat	12×2
	DataVsMC_MC16	2
	N_constscale_noise_up_central	2
	N_constscale_noise_up_forward	2
	N_fit_conversion_hist_variation_central	2
	N_fit_conversion_hist_variation_forward	2
	N_fit_error_up_central	2
	N_fit_error_up_forward	2
	N_fitrange_high_up_central	2
	N_fitrange_high_up_forward	2
	N_mc_non_closure_variation_central	2
	N_mc_non_closure_variation_forward	2
	N_no_pu_N_fit_instability_variation_central	2
	N_no_pu_N_fit_instability_variation_forward	2
	N_no_pu_N_unc_up_central	2
	N_no_pu_N_unc_up_forward	2
	dijet_closure	2
	dijet_jesnp1	2
	dijet_jesnp2	2
	dijet_jesnp3	2
	dijet_jv	2
	dijet_mcgenerator	2
	dijet_pt3dphi	2

Table B.2: Breakdown of the systematic uncertainties from Z+jet balance and JER. The $\times 2$ everywhere accounts for the up and down variations.